NeuroImage 112 (2015) 138–151

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Functional brain networks underlying detection and integration of disconfirmatory evidence

Katie M. Lavigne, Paul D. Metzak, Todd S. Woodward *

Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada
BC Mental Health and Addictions Research Institute, Provincial Health Services Authority, Vancouver, BC, Canada

ARTICLE INFO

Article history:
Accepted 20 February 2015
Available online 28 February 2015

Keywords:
Disconfirmatory evidence integration
Functional connectivity
fMRI

ABSTRACT

Processing evidence that disconfirms a prior interpretation is a fundamental aspect of belief revision, and has clear social and clinical relevance. This complex cognitive process requires (at minimum) an alerting stage and an integration stage, and in the current functional magnetic resonance imaging (fMRI) study, we used multivariate analysis methodology on two datasets in an attempt to separate these sequentially-activated cognitive stages and link them to distinct functional brain networks. Thirty-nine healthy participants completed one of two versions of an evidence integration experiment involving rating two consecutive animal images, both of which consisted of two intact images of animal faces morphed together at different ratios (e.g., 70/30 bird/dolphin followed by 10/90 bird/dolphin). The two versions of the experiment differed primarily in terms of stimulus presentation and timing, which facilitated functional interpretation of brain networks based on differences in the hemodynamic response shapes between versions. The data were analyzed using constrained principal component analysis for fMRI (fMRI-CPCA), which allows distinct, simultaneously active task-based networks to be separated, and these were interpreted using both temporal (task-based hemodynamic response shapes) and spatial (dominant brain regions) information. Three networks showed increased activity during integration of disconfirmatory relative to confirmatory evidence: (1) a network involved in alerting to the requirement to revise an interpretation, identified as the salience network (dorsal anterior cingulate cortex and bilateral insula); (2) a sensorimotor response-related network (pre- and post-central gyri, supplementary motor area, and thalamus); and (3) an integration network involving rostral prefrontal, orbitofrontal and posterior parietal cortex. These three networks were staggered in their peak activity (alerting, responding, then integrating), but at certain time points (e.g., 17 s after trial onset) the hemodynamic responses associated with all three networks were simultaneously active. These findings highlight distinct cognitive processes and corresponding functional brain networks underlying stages of disconfirmatory evidence integration, and demonstrate the power of multivariate and multi-experiment methodology in cognitive neuroscience.

The evaluation and integration of evidence that disconfirms a prior belief is a fundamental aspect of belief revision. Failures in evidence integration, and particularly in the ability to integrate disconfirmatory evidence, has social relevance as it can lead to resistance in modifying outdated or unhelpful beliefs (Turner and Pratkanis, 1998), and has clinical relevance as it has been linked to delusions in schizophrenia (Sanford et al., 2014; Speechley et al., 2011; Woodward et al., 2006), and to self-regulation deficits in traumatic brain injury (Flashman and McAllister, 2002) and obsessive–compulsive disorder (Marsh et al., 2014).

Evidence integration involves multiple cognitive processes, including alerting to the piece of evidence in question, and integration of that evidence into the current belief. When evidence contradicts a currently-held belief (i.e., disconfirmatory evidence), this would increase demand for alerting and integrating processes, as the initial belief must either be revised or discarded in order to assimilate the newly-accepted evidence and maintain a coherent belief system. When the evidence is neutral, or consistent with a belief (i.e., confirmatory evidence), these cognitive processes would be expected to have a reduced role. To date, there have been few investigations into the functional brain networks underlying disconfirmatory evidence integration, and it is not known whether distinct, sequentially-active brain networks that correspond to alerting and integration processes can be measured. However, the left inferior frontal gyrus (IFG) has been implicated in disconfirmatory evidence integration, with previous studies finding improved integration following transcranial magnetic stimulation (Sharot et al., 2012). The dorsal anterior cingulate cortex (dACC), with regard to its role in adjusting behavior and changing mental set (Behrens et al., 2007; Whitman et al., 2013; Woodward et al., 2008), may play a role

* Corresponding author at: Room A3-A117, BC Mental Health & Addictions Research Institute, Translational Research Building, 3rd Floor, 938W. 28th Avenue, Vancouver, British Columbia, Canada, V5Z 4H4. Fax: +1 604 875 3871.
E-mail address: Todd.S.Woodward@gmail.com (T.S. Woodward).

http://dx.doi.org/10.1016/j.neuroimage.2015.02.043
1053-8119/

in alerting. In the current functional magnetic resonance imaging (fMRI) study, we used multivariate analysis methodology on two datasets to attempt to identify functional brain networks underlying different stages of disconfirmatory evidence integration.

In order to assess spatial and temporal replication of network configurations, and take advantage of spatial replication combined with temporal differences to interpret function of brain networks, two versions of an evidence integration experiment were run and analyzed simultaneously using constrained principal component analysis for fMRI (fMRI-CPCA; Lavigne et al., 2014; Metzak et al., 2011, 2012; Whitman et al., 2013; Woodward et al., 2013). fMRI-CPCA allows observation of coordinated task-based activity of multiple distinct, sequentially-active functional brain networks based on distinct hemodynamic response (HDR) shapes and spatial distributions. fMRI-CPCA determines the degree to which each functional brain network replicates across experiments by comparing the magnitude and pattern of the HDR shape associated with each network. When two (or more) experiment versions elicit the same underlying cognitive operation (e.g., evidence integration), spatial and temporal replication would be observed if HDR shapes were not distinguishable between the two experiment versions, and this should be the case if the timing of the cognitive operation does not differ between experiments. In contrast, spatial but not temporal replication would be observed if HDR shapes were reliably different between the two experiment versions, and this should be the case if the timing of the cognitive operation differs between experiments. This case (spatial but not temporal replication) provides an important scientific opportunity to use differences between experiments to help interpret the cognitive function of brain networks. Finally, if a cognitive operation is elicited by only one version of the experiment but not the other, the version not eliciting this cognitive operation would show a flat HDR shape for that functional brain network, and therefore it could be concluded that neither spatial nor temporal replication has been observed.

In the current study, we examined the functional brain networks underlying disconfirmatory evidence integration by combining data from two versions of an evidence integration task. The main distinction between the two experiment versions was a persistent visual display throughout the trial in version 1, and the removal of the visual display during rating in version 2. This was expected to elicit distinct HDR shapes for visual-processing brain networks between versions, producing spatial but not temporal replication for visual-processing networks, but similar HDR shapes for evidence integration brain networks, producing spatial and temporal replication for evidence integration brain networks. This method will facilitate separation of cognitive processes underlying visual processing from those related specifically to the alerting to and integration of disconfirmatory evidence. In accordance with the two-stage process mentioned above, we hypothesized that two separable and sequentially active functional networks (viz., alerting followed by integration), would be associated with disconfirmatory evidence integration to a greater degree than confirmatory evidence integration, and would not be associated with pure visual processing.

Material and methods

2.1. Participants

Participants were 39 healthy volunteers (Version 1: 10 male, 10 female, mean age = 24.90, SD = 6.87; Version 2: 9 male, 10 female, mean age = 26.84, SD = 7.34), most of which were native English speakers (Version 1: 17 participants; Version 2: 15 participants). Non-native English speakers had been using English daily for at least the past five years and responded accurately to questions about the consent form designed to confirm their ability to read and understand English. All participants were right-handed (Annett, 1970), with the exception of one left-handed and two mixed-handed participants who completed Version 2. Participants were recruited via advertisements and word-of-mouth from Vancouver, British Columbia, and participated in exchange for $10/h and a copy of their structural brain images. All were screened for MRI compatibility, and gave written informed consent prior to participation. All experimental procedures were approved by the University of British Columbia clinical research ethics board.

2.2. Experimental design

Participants completed one of two versions of a novel perceptual interpretation task while undergoing functional magnetic resonance imaging (fMRI). In Version 1, each trial began with a brief (500 ms) presentation of a heavily distorted image (Adobe Photoshop effects: 50 random noise, brightness −80, mosaic 8 & 8, ripple 5, 5, 50, 50; see Figs. 1A and B) of two animals (e.g., animal A = bird; animal B = dolphin) morphed together at a ratio of 60:40 or 40:60 (animal A/animal B). Participants were presented with a 16-point rating scale and were asked to indicate the degree to which the image appeared to be of one animal or the other. After 6 s, or once a rating was made, a mildly distorted image (brightness −50, mosaic 8 & 8) of the same animals morphed together at a ratio of 60:40 (animal A/animal B) was displayed on screen for 3 s, and participants were asked to re-rate the image. This led to the design of two types of trials: confirm (image 1: 60% animal A; image 2: 60% animal A); and disconfirm (image 1: 40% animal A; image 2: 60% animal A).

Version 2 differed from version 1 primarily in the following respects (see Fig. 1B): (1) removal of images during presentation of the rating scales; and (2) the addition of a backwards mask lasting 250 ms between the offset of the first image and the onset of the first rating scale. These changes removed the ability to visually process the images when responding, facilitating separation of visual-processing networks from those underlying alerting and evidence integration. In addition, (3) the morphing ratios were increased to 70:30 for image 1 and 10:90 for image 2 in an attempt to intensify the disconfirmatory evidence presented in image 2; (4) the name of either animal A or animal B was centered above the rating scale in version 2 rather than both names appearing at opposite ends of the scale, which ensured greater variability in participants' responses (i.e., selecting a degree of belief towards one animal rather than choosing between one or the other); and (5) jittered inter-trial intervals (ITIs) of 2, 4, 6, 8, 10, and 20 s (rather than the 2 s ITI in version 1) were included to optimize the deconvolution of the BOLD signal (Serences, 2004).

2.3. Response conditions

For each trial, participants rated each of the two images on a 16-point scale to describe the degree to which they believed the image depicted the queried animal(s). In order to emphasize that participants were to revise their initial ratings after viewing the second image, participants' ratings on the first image were preserved on the second rating scale, and ratings were modified from that point. Assignment of all experimental conditions (for both versions) was based on participants' rating changes from image 1 to image 2. These response-based conditions were labeled no change, confirm, and disconfirm. The *no change* response condition included trials in which participants' ratings changed by less than or equal to two points on the rating scale in either direction (e.g., image 1 rating = 9, image 2 rating = 7). The *confirm* response condition consisted of trials in which the initial rating was supported by the second rating. Specifically, this refers to trials in which ratings did not cross the mid-point of the scale (8) *and* where image 2 was rated closer to the extremes of the scale (e.g., image 1 rating = 6, image 2 rating = 3; or image 1 rating = 9, image 2 rating = 14). The *disconfirm* response condition consisted of trials in which the second rating contradicted the initial rating, such that ratings *either* crossed the mid-point of the scale (8) or image 2 was rated closer to the middle of the scale (e.g., image 1 rating = 4, image 2 rating = 9; or image 1 rating = 15, image 2 rating = 11). All response conditions were created such that they were mutually-exclusive (i.e., trials with rating changes

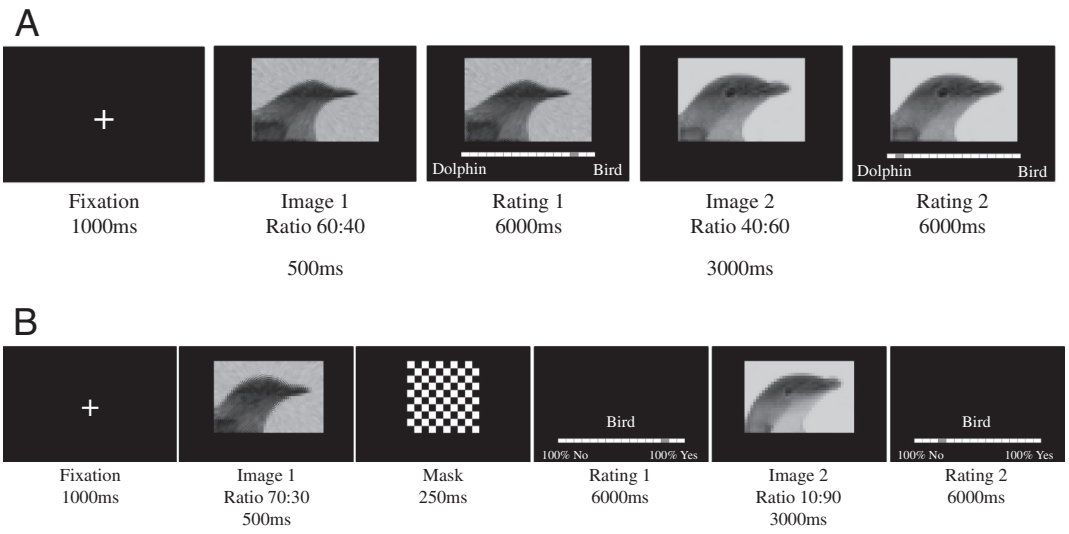


Fig. 1. A–B. Timeline of the evidence integration tasks (disconfirm condition). Each trial began with the presentation of a distorted image of two animals (e.g., bird and dolphin) morphed together at a ratio of 60:40 (Version 1) or 70:30 (Version 2) for 500 ms. After a 250 ms mask (Version 2 only), participants were presented with a 16-point rating scale and were asked to indicate the degree to which the image appeared to be of one animal or the other. After 6 s, or once a rating was made, a less distorted image of the same animals morphed together at a different ratio was displayed on screen for 3 s, and participants were asked to re-rate the image. A = Version 1; B = Version 2.

of less than two that fit under either *confirm* or *disconfirm* were classified as *no change*).

2.4. Image acquisition and processing

Imaging was performed at the University of British Columbia MRI Research Centre on a Philips Achieva 3.0 Tesla (T) MRI scanner with quasar dual gradients (maximum gradient amplitude, 80 mT/m; maximum slew rate, 200 mT/m/s). The participant's head was firmly secured using a customized head holder. Functional image volumes were collected using a T2*-weighted gradient-echo spin pulse sequence with 36 axial slices; thickness/gap, 3/1 mm; matrix, 80 × 80; repetition time (TR), 2000 ms; echo time (TE), 30 ms; flip angle (FA), 90°, field of view (FOV), 240 × 240 mm, effectively covering the whole brain. In version 1, between 288 and 296 images were acquired in each of 3 runs lasting approximately 9 min and 52 s each. In Version 2, 350 volumes were acquired in each of two runs lasting 11 min and 40 s each. For both versions, run order was randomly assigned for each participant in order to minimize order effects.

Functional images were pre-processed using Statistical Parametric Mapping 8 (SPM8; Wellcome Trust Centre for Neuroimaging, UK). For each participant, each functional run was corrected for slice-timing, realigned, co-registered to their structural (T1) image, and subsequently normalized to the Montreal Neurological Institute (MNI) T1 brain template (voxel size = 2 × 2 × 2 mm). All images were spatially smoothed with an 8 × 8 × 8 mm full width at half maximum Gaussian filter. Runs for which motion correction exceeded 4 mm or degrees were excluded from analysis. This led to the exclusion of four runs across four participants, two in each experiment version.

2.5. Data analysis

2.5.1. Functional connectivity

fMRI data analysis was carried out using constrained principal component analysis for fMRI (fMRI-CPCA) with orthogonal rotation (Lavigne et al., 2014; Metzak et al., 2011, 2012; Whitman et al., 2013; Woodward et al., 2013). The theory and proofs of CPCA are detailed in previously published work (Hunter and Takane, 2002; Takane and Hunter, 2001; Takane and Shibayama, 1991) and the fMRI-CPCA application is available on-line, free of charge (www.nitrc.org/projects/fmricpca). Briefly, fMRI-CPCA combines multivariate multiple regression analysis and principal component analysis into a unified framework to reveal multiple independent sources of poststimulus fluctuations in brain activity. fMRI-CPCA is able (1) to identify multiple functional brain networks simultaneously involved in executing a cognitive task, (2) to estimate the task-related time course of coordinated BOLD activity fluctuations associated with each functional network, and (3) to statistically test the effect of experimental manipulations and group differences on BOLD activity associated with each functional brain network.

2.5.2. Matrix equations

We now present a brief summary of the logic and matrix equations for fMRI-CPCA. Broadly speaking, whole-brain BOLD activity variance was partitioned into (i.e., constrained to) task-related fluctuations using multivariate multiple regression. Orthogonal sources (components) of task-related BOLD activity fluctuations were then determined using PCA. Functional brain networks associated with each orthogonal source of BOLD variance were spatially interpreted by viewing the networks represented by voxels dominating each component, and temporally interpreted by viewing the HDR shape associated with each component.

To begin, two matrices were prepared for further analysis. The first matrix, *Z*, contained the intensity values for normalized and smoothed BOLD time-series of each voxel, with one column per voxel and one row per repetition time (TR) or scan. Subject-specific datasets were stacked vertically to produce *Z*. The second matrix, *G*, consisted of a finite impulse response (FIR) basis set, which was used to estimate the change in BOLD signal at specific poststimulus scans relative to all other scans. The value 1 is placed in rows of *G* for which BOLD signal amplitude is to be estimated, and the value 0 in all other rows ("mini boxcar" functions). The time points for which a basis function was specified in the current study were the 1st to 12th scans following stimulus presentation. Since the TR for these data was 2 s, this resulted in estimating BOLD signal over a 24 s window, with the start of the first time point (time = 0) corresponding to stimulus onset. In this analysis we created a *G* matrix for estimating subject-and-condition specific effects by including a separate FIR basis set for each condition and for each subject. The columns in this subject-and-condition based *G* matrix code 12 poststimulus time points for each of the three conditions (viz., no change, confirm, and disconfirm) for each of the 39 subjects, totaling 1404 columns (12 × 3 × 39 = 1404). Each column of *Z* and *G* was standardized for each subject separately.

The matrix of BOLD time series (Z) and the design matrix (G) were input to fMRI-CPCA, with BOLD signal in Z being predicted from the FIR model in G. In order to achieve this, multivariate least-squares linear multiple regression was carried out, whereby the BOLD time series (Z) was regressed onto the design matrix (G):

$$Z = GC + E, \quad (1)$$

where $C = (G'G)^{-1} G'Z$. The C matrix represents condition-specific regression weights, which are akin to the beta images produced by conventional univariate fMRI analyses. GC represents the variability in Z that was predictable from the design matrix G, that is to say, the task-related variability in Z.

The next step used singular value decomposition (of which PCA is a special case) to extract components in GC that represented temporally orthogonal functional brain networks in which BOLD activity fluctuated coherently with experimental stimuli. The singular value decomposition of GC resulted in:

$$UDV' = GC \quad (2)$$

where U = matrix of left singular vectors; D = diagonal matrix of singular values; and V = matrix of right singular vectors. After reduction of dimensionality (discussed in more detail below) and orthogonal rotation (Metzak et al., 2011) each column of $VD/\sqrt{(m-1)}$, where m = number of rows in Z, was overlaid on a structural brain image to allow spatial visualization of the brain regions dominating each functional network. $VD/\sqrt{(m-1)}$ is referred to as a *loading matrix*, and the values are correlations between the component scores (in U) and the variables in GC.

2.5.3. Predictor weights

To interpret the functional brain networks with respect to the conditions represented in G, *predictor weights* in matrix P are produced. These are the weights that, when applied to each column of the matrix of predictor variables (G), create U ($U = GP$). Thus, the P matrix relates each column of the G matrix to the component scores in U, and provides information about the similarity of the fluctuation of the BOLD signal over all scans to the FIR model coded into G. For the current analysis, this would provide 1404 values per functional brain network, one for each combination of poststimulus time (12), subject (39), and condition (3). Each subject- and condition-specific set of predictor weights is expected to take the shape of a HDR, with the highest values corresponding to the HDR peaks.

These predictor weights provide estimates of the engagement of functional networks at each point in poststimulus time, and can be submitted to an analysis of variance (ANOVA) to test for (1) reliability of each component over subjects, (2) differences between conditions in the activation of each functional brain network, and (3) differences between experiment versions in the activation of each functional network. These analyses were carried out as 12 × 3 × 2 mixed-model ANOVAs (one for each component extracted), with the within-subjects factors of Poststimulus Time (12 whole-brain scans after the onset of each trial were estimated in the FIR model) and Response Condition (no change, confirm, and disconfirm), and the between-subjects factor of Version (version 1, version 2). Any impact of Version or Response Condition would typically be reflected by a significant interaction with Poststimulus Time for the measure of estimated HDR (i.e., the predictor weights), suggesting that the HDR shape depends on Version or Response Condition, although main effects are also possible. Significant interactions were interpreted using analysis of simple main effects involving the relevant factors. Spatial and temporal replication would be indicated by a reliable HDR shape over subjects (i.e., a significant Poststimulus Time effect) and no difference between experiment versions (i.e., no significant Version main effect or Version × Poststimulus Time interaction effect). Spatial but not temporal replication would be indicated by a reliable HDR shape over subjects (i.e., a significant Poststimulus Time effect) and a significant difference between experiment versions (i.e., a significant Version main effect or Version × Poststimulus Time interaction effect). Spatial (and temporal) non-replication would be indicated by a reliable HDR shape over subjects in only one experiment version (i.e., a non significant Poststimulus Time effect at one but not the other level of Version) and a difference between experiment versions (i.e., a significant Version × Poststimulus Time interaction effect). Tests of sphericity were carried out for all ANOVAs, and adjustment in degrees of freedom for violations of sphericity did not affect the results; therefore, the original degrees of freedom are reported.

Results

Inspection of the scree plot of singular values (Cattell, 1966; Cattell and Vogelmann, 1977) suggested that five components should be extracted. The percentages of task-related variance accounted for by each rotated component were 10.88%, 10.43%, 9.00%, 7.64%, and 6.20%, for Components 1 to 5, respectively. For Component 3,[1] no main effects or interactions involving Response Condition or Version were significant so it is not discussed further, but details about this component are available from the corresponding author. Visual inspection of the predictor weights for each component confirmed a HDR shape (see Figs. 2 to 5, for Components 1, 2, 4, and 5 respectively). Components 1, 2, 4, and 5 showed a significant effect of Poststimulus Time, $F(11,374) = 47.47$, $p < .001$, $\eta_p^2 = 0.58$; $F(11,374) = 55.99$, $p < .001$, $\eta_p^2 = 0.62$, $F(11,374) = 56.30$, $p < .001$, $\eta_p^2 = 0.62$; $F(11,374) = 38.70$, $p < .001$, $\eta_p^2 = 0.53$, respectively, demonstrating detection of a biologically plausible and reliable HDR signal for each functional brain network (Metzak et al., 2011, 2012; Woodward et al., 2013).

3.1. Anatomical descriptions and relations to experimental conditions

The brain regions associated with Components 1, 2, 4, and 5 are displayed in Figs. 2A to 5A, with anatomical descriptions in Tables 1 to 4, respectively. All components showed spatial but not temporal replication, described in detail below.

3.1.1. Component 1: Integration Network

Component 1 was characterized by a functional network that included activations in rostral prefrontal and orbitofrontal cortex (rPFC & OFC; BAs 10, 11, 47), bilateral inferior frontal gyrus (IFG; BAs 6, 38), right dorsolateral prefrontal cortex (DLPFC; BA 46), superior parietal cortex (extending into angular and supramarginal gyri; BAs 2, 40), and bilateral cerebellar and occipital (BAs 17, 18, 19) regions. This network showed significant Poststimulus Time × Version, $F(11,374) = 10.52$, $p < .001$, $\eta_p^2 = 0.24$, and Poststimulus Time × Response Condition, $F(22,748) = 6.29$, $p < .001$, $\eta_p^2 = 0.16$, interactions, but no significant three-way interaction. This suggests that the HDR shape associated with Component 1 depended on Version and Response Condition, but that each could be interpreted independently. In order to interpret the Version effect, simple contrasts averaging over Response Condition were observed (Fig. 2B), and revealed significant differences between versions at 9, 11, and 21 s (all $ps < .005$), due to higher activity for version 2 relative to version 1. In order to interpret the Response Condition effect, simple contrasts averaging over Version were observed (Fig. 2C), and revealed significantly greater activity for (1) the confirm relative to no change response conditions at 17 s, (2) the disconfirm relative to no change response conditions from 17 to 21 s, and (3) the disconfirm relative to

[1] Component 3 included activations in bilateral intracalcarine cortex (BAs 17, 18), lingual gyrus (BA 19), pre- and post-central gyri (BAs 3, 6), ventromedial prefrontal cortex (BAs 9, 10), and posterior cingulate cortex (BAs 23, 30). This component showed a significant main effect of peristimulus time, $F(11,374) = 28.75$, $p < .001$, $\eta_p^2 = 0.46$, but no other significant main effects or interactions were present, suggesting that although it was a biologically plausible network, activity was not related to the experimental conditions of interest.

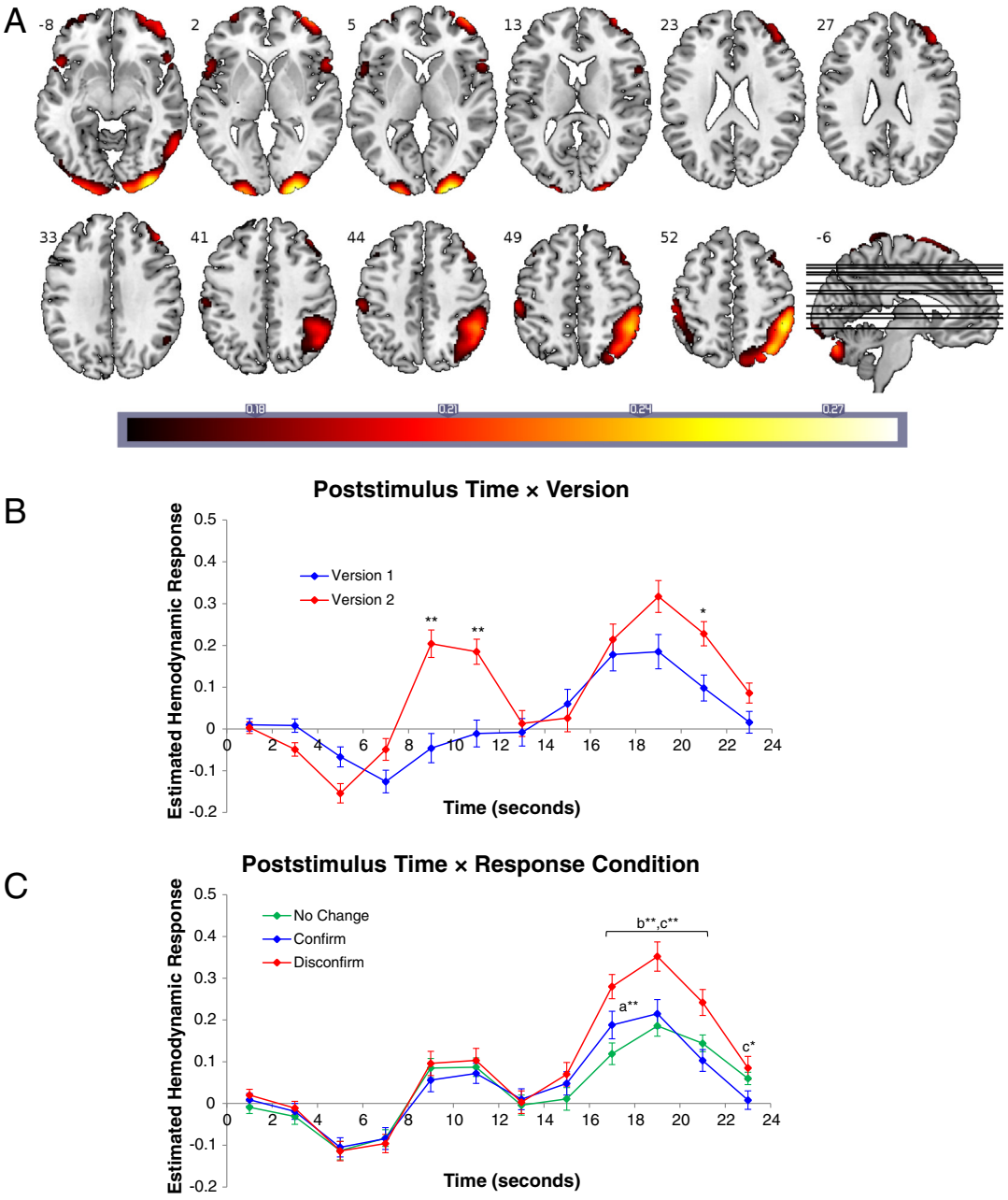


Fig. 2. A–C. A: Dominant 10% of component loadings for Component 1 (Integration Network; red/yellow = positive loadings, threshold = 0.16, max = 0.28; no negative loadings passed threshold). Montreal Neurological Institute Z-axis coordinates are displayed. B: Mean finite impulse response (FIR)-based predictor weights for Component 1 averaged over conditions and plotted as a function of poststimulus time. C: Mean FIR-based predictor weights for Component 1 averaged over versions and plotted as a function of poststimulus time; [a] = confirm > no change; [b] = disconfirm > no change; [c] = disconfirm > confirm. Error bars are standard errors. * = $p < .005$, ** = $p < .001$.

confirm response conditions from 17 to 23 s (all *ps* < .005). Thus, activity in this network was highest for the disconfirm response condition after the onset of the second image, when the disconfirmatory evidence was presented, and remained elevated throughout the remainder of the trial. Since Response Condition did not interact with Version, this pattern can be considered present in both experiments. Based on these differences between response conditions and the spatial distribution of the network, this network was labeled *Integration Network*.

3.1.2. Component 2: Visual/Default-Mode Network

Component 2 was characterized by a functional network including activations in bilateral occipital cortex (BAs 17, 18, 19), superior parietal (BA 7) regions, and bilateral middle frontal gyrus (BAs 44, 45). This network also included deactivations (negative loadings) in ventromedial prefrontal cortex (VmPFC; BAs 9, 10), precuneus and posterior cingulate gyrus (BA 23), bilateral anterior middle temporal gyrus (BA 21), and bilateral angular/supramarginal gyri (BAs 39, 40), regions commonly associated with the default-mode network (DMN; Buckner et al., 2008; Raichle and MacLeod, 2001). Component 2 showed significant Poststimulus Time × Version, $F(11,374) = 9.35$, $p < .001$, $\eta_p^2 = 0.22$, and Poststimulus Time × Response Condition, $F(22,748) = 3.57$, $p < .001$, $\eta_p^2 = 0.10$, interactions, but no significant three-way interaction. In order to interpret the Version effect, simple contrasts averaging over Response Condition were observed (Fig. 3B), and revealed significant differences between versions at 1, 3, and 7–23 s (all *ps* < .01). This was attributable to greater activity in version 1 relative to version 2 across

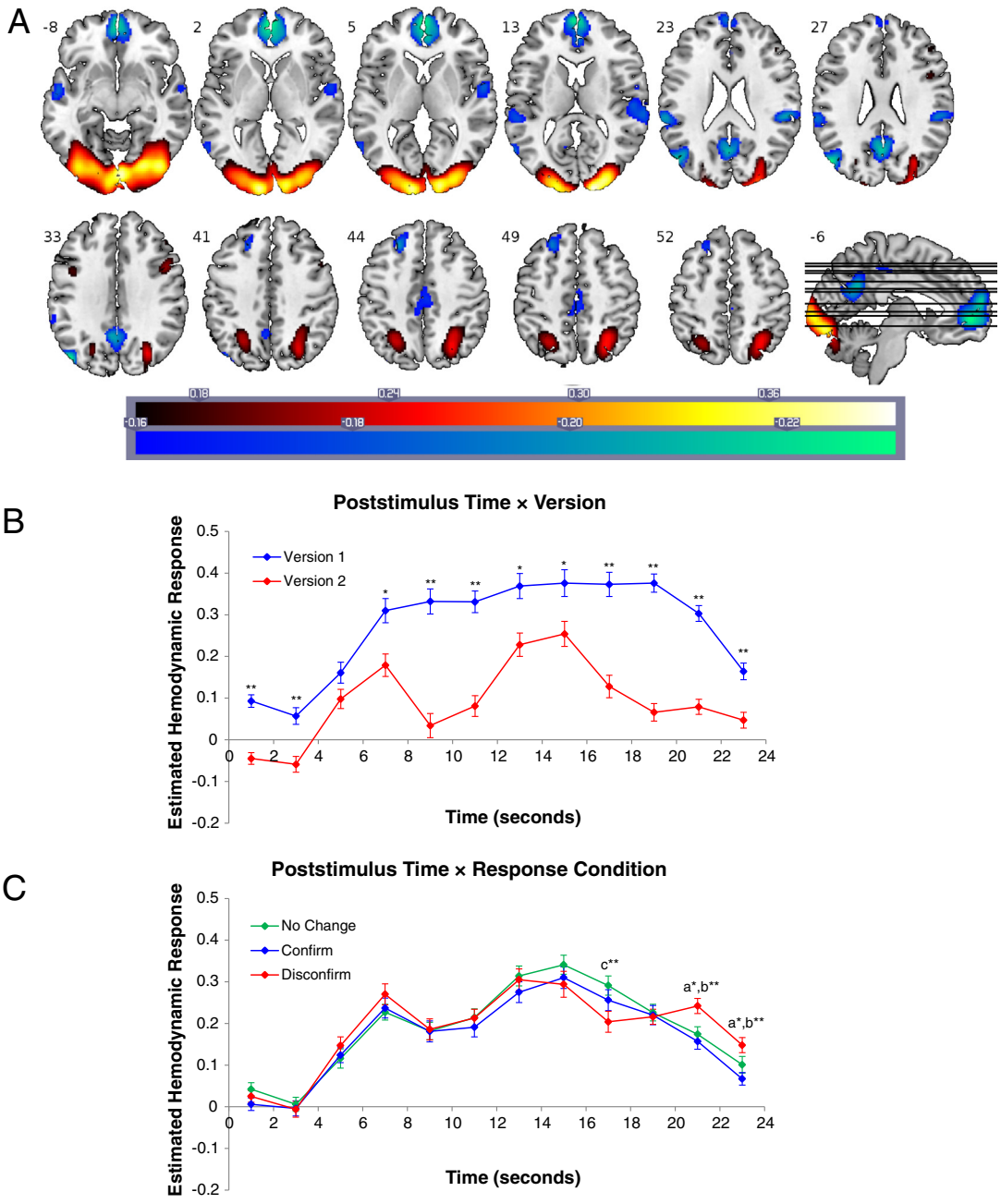
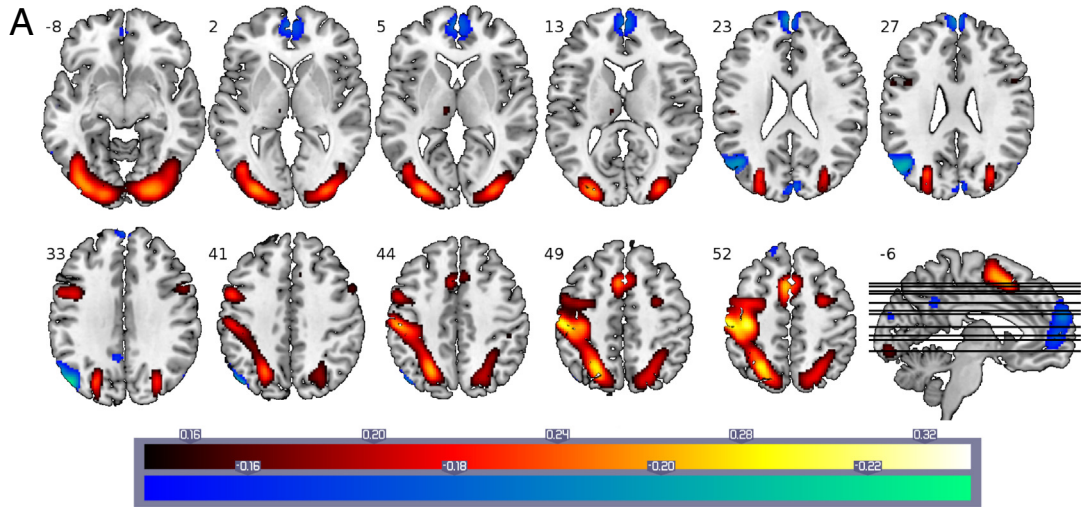


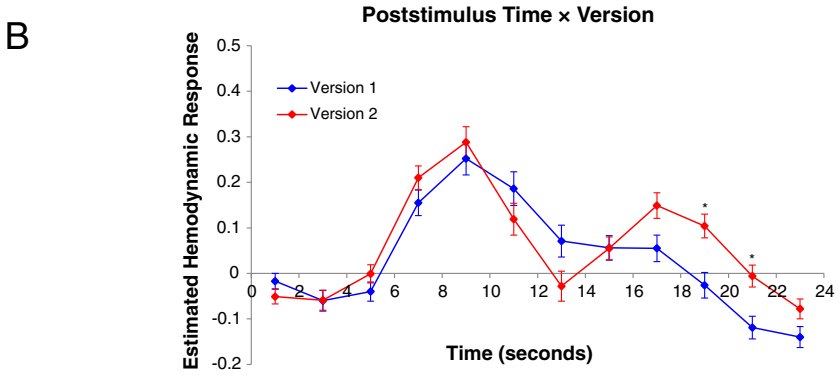
Fig. 3. A–C. A: Dominant 10% of component loadings for Component 2 (Visual/Default-Mode Network; red/yellow = positive loadings, threshold = 0.16, max = 0.40; blue/green = negative loadings, threshold = −0.13, min = −0.23). Montreal Neurological Institute Z-axis coordinates are displayed. B: Mean finite impulse response (FIR)-based predictor weights for Component 2 averaged over conditions and plotted as a function of poststimulus time. C: Mean FIR-based predictor weights for Component 2 averaged over versions and plotted as a function of poststimulus time; [a] = disconfirm > no change; [b] = disconfirm > confirm; [c] = no change > disconfirm. Error bars are standard errors. * = $p < .01$, ** = $p < .001$.

all significant time points. In order to interpret the Response Condition effect, simple contrasts averaging over Version were observed (Fig. 3C), and revealed significantly increased activity for (1) the disconfirm relative to no change response conditions at 21 and 23 s, (2) the disconfirm relative to confirm response conditions at 21 and 23 s, and for (3) the no change relative to disconfirm response conditions at 17 s (all $ps < .005$). This network showed the largest Version, rather than Response Condition, effect with greater activity across the trial for version 1 (in which the images were continuously displayed) than version 2. Due to this sustained activity during version 1 versus the two peaks observed in version 2, as well as to the involvement of primary visual cortex and DMN regions, this functional network was labeled *Visual/Default-Mode Network*.

3.1.3. Component 4: Response Network

Component 4 was characterized by a functional network including activations in bilateral cerebellum and occipital (BAs 18, 19) regions, left-dominant pre- and post-central gyri and supplementary motor area (BAs 3, 4, 6), and left thalamus. This network also included deactivations in VmPFC (BA 10), precuneus (BA 23), and left posterior middle temporal gyrus (BA 21). Component 4 showed significant Poststimulus Time × Version, $F(11,374) = 5.43$, $p < .001$, $\eta_p^2 = 0.14$, and Poststimulus Time × Response Condition, $F(22,748) = 10.19$, $p < .001$, $\eta_p^2 = 0.23$, interactions, but no significant three-way interaction. In order to interpret the Version effect, simple contrasts averaging over Response Condition were observed (Fig. 4B), and revealed significant differences between versions at





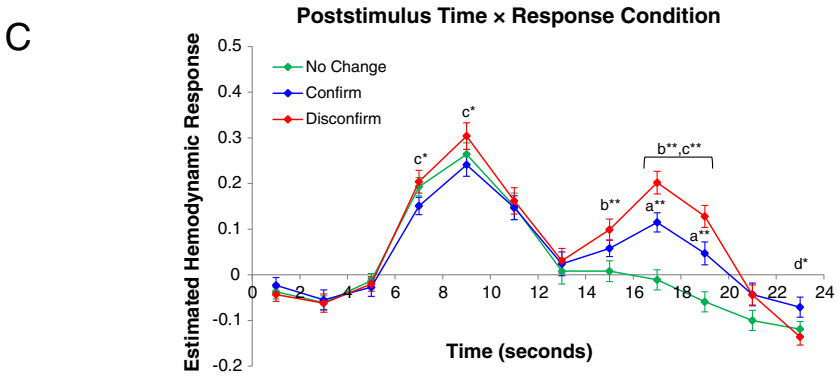


Fig. 4. A–C. A: Dominant 10% of component loadings for Component 4 (Response Network; red/yellow = positive loadings, threshold = 0.15, max = 0.33; blue/green = negative loadings, threshold = −0.23, max = −0.15). Montreal Neurological Institute Z-axis coordinates are displayed. B: Mean finite impulse response (FIR)-based predictor weights for Component 4 averaged over conditions and plotted as a function of poststimulus time. C: Mean FIR-based predictor weights for Component 4 averaged over versions and plotted as a function of poststimulus time; [a] = confirm > no change; [b] = disconfirm > no change; [c] = disconfirm > confirm; [d] = confirm > disconfirm. Error bars are standard errors. * = $p < .01$, ** = $p < .001$.

19 and 21 s ($ps < .005$), due to increased activity in version 2 relative to version 1. In order to interpret the Response Condition effect, simple contrasts averaging over Version were observed (Fig. 4C), and revealed significantly increased activity for (1) the confirm relative to no change response conditions at 17 and 19 s ($ps < .001$), (2) the disconfirm relative to no change response conditions from 15 to 19 s ($ps < .001$), (3) the disconfirm relative to confirm response conditions at 7, 9, 17, and 19 s ($ps < .005$), and for (4) the confirm relative to disconfirm response conditions at 23 s ($p < .005$). This functional network displayed two peaks of activity, corresponding to the time at which ratings were made. Two peaks of activation would be necessary for a response network, since ratings were made for each of the two images displayed. Due to this, and the spatial distribution of the network, it was labeled *Response Network*.

3.1.4. Component 5: Alerting/Salience Network

Component 5 was characterized by a functional network including activations in superior frontal gyrus (BA 8) extending into dACC, right lateral prefrontal cortex extending into IFG (BAs 44, 45), bilateral anterior insula (BA 47) and bilateral occipital cortex (BAs 18, 19). Component 5 showed significant Poststimulus Time × Version, $F(11,374) = 5.30$, $p < .001$, $\eta_p^2 = 0.14$, and Poststimulus Time × Response Condition, $F(22,748) = 6.46$, $p < .001$, $\eta_p^2 = 0.16$, interactions, but no significant three-way interaction. In order to interpret the Version effect, simple

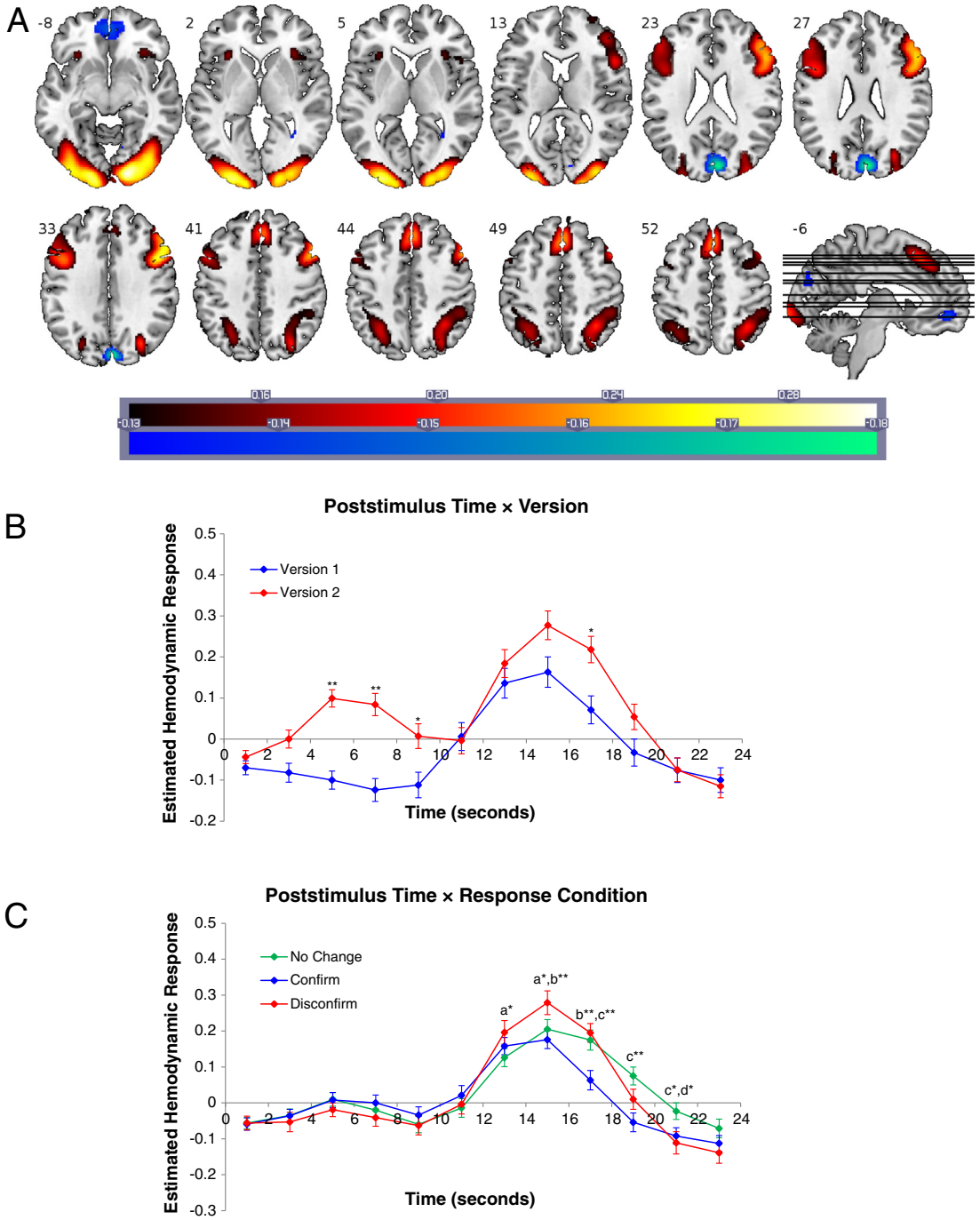


Fig. 5. A–C. A: Dominant 10% of component loadings for Component 5 (Salience Network; red/yellow = positive loadings, threshold = 0.13, max = 0.30; blue/green = negative loadings, threshold = −0.13, min = −0.18). Montreal Neurological Institute Z-axis coordinates are displayed. B: Mean finite impulse response (FIR)-based predictor weights for Component 5 averaged over conditions and plotted as a function of poststimulus time. C: Mean FIR-based predictor weights for Component 5 averaged over versions and plotted as a function of poststimulus time; [a] = disconfirm > no change; [b] = disconfirm > confirm; [c] = no change > confirm; [d] = no change > disconfirm. Error bars are standard errors. * = $p < .01$, ** = $p < .001$.

contrasts averaging over Response Condition were observed (Fig. 5B), and revealed significant differences between versions from 5 to 9 s, and at 17 s ($ps < .01$) due to greater activity in version 2 relative to version 1. In order to interpret the Response Condition effect, simple contrasts averaging over Version were observed (Fig. 5C), and revealed significantly greater activity for (1) the disconfirm relative to no change response conditions at 13 and 15 s ($ps < .005$), (2) the disconfirm relative to confirm response conditions at 15 and 17 s ($ps < .001$), (3) the no change relative to confirm response conditions from 17 to 21 s ($ps < .01$), and for (4) the no change relative to disconfirm response conditions at 21 s ($p < .01$). This functional network peaked briefly at the onset of the second image, when the evidence was first presented, was highest in the disconfirm condition, and was present in both experiments. For this reason, and due to the spatial distribution of the network, it was labeled *Alerting/Salience Network*.

Discussion

In the current study we used multivariate methodology on two datasets in an attempt to link two sequential cognitive stages involved in integrating disconfirmatory evidence to distinct functional brain networks. Three functional networks showed greater intensity

Table 1
Cluster volumes for the most extreme 10% of Component 1 (Integration Network) loadings, with anatomical descriptions, Montreal Neurological Institute (MNI) coordinates, and Brodmann's area (BA) for the peaks within each cluster.

Brain regions	Cluster volume (voxels)	BA for peak locations	MNI coordinate for peak locations		
			x	y	z
Positive loadings					
Cluster 1: bilateral	14,271				
Cerebellum crus I		n/a	40	−76	−24
Lateral occipital cortex, inferior division		19	38	−88	−16
Occipital pole		17/18	26	−100	2
Cerebellum crus I		n/a	−18	−82	−30
Lateral occipital cortex, inferior division		19	−40	−80	−22
Occipital pole		17	−22	−102	0
Occipital pole		18	−34	−94	−14
Cerebellum crus II		n/a	−38	−64	−50
Cerebellum V		n/a	16	−54	−22
Cerebellum VI		n/a	−32	−44	−40
Cerebellum VIIIa		n/a	−30	−40	−42
Cluster 2: bilateral	11,626				
Supramarginal gyrus, posterior division		40	46	−44	58
Lateral occipital cortex, superior division		7	14	−70	64
Frontal orbital cortex		10	40	62	−2
Superior parietal lobule		40	−40	−46	62
Middle frontal gyrus		8	30	14	60
Frontal pole		46	44	52	−12
Frontal pole		45	42	42	28
Superior frontal gyrus		6	−6	−2	78
Postcentral gyrus		2	−56	−26	50
Lateral occipital cortex, superior division		7	−20	−68	60
Supramarginal gyrus, anterior division		2	−54	−30	52
Precentral gyrus		6	−30	−22	72
Superior frontal gyrus		6	30	4	66
Middle frontal gyrus		9/46	42	30	44
Middle frontal gyrus		9	44	28	46
Superior frontal gyrus		8	16	20	66
Postcentral gyrus		3	−32	−36	70
Precentral gyrus		4	−2	−24	82
Cluster 3: left hemisphere	380				
Frontal pole		10	−32	64	2
Frontal orbital cortex		11	−30	64	−8
Frontal orbital cortex		47	−40	50	−14
Cluster 4: right hemisphere	355				
Inferior frontal gyrus/pars opercularis		38	54	20	−2
Inferior frontal gyrus/pars opercularis		6	56	12	10
Cluster 5: left hemisphere	343				
Inferior frontal gyrus/pars opercularis		38	−52	18	−4
Cluster 6: left hemisphere	5				
Middle frontal gyrus		9	−42	24	46

(i.e., increased activations and/or increased deactivations) during integration of disconfirmatory relative to confirmatory evidence for both experiment versions. In order of peak timing (see Figs. 5B, 4B, and 2B, respectively), these reflected (1) an alerting/salience network including dACC and bilateral insulae; (2) a sensorimotor response-related network; and (3) an integration network including bilateral rPFC, OFC, posterior parietal cortex, and IFG. Activity and deactivity associated with visual processing and the DMN separated out from other networks based on HDR shape differences due to stimulus timing differences between the two versions of the experiment.

4.1. Alerting/Salience Network (Component 5)

The dACC (e.g., 2, 28, 48) and bilateral insula (e.g., −30, 24, −4) were the dominant regions of the alerting/salience network (Component 5), which became active during the onset of the second image, when the confirmatory/disconfirmatory evidence was presented. This network can be described as the well-documented salience network, which is involved in attending to environmentally-salient stimuli and has been hypothesized to be responsible for switching between large-scale brain networks to allow access to relevant cognitive and sensory systems (Goulden et al., 2014; Menon and Uddin, 2010). Relating the current version of the salience network to the 7-network brain parcellation derived from resting state data (Buckner et al., 2011; Choi et al., 2012; Yeo et al., 2011), the dACC (e.g., 2, 28, 48), prefrontal (e.g., −46, 10, 32), caudate (subthreshold) (e.g., 16, 14, 10), and parietal activations (e.g., −46, −40, 48) were all located on the frontoparietal network, and the occipital activations (e.g., 18, −92, −8) on the visual network. All deactivations were located on the DMN. The dACC and insula have been implicated in the "moment of recognition" of an object during evidence accumulation (Liu and Pleskac, 2011; Ploran et al., 2007); the dACC specifically is involved in surprise and error detection, and has been suggested to play a role in the "aha! moment", or to alert when behavioral adjustment is required (Cameron and Vincent van, 2007; Egner, 2011; Walsh et al., 2011; Whitman et al., 2013; Woodward et al., 2008). The right lateral prefrontal cortex, also involved in this network in the current study (e.g., 48, 12, 34), has been shown to activate in response to prediction error during associative learning (Corlett et al., 2004; Fletcher et al., 2001; Turner et al., 2004), and has been identified as important to hypothesis evaluation in clinical settings (Coltheart, 2010). The significantly higher HDR peak in the disconfirm (relative to confirm) condition may be interpreted as salience network activation due to the conflict between the initial belief (formed during the presentation of image 1) and the disconfirmatory evidence presented at image 2. Detection of conflict between a held belief and presented evidence is a crucial first step in the process of belief revision, and

Table 2
Cluster volumes for the most extreme 10% of Component 2 (Visual/Default-Mode Network) loadings, with anatomical descriptions, Montreal Neurological Institute (MNI) coordinates, and Brodmann's area (BA) for the peaks within each cluster.

Brain regions	Cluster volume (voxels)	BA for peak locations	MNI coordinate for peak locations		
			x	y	z
Positive loadings					
Cluster 1: bilateral	19,622				
Occipital fusiform gyrus		18	26	−78	−14
Occipital pole		17	16	−100	14
Lingual gyrus		17	2	−86	−10
Occipital pole		17	−12	−100	2
Occipital fusiform gyrus		18	−22	−80	−16
Occipital pole		18	24	−94	12
Lateral occipital cortex, superior division		7	28	−64	48
Lateral occipital cortex, superior division		19	28	−72	32
Superior parietal lobule		7	−26	−62	46
Cluster 2: right hemisphere	208				
Middle frontal gyrus		44	48	12	34
Cluster 3: left hemisphere	61				
Middle frontal gyrus		44	−44	6	34
Cluster 4: right hemisphere	39				
Frontal pole/middle frontal gyrus		45	48	36	32
Negative loadings					
Cluster 1: bilateral	2479				
Ventromedial prefrontal cortex		10	−4	52	−2
Ventromedial prefrontal cortex		10	0	62	−2
Cluster 2: left hemisphere	1036				
Lateral occipital cortex, superior division		39	−50	−74	26
Lateral occipital cortex, superior division		37	−60	−64	14
Cluster 3: right hemisphere	924				
Parietal operculum cortex		48	54	−30	22
Central operculum cortex		48	58	−2	6
Cluster 4: bilateral	851				
Precuneus cortex		23	−2	−62	26
Cluster 5: left hemisphere	516				
Parietal operculum cortex		42	−60	−32	20
Supramarginal gyrus, posterior division		40	−64	−44	36
Cluster 6: left hemisphere	428				
Middle temporal gyrus, anterior division		21	−54	−6	−16
Cluster 7: right hemisphere	268				
Middle temporal gyrus, anterior division		21	56	−6	−20
Cluster 8: left hemisphere	237				
Superior frontal gyrus		9	−24	32	46
Cluster 9: bilateral	189				
Cingulate gyrus, posterior division		23	2	−24	46
Cluster 10: right hemisphere	91				
Postcentral gyrus		3	28	−38	66
Cluster 11: right hemisphere	16				
Angular gyrus		39	58	−66	18

dysfunction in the detection of this conflict could contribute to resistance in modifying outdated beliefs.

4.2. Integration Network (Component 1)

Like the salience network, the integration network (Component 1; bilateral rPFC/OFC, posterior parietal cortex, and IFG) distinguished between confirmatory and disconfirmatory evidence, and peaked during confirmatory/disconfirmatory evidence presentation; however, its peak was noticeably later than that of the salience network (19 vs. 15 s), suggesting sequential activation. Relating the integration network to the 7-network brain parcellation derived from resting state data (Buckner et al., 2011; Choi et al., 2012; Yeo et al., 2011), the rostral and superior prefrontal (e.g., 2, 28, 48), caudate (subthreshold) (e.g., 20, 17, 14), and cerebellar activations (e.g., −38, −64, −50) were located on the frontoparietal network, inferior frontal gyrus/pars opercularis (e.g., −38, −64, −50) and putamen (subthreshold) (e.g., 25, 4, 6) on the ventral attention network, and the occipital activations on the visual network. Parietal activations (e.g., 46, −44, 58) were located on the dorsal attention network. The rPFC is involved in the evaluation of self-generated information (Christoff et al., 2003), and has been proposed as a key region involved in the balance between self-generated and externally-generated information (Burgess et al., 2005; Gilbert et al., 2006). Together, the IFG and posterior parietal regions (e.g., supramarginal and angular gyri), are involved in semantic processing and visual word recognition (Binder et al., 2009), and are recruited during perceptual decision-making tasks, such as in the current study. The IFG has also been implicated in belief formation and updating (d'Acremont et al., 2013; Sharot et al., 2011), and there is evidence that disruption of the left IFG improves integration of unfavorable evidence (Sharot et al., 2012), suggesting that it may play a key role in disconfirmatory evidence integration in particular.

Much like the salience network, functional brain activity in the integration network was highest during disconfirmatory evidence integration. Taken together with its delayed peak relative to the salience network, this suggests a role in evaluating the presented evidence relative to the initial belief (formed at image 1). This would also be necessary during confirmatory evidence integration, but might be expected to elicit lesser and less sustained activity than during disconfirmatory evidence integration, as was evident in the current findings (see Fig. 2C). Evaluating presented evidence and comparing it to prior knowledge is another crucial aspect of evidence integration, and dysfunction in this network could also contribute to resistance in modifying beliefs.

Table 3
Cluster volumes for the most extreme 10% of Component 4 (Response Network) loadings, with anatomical descriptions, Montreal Neurological Institute (MNI) coordinates, and Brodmann's area (BA) for the peaks within each cluster.

Brain regions	Cluster volume (voxels)	BA for peak locations	MNI coordinate for peak locations		
			x	y	z
Positive loadings					
Cluster 1: bilateral	20,765				
Precentral gyrus		4	−40	−20	58
Postcentral gyrus		3	−54	−22	50
Lateral occipital cortex, superior division		7	−26	−60	50
Lateral occipital cortex, inferior division		18	−30	−88	6
Lateral occipital cortex, inferior division		19	−32	−88	0
Occipital fusiform gyrus		18	22	−86	−12
Cerebellum VI		n/a	32	−56	−22
Temporal occipital fusiform cortex		19	−40	−64	−16
Occipital fusiform gyrus		18	−20	−90	−12
Temporal occipital fusiform cortex		37	−38	−54	−20
Lateral occipital cortex, inferior division		18	32	−88	6
Lateral occipital cortex, superior division		19	−26	−74	28
Precentral gyrus		6	−52	4	38
Lateral occipital cortex, superior division		19	28	−74	28
Lateral occipital cortex, superior division		7	24	−64	52
Cerebellum VI		18	10	−74	−22
Cluster 2: bilateral	1428				
Supplementary motor area		6	−4	8	54
Cluster 3: right hemisphere	491				
Middle frontal gyrus		6	28	−2	54
Cluster 4: right hemisphere	222				
Precentral gyrus		44	52	8	32
Cluster 5: left hemisphere	78				
Thalamus		n/a	−10	−18	8
Cluster 6: left hemisphere	46				
Central opercular cortex		48	−48	−22	20
Cluster 7: right hemisphere	12				
Supramarginal gyrus, anterior division		2	46	−32	44
Negative loadings					
Cluster 1: bilateral	1869				
Frontal pole		10	−4	60	22
Cluster 2: left hemisphere	1356				
Lateral occipital cortex, superior division		39	−50	−74	34
Cluster 3: bilateral	201				
Cuneal cortex		18	4	−86	24
Cluster 4: left hemisphere	167				
Middle temporal gyrus, posterior division		21	−60	−12	−16
Cluster 5: right hemisphere	145				
Lateral occipital cortex, superior division		39	54	−70	32
Cluster 6: left hemisphere	91				
Precuneus cortex		23	−8	−52	34
Cluster 7: left hemisphere	64				
Middle temporal gyrus, temporooccipital part		37	−64	−56	−4
Cluster 8: left hemisphere	47				
Frontal pole		9	−16	42	52

4.3. Visual/Default-Mode Network (Component 2)

The visual/default-mode network showed sustained activity during version 1 and two peaks in version 2, and was characterized by deactivations in DMN regions, and activations in visual cortex regions. Relating the visual/default-mode network to the 7-network brain parcellation derived from resting state data (Buckner et al., 2011; Choi et al., 2012; Yeo et al., 2011), the occipital activations were all located on the visual network. The parietal (e.g., −26, −62, 46) and lateral prefrontal (e.g., 48, 36, 32) activations were located on the dorsal attention network. All deactivations were located on the DMN, but the superior temporal deactivations (which included primary auditory cortices, e.g., −58, −32, 16) were located on the somatosensory network. The auditory cortex deactivations present on this component have been shown to be sensitive to load-dependent task-related decreases in activity in working memory and source experiments that employing visual encoding (Metzak et al., 2011, 2012, submitted; Woodward et al., 2013). This coordinated decrease in bilateral primary auditory cortex activity could relate to reduced activation during inner speech (Buchsbaum et al., 2005; Frith et al., 1991), or a more general phenomenon whereby task-irrelevant primary sensory cortices (with visual cortices being task-relevant) are deactivated during task performance (Laurienti et al., 2002; Shulman et al., 1997). The fact that the bilateral primary auditory cortex deactivity emerged uniquely on the visual processing component provides support for the latter interpretation.

4.4. Response Network (Component 4)

Relating the response network to the 7-network brain parcellation derived from resting state data (Buckner et al., 2011; Choi et al., 2012; Yeo et al., 2011), the supplementary motor area/dACC (e.g., −4, 8, 54) and cerebellar (e.g., 32, −56, −22) peak activations were located on the ventral attention network, the frontal (e.g., 28, −2, 54) and parietal (e.g., −26, −60, 50) activations on the dorsal attention network, the left sensorimotor activation (e.g., −40, −20, 58) on the somatosensory network, and the occipital activations on the visual network. All deactivations were located on the DMN. The sensorimotor response network identified in the current study peaked during responses made to both the first and second images, and showed significantly greater activity

Table 4
Cluster volumes for the most extreme 10% of Component 5 (Salience Network) loadings, with anatomical descriptions, Montreal Neurological Institute (MNI) coordinates, and Brodmann's area (BA) for the peaks within each cluster.

Brain regions	Cluster volume (voxels)	BA for peak locations	MNI coordinate for peak locations		
			x	y	z
Positive loadings					
Cluster 1: bilateral	16,843				
Occipital pole		18	18	−92	−8
Occipital fusiform gyrus		19	34	−78	−14
Occipital pole		18	−28	−96	4
Occipital fusiform gyrus		18	−20	−90	−12
Occipital fusiform gyrus		19	−36	−78	−14
Occipital pole		18	32	−90	8
Temporal occipital fusiform cortex		37	40	−60	−18
Occipital pole		18	−32	−92	−6
Lateral occipital cortex, superior division		19	30	−72	32
Superior parietal lobule		7	34	−56	48
Lateral occipital cortex, superior division		7	−30	−62	44
Supramarginal gyrus, posterior division		40	48	−42	46
Lateral occipital cortex, superior division		19	−28	−74	30
Supramarginal gyrus, posterior division		40	−46	−40	48
Cluster 2: right hemisphere	4694				
Middle frontal gyrus		44	48	12	34
Middle frontal gyrus		6	38	8	62
Insular cortex		47	32	26	−2
Cluster 3: left hemisphere	2329				
Middle frontal gyrus		44	−46	10	32
Middle frontal gyrus		45	−50	32	24
Middle frontal gyrus		6	−42	6	60
Cluster 4: bilateral	1676				
Superior frontal gyrus/Dorsal anterior cingulate cortex		8	2	28	48
Cluster 5: left hemisphere	160				
Insular cortex		47	−30	24	−4
Cluster 6: right hemisphere	75				
Frontal pole		10	30	58	10
Cluster 7: left hemisphere	31				
Frontal pole		46	−44	48	−2
Cluster 8: left hemisphere	1				
Cerebellum crus II		n/a	−10	−76	−32
Negative loadings					
Cluster 1: bilateral	684				
Cuneal cortex		18	6	−84	26
Cluster 2: bilateral	425				
Frontal medial cortex		11	−2	50	−8
Cluster 3: right hemisphere	36				
Lingual gyrus		18	10	−66	−4
Cluster 4: right hemisphere	28				
Lingual gyrus		37	32	−52	2

for disconfirmatory evidence integration during the response to the second image. These differences were likely the result of the way in which the conditions were encoded, given that the disconfirm responses included greater rating changes between image 1 and 2 (e.g., rating change from 14 to 4) than the confirm condition (e.g., rating change from 9 to 15). Consistent with this interpretation, the no change condition (which included rating changes of two steps or fewer) showed the least activity after the onset of the second image (see Fig. 4C).

It should be noted that there was a small, but significant difference between the confirm and disconfirm conditions during the response to image 1. Given that one of the parameters for the disconfirm condition was that image 2 be rated closer to the middle of the scale than image 1 (whereas the opposite was true for confirm), this unexpected finding likely reflects initial ratings closer to the extremes of the scale in the disconfirm condition. This would result in greater response-related activity in the disconfirm relative to confirm condition during image 1, since the initial point from which participants made their ratings was at the middle of the scale.

Although activity in Component 1 (integration) and 4 (responding) began increasing simultaneously after evidence presentation, the integration network peaked later than the response network (19 vs 17 s post-stimulus), and activation extended past the time at which the response network returned to baseline. This is explained by the fact that participants were able to modify their responses throughout the 6 s time window during which the response options were displayed. This earlier peak for responding versus integrating may be because participants began responding based on the aha! moment elicited at the onset of the second image (which would guide their decision to begin either down-rating or up-rating their initial responses), and then continued to evaluate their decision during and following finalization of their responses throughout the 6 s time window. This HDR pattern (viz., the response network peaking earlier than cognitive networks) was also observed in a previous study from our lab on controlled semantic association (Woodward et al., in press).

4.5. Differences between experiment versions

In addition to differences between the response conditions, all functional networks also demonstrated distinct activation patterns across the two versions of the experiment, meaning that all components showed spatial but not temporal replication, providing an opportunity to use the differences between experiments to help interpret the cognitive function of the networks. For example, for Component 2 (visual/default mode network) there was a higher, sustained peak in version 1 relative to version 2, which corresponded to the sustained presentation of the images. In addition, on Components 1 (integration network)

and 5 (salience network), version 2 showed an early peak not present in version 1. Although these networks showed higher activity during disconfirmatory evidence integration, the time points on which a version effect was observed (5 to 11 s) were not the same as those on which a response conditions effect was observed (15 to 23 s). Given that both of these networks included brain regions related to visual processing, it is possible that these differences between versions were also driven by the visual stimuli, as with Component 2. Finally, some late trial differences between Versions 1 and 2 were present on Component 1 (21 s), Component 4 (19 to 21 s), and Component 5 (17 s), due to a higher or more sustained peak in version 2 relative to version 1 for these components. These higher, prolonged peaks on the integration (Component 1) and salience (Component 5) networks could be the result of the increased morphing ratios and improved experiment design (jittered ITIs), which presumably served to increase the magnitude of the experimental effects in Version 2. In the case of the response network (Component 4), the difference between versions is likely due to the single animal name in version 2 (compared to both names in version 1), which would lead to larger and more extreme response changes after evidence presentation. Importantly, these version effects were statistically independent from the response condition effects, due to the absence of both three-way interactions and Version × Response Condition interactions.

Inclusion of the no change (rating changes less than 3) response condition in the current study was intended as a control; however, there was evidence of *increased* activity during evidence integration in the no change condition relative to the other response conditions within two functional networks: the visual/DMN network (Component 2) and the salience network (Component 5). While an unexpected finding, this might be a consequence of uncertainty about the nature of the second image on the part of participants. If participants were unable to identify the animal in the second image, they would likely have had difficulty determining whether they should respond in a clearly confirmatory or disconfirmatory manner, and might be less inclined to change their initial ratings. The increased attention associated with uncertainty could account for the heightened activity in functional networks associated with visual and cognitive attention.

4.6. Limitations

One limitation of this study was that this comparison was carried out between subjects. Ideally, one would conduct multi-experiment comparisons within-subjects (e.g., de Zubicaray et al., 2013; Metzak et al., 2013); however, fMRI data is expensive to collect, and testing time is limited, such that comparing experiment versions often must be carried out between subjects. Combining versions of an experiment with a number of differences in the experimental design (e.g., persistence of the visual stimuli, differences in response interface, etc.) facilitates powerful comparisons for interpreting network functions, and it is the analysis of differences and commonalities between HDR shapes across experiment versions that is of scientific interest. In the current analysis, differences between experimental versions did not appear to substantially impact interpretation of the response conditions, evidenced by the absence of three-way interactions. For example, only minor variations between experiment versions were present on the response component, despite substantial differences in the nature of the response interface. In addition, there was an early peak present on the salience and integration networks in version 2 that was not present in version 1, which may have been driven by visual processing regions that were part of each network, or may have been due to the increased morphing ratios and other improvements implemented in version 2. In either case, these early differences between versions did not affect interpretation of the differences between response conditions, which occurred later in the trial. The jittered ITIs in version 2, designed to increase power of the manipulations, did appear to impact the magnitude of the HDR shapes for alerting and integration of disconfirmatory evidence, which were greater in version 2. However, the conclusions reached for this study would ideally be tested by conducting multi-experiment comparisons within-subjects.

4.7. Conclusion

In the current study, we examined the functional networks associated with the processing of disconfirmatory relative to confirmatory evidence. By combining data from two versions of the same experiment that differed primarily in terms of stimulus timing, we were able to distinguish between functional brain activity associated with detection and integration of evidence, as well as others associated with responding and visual processing. We identified three functional networks that showed increased activity during disconfirmatory relative to confirmatory evidence integration: a salience network involved in detecting a mismatch between the presented evidence and the initially-formed belief, a response network, and an integration network involved in the evaluation of the evidence and in comparison of that evidence to the initial belief.

These findings highlight two distinct functional networks underlying disconfirmatory evidence integration that correspond to two important cognitive processes underlying belief revision: (1) detection of a conflict between an initial belief and a piece of (disconfirmatory) evidence; and (2) evaluation of that evidence in light of the initial belief in order to determine whether it should be integrated into the current belief system and the belief modified or dropped. In cases where an individual is consistently resistant to disconfirmatory evidence (e.g., groupthink, stereotyping), or in clinical settings (psychotic delusions), one or both of these mechanisms may play a role. For example, delusional schizophrenia patients who demonstrate a bias against disconfirmatory evidence (Sanford et al., 2014; Speechley et al., 2011; Woodward et al., 2006) may show decreased activity in both the salience and integration networks when faced with disconfirmatory evidence. Future research is necessary to determine whether resistance to modifying beliefs when faced with disconfirmatory evidence is due to a lack of attention towards/detection of that evidence, to an inability to integrate the evidence into the current belief system, or some combination of both processes.

Acknowledgments

This project was supported by a Scholar Award from the Michael Smith Foundation for Health Research (MSFHR; CI-SCH-00073), a New Investigator Award from the Canadian Institutes of Health Research (CIHR; MMS8770) to TSW, as well as CIHR Doctoral Research Awards for KML (DSZ-128637) and PDM (DPO-128616). Operating costs were supported by a grant to TSW from the British Columbia Schizophrenia Society (BCSS; formerly Mind Foundation of BC). The authors acknowledge the UBC High Field Magnetic Resonance Imaging Centre, and thank John Paiement for assistance with computer programming, Teresa Dahm for assistance with stimulus preparation, and Nicole Sanford, Jennifer Riley, and Jennifer Whitman for contributions to data processing.

References

Annett, M., 1970. A classification of hand preference by association analysis. Br. J. Psychol. 61 (3), 303–321.

Behrens, T.E.J., Woolrich, M.W., Walton, M.E., Rushworth, M.F.S., 2007. Learning the value of information in an uncertain world. Nat. Neurosci. 10 (9), 1214–1221. http://dx.doi.org/10.1038/nn1954.

Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb. Cortex 19 (12), 2767–2796. http://dx.doi.org/10.1093/cercor/bhp055.

Buchsbaum, B.R., Olsen, R.K., Koch, P.F., Kohn, P., Kippenhan, J.S., Berman, K.F., 2005. Reading, hearing, and the planum temporale. NeuroImage 24 (2), 444–454. http://dx.doi.org/10.1016/j.neuroimage.2004.08.025.

Buckner, R.L., Andrews-Hanna, J.R., Schacter, D.L., 2008. The brain's default network. Ann. N. Y. Acad. Sci. 1124, 1–38. http://dx.doi.org/10.1196/annals.1440.011.

Buckner, R.L., Krienen, F.M., Castellanos, A., Diaz, J.C., Yeo, B.T., 2011. The organization of the human cerebellum estimated by intrinsic functional connectivity. J. Neurophysiol. 106 (5), 2322–2345. http://dx.doi.org/10.1152/jn.00339.2011.

Burgess, P.W., Simons, J.S., Dumontheil, I., Gilbert, S.J., 2005. The gateway hypothesis of rostral prefrontal cortex (area 10) function. In: Duncan, J., McLeod, P., Phillips, L. (Eds.), Measuring the Mind: Speed, Control, and Age. Oxford University Press, Oxford, pp. 215–246.

Cameron, S.C., Vincent van, V., 2007. Anterior cingulate cortex and conflict detection: an update of theory and data. Cogn. Affect. Behav. Neurosci. 7 (4), 367–379.

Cattell, R.B., 1966. The scree test for the number of factors. Multivar. Behav. Res. 1 (2), 245–276 (doi: citeulike-article-id:3574985).

Cattell, R.B., Vogelmann, S., 1977. A comprehensive trial of the scree and kg criteria for determining the number of factors. Multivar. Behav. Res. 12 (3), 289–325.

Choi, E.Y., Yeo, B.T., Buckner, R.L., 2012. The organization of the human striatum estimated by intrinsic functional connectivity. J. Neurophysiol. 108 (8), 2242–2263. http://dx.doi.org/10.1152/jn.00270.2012.

Christoff, K., Ream, J.M., Geddes, L.P.T., Gabrieli, J.D.E., 2003. Evaluating self-generated information: anterior prefrontal contributions to human cognition. Behav. Neurosci. 117 (6), 1161–1168.

Coltheart, M., 2010. The neuropsychology of delusions. Ann. N. Y. Acad. Sci. 1191, 16–26.

Corlett, P.R., Aitken, M.R., Dickinson, A., Shanks, D.R., Honey, G.D., Honey, R.A., Fletcher, P.C., 2004. Prediction error during retrospective revaluation of causal associations in humans: fMRI evidence in favor of an associative model of learning. Neuron 44 (5), 877–888. http://dx.doi.org/10.1016/j.neuron.2004.11.022.

d'Acremont, M., Schultz, W., Bossaerts, P., 2013. The human brain encodes event frequencies while forming subjective beliefs. J. Neurosci. 33 (26), 10887–10897.

de Zubicaray, G.I., Hansen, S., McMahon, K.L., 2013. Differential processing of thematic and categorical conceptual relations in spoken word production. J. Exp. Psychol. Gen. 142 (1), 131–142. http://dx.doi.org/10.1037/a0028717.

Egner, T., 2011. Surprise! A unifying model of dorsal anterior cingulate function? Nat. Neurosci. 14 (10), 1219–1220. http://dx.doi.org/10.1038/nn.2932.

Flashman, L.A., McAllister, T.W., 2002. Lack of awareness and its impact in traumatic brain injury. NeuroRehabilitation 17 (4), 285–296.

Fletcher, P.C., Anderson, J.M., Shanks, D.R., Honey, R., Carpenter, T.A., Donovan, T., Bullmore, E.T., 2001. Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. Nat. Neurosci. 4 (10), 1043–1048. http://dx.doi.org/10.1038/nn733.

Frith, C.D., Friston, K.J., Liddle, P.F., Frackowiak, R.S., 1991. A PET study of word finding. Neuropsychologia 29 (12), 1137–1148.

Gilbert, S.J., Spengler, S., Simons, J.S., Steele, J.D., Lawrie, S.M., Frith, C.D., Burgess, P.W., 2006. Functional specialization within rostral prefrontal cortex (area 10): a meta-analysis. J. Cogn. Neurosci. 18 (6), 932–948.

Goulden, N., Khusnulina, A., Davis, N.J., Bracewell, R.M., Bokde, A.L., McNulty, J.P., Mullins, P.G., 2014. The salience network is responsible for switching between the default mode network and the central executive network: replication from DCM. NeuroImage 99, 180–190. http://dx.doi.org/10.1016/j.neuroimage.2014.05.052.

Hunter, M.A., Takane, Y., 2002. Constrained principal component analysis: various applications. J. Educ. Behav. Stat. 27 (2), 105–145.

Laurienti, P.J., Burdette, J.H., Wallace, M.T., Yen, Y.F., Field, A.S., Stein, B.E., 2002. Deactivation of sensory-specific cortex by cross-modal stimuli. J. Cogn. Neurosci. 14 (3), 420–429. http://dx.doi.org/10.1162/089892902317361930.

Lavigne, K.M., Rapin, L.A., Metzak, P.D., Whitman, J.C., Jung, K., Dohen, M., Woodward, T.S., 2014. Left-dominant temporal–frontal hypercoupling in schizophrenia patients with hallucinations during speech perception. Schizophr. Bull. http://dx.doi.org/10.1093/schbul/sbu004.

Liu, T., Pleskac, T.J., 2011. Neural correlates of evidence accumulation in a perceptual decision task. J. Neurophysiol. 106 (5), 2383–2398. http://dx.doi.org/10.1152/jn.00413.2011.

Marsh, R., Horga, G., Parashar, N., Wang, Z., Peterson, B.S., Simpson, H.B., 2014. Altered activation in fronto-striatal circuits during sequential processing of conflict in unmedicated adults with obsessive–compulsive disorder. Biol. Psychiatry 75 (8), 615–622. http://dx.doi.org/10.1016/j.biopsych.2013.02.004.

Menon, V., Uddin, L.Q., 2010. Saliency, switching, attention and control: a network model of insula function. Brain Struct. Funct. 214 (5–6), 655–667. http://dx.doi.org/10.1007/s00429-010-0262-0.

Metzak, P.D., Feredoes, E., Takane, Y., Wang, L., Weinstein, S., Cairo, T., Woodward, T.S., 2011. Constrained principal component analysis reveals functionally connected load-dependent networks involved in multiple stages of working memory. Hum. Brain Mapp. 32 (6), 856–871. http://dx.doi.org/10.1002/hbm.21072.

Metzak, P.D., Riley, J.D., Wang, L., Whitman, J.C., Ngan, E.T.C., Woodward, T.S., 2012. Decreased efficiency of task-positive and task-negative networks during working memory in schizophrenia. Schizophr. Bull. 38 (4), 803–813.

Metzak, P.D., Meier, B., Graf, P., Woodward, T.S., 2013. More than a surprise: the bivalency effect in task switching. J. Cogn. Psychol. 25 (7), 833–842. http://dx.doi.org/10.1080/20445911.2013.832196.

Metzak, P.D., Lavigne, K.M., Woodward, T.S., 2015. Functional brain networks involved in reality monitoring (submitted).

Ploran, E.J., Nelson, S.M., Velanova, K., Donaldson, D.I., Petersen, S.E., Wheeler, M.E., 2007. Evidence accumulation and the moment of recognition: dissociating perceptual recognition processes using fMRI. J. Neurosci. 27 (44), 11912–11924. http://dx.doi.org/10.1523/jneurosci.3522-07.2007.

Raichle, M.E., MacLeod, A.M., 2001. A default mode of brain function. Proc. Natl. Acad. Sci. U. S. A. 98 (2), 676–682.

Sanford, N., Veckenstedt, R., Moritz, S., Balzan, R., Woodward, T.S., 2014. Impaired integration of disambiguating evidence in delusional schizophrenia patients. Psychol. Med. 44 (13), 2729–2738. http://dx.doi.org/10.1017/S0033291714000397.

Serences, J.T., 2004. A comparison of methods for characterizing the event-related BOLD time series in rapid fMRI. NeuroImage 21 (4), 1690–1700. http://dx.doi.org/10.1016/j.neuroimage.2003.12.021.

Sharot, T., Korn, C.W., Dolan, R.J., 2011. How unrealistic optimism is maintained in the face of reality. Nat. Neurosci. 14 (11), 1475–1479.

Sharot, T., Kanai, R., Marston, D., Korn, C.W., Rees, G., Dolan, R.J., 2012. Selectively altering belief formation in the human brain. Proc. Natl. Acad. Sci. U. S. A. 109 (42), 17058–17062. http://dx.doi.org/10.1073/pnas.1205828109.

Shulman, G.L., Corbetta, M., Buckner, R.L., Raichle, M.E., Fiez, J.A., Miezin, F.M., Petersen, S.E., 1997. Top-down modulation of early sensory cortex. Cereb. Cortex 7 (3), 193–206.

Speechley, W.J., Moritz, S., Ngan, E.T.C., Woodward, T.S., 2011. Impaired evidence integration and delusions in schizophrenia. J. Exp. Psychopathol. 3 (4), 688–701.

Takane, Y., Hunter, M.A., 2001. Constrained principal component analysis: a comprehensive theory. AAECC 12, 391–419.

Takane, Y., Shibayama, T., 1991. Principal component analysis with external information on both subjects and variables. Psychometrika 56 (1), 97–120. http://dx.doi.org/10.1007/bf02294589.

Turner, M.E., Pratkanis, A.R., 1998. Twenty-five years of groupthink theory and research: lessons from the evaluation of a theory. Organ. Behav. Hum. Decis. Process. 73 (2/3), 105–115.

Turner, D.C., Aitken, M.R., Shanks, D.R., Sahakian, B.J., Robbins, T.W., Schwarzbauer, C., Fletcher, P.C., 2004. The role of the lateral frontal cortex in causal associative learning: exploring preventative and super-learning. Cereb. Cortex 14 (8), 872–880. http://dx.doi.org/10.1093/cercor/bhh046.

Walsh, B.J., Buonocore, M.H., Carter, C.S., Mangun, G.R., 2011. Integrating conflict detection and attentional control mechanisms. J. Cogn. Neurosci. 23 (9), 2191–2201.

Whitman, J.C., Metzak, P.D., Lavigne, K.M., Woodward, T.S., 2013. Functional connectivity in a frontoparietal network involving the dorsal anterior cingulate cortex underlies decisions to accept a hypothesis. Neuropsychologia 51 (6), 1132–1141. http://dx.doi.org/10.1016/j.neuropsychologia.2013.02.016.

Woodward, T.S., Moritz, S., Cuttler, C., Whitman, J.C., 2006. The contribution of a cognitive bias against disconfirmatory evidence (BADE) to delusions in schizophrenia. J. Clin. Exp. Neuropsychol. 28, 605–617.

Woodward, T.S., Metzak, P.D., Meier, B., Holroyd, C.B., 2008. Anterior cingulate cortex signals the requirement to break inertia when switching tasks: a study of the bivalency effect. NeuroImage 40 (3), 1311–1318. http://dx.doi.org/10.1016/j.neuroimage.2007.12.049.

Woodward, T.S., Feredoes, E., Metzak, P.D., Takane, Y., Manoach, D.S., 2013. Epoch-specific functional networks involved in working memory. NeuroImage 65, 529–539.

Woodward, T.S., Tipper, C., Leung, A., Lavigne, K.M., Metzak, P.D., 2015. Reduced functional connectivity during controlled semantic integration in schizophrenia: a multivariate approach. Hum. Brain Mapp. (in press).

Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Buckner, R.L., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. 106 (3), 1125–1165. http://dx.doi.org/10.1152/jn.00338.2011.