Causal Inference in Audiovisual Perception

Agoston Mihalik^{1,2} and Uta Noppeney^{1,3}

¹Computational Neuroscience and Cognitive Robotics Centre, University of Birmingham, Birmingham B15 2TT, United Kingdom, ²Centre for Medical Image Computing, Department of Computer Science, University College London, London WC1V 6LJ, United Kingdom, and ³Donders Institute for Brain, Cognition and Behaviour, Radboud University, 6525, Nijmegen, The Netherlands

In our natural environment the senses are continuously flooded with a myriad of signals. To form a coherent representation of the world, the brain needs to integrate sensory signals arising from a common cause and segregate signals coming from separate causes. An unresolved question is how the brain solves this binding or causal inference problem and determines the causal structure of the sensory signals. In this functional magnetic resonance imaging (fMRI) study human observers (female and male) were presented with synchronous auditory and visual signals at the same location (i.e., common cause) or different locations (i.e., separate causes). On each trial, observers decided whether signals come from common or separate sources (i.e., “causal decisions”). To dissociate participants’ causal inference from the spatial correspondence cues we adjusted the audiovisual disparity of the signals individually for each participant to threshold accuracy. Multivariate fMRI pattern analysis revealed the lateral prefrontal cortex as the only region that encodes predominantly the outcome of observers’ causal inference (i.e., common vs separate causes). By contrast, the frontal eye field (FEF) and the intraparietal sulcus (IPS0–4) form a circuitry that concurrently encodes spatial (auditory and visual stimulus locations), decisional (causal inference), and motor response dimensions. These results suggest that the lateral prefrontal cortex plays a key role in inferring and making explicit decisions about the causal structure that generates sensory signals in our environment. By contrast, informed by observers’ inferred causal structure, the FEF–IPS circuitry integrates auditory and visual spatial signals into representations that guide motor responses.

Key words: audiovisual; causal inference; fMRI; multisensory; multivariate; prefrontal cortex

Significance Statement

In our natural environment, our senses are continuously flooded with a myriad of signals. Transforming this barrage of sensory signals into a coherent percept of the world relies inherently on solving the causal inference problem, deciding whether sensory signals arise from a common cause and should hence be integrated or else be segregated. This functional magnetic resonance imaging study shows that the lateral prefrontal cortex plays a key role in inferring the causal structure of the environment. Crucially, informed by the spatial correspondence cues and the inferred causal structure the frontal eye field and the intraparietal sulcus form a circuitry that integrates auditory and visual spatial signals into representations that guide motor responses.

Introduction

In our natural environment our senses are continuously flooded with myriads of signals. To form a coherent representation of the world, the brain needs to integrate sensory signals arising from a common cause and segregate signals coming from different causes (Noppeney, 2020). Multisensory perception thus implicitly relies on solving the so-called causal inference or binding

problem (i.e., deciding whether or not signals originate from a common cause based on spatiotemporal or higher order correspondence cues; Munhall et al., 1996; Welch, 1999; Slutsky and Recanzone, 2001; Lewald and Gusk, 2003; Wallace et al., 2004b; Noesselt et al., 2007; van Wassenhove et al., 2007; Recanzone, 2009; Lee and Noppeney, 2011a; Parise and Ernst, 2016).

Accumulating evidence suggests that human observers arbitrate between sensory integration and segregation in perception consistent with Bayesian causal inference (Körding et al., 2007; Shams and Beierholm, 2010; Rohe and Noppeney, 2015a; Acerbi et al., 2018). Most notably, observers integrate synchronous audiovisual (AV) signals when they are presented with a small spatial disparity but segregate them at large spatial disparities. As a result, they perceive the sound location biased or shifted toward the visual signal location and vice versa, depending on the relative auditory and visual reliabilities (Bertelson and Radeau, 1981;

Received Jan. 7, 2020; revised June 26, 2020; accepted July 1, 2020.

Author contributions: A.M. and U.N. designed research; A.M. performed research; A.M. and U.N. analyzed data; A.M. and U.N. wrote the paper.

This work was supported by the European Research Council (Grant ERC-2012-StG_20111109 multisens).

The authors declare no competing financial interests.

Correspondence should be addressed to Agoston Mihalik at axm676@alumni.bham.ac.uk.

<https://doi.org/10.1523/JNEUROSCI.0051-20.2020>

Copyright © 2020 the authors

Driver, 1996; Ernst and Banks, 2002; Alais and Burr, 2004; Bonath et al., 2007; Meijer et al., 2019). Crucially, these cross-modal biases taper off at large spatial disparities when it is unlikely that auditory and visual signals come from a common source.

At the neural level, functional magnetic resonance imaging (fMRI), magnetoencephalography, and electroencephalography research (Rohe and Noppeney, 2015b, 2016; Aller and Noppeney, 2019; Cao et al., 2019; Rohe et al., 2019) has recently suggested that the brain flexibly combines sensory signals by dynamically encoding multiple perceptual estimates at distinct cortical levels along the visual and auditory processing hierarchies. For instance, early (50–100 ms) neural processes in primary sensory areas encoded predominantly the spatial locations independently for auditory and visual signals, while later processes (100–200 ms) in posterior intraparietal sulcus (IPS) (IPS1–2) formed spatial representations by combining audiovisual signals. Critically, only at the top of the hierarchy in anterior IPS (IPS3–IPS4, 350–450 ms) were audiovisual signals integrated weighted by their bottom-up sensory reliabilities and top-down task relevance into spatial priority maps that take into account the causal structure of the world.

While previous research has thus convincingly demonstrated that causal inference implicitly influences how observers flexibly combine signals into representations of the environment, it remains unknown which brain systems are critical for solving this causal inference problem. How does the brain determine whether signals arise from common or independent causes based on spatiotemporal correspondence cues? Previous research (Rohe and Noppeney, 2015b, 2016; Aller and Noppeney, 2019; Cao et al., 2019; Rohe et al., 2019) could not address this critical question because observers' implicit causal inference was inherently correlated with the physical correspondence cues (e.g., spatial, temporal, or rate). To define the neural systems underlying causal inference, we need to dissociate the decisional outcome of observers' causal inference from the underlying physical correspondence cues such as, for example, the spatial congruency of audiovisual signals.

This fMRI study investigated how the brain infers the environment's causal structure. Human observers were presented with auditory and visual signals in synchrony at the same (spatially congruent) or separate (spatially incongruent) locations. On each trial, participants decided in an explicit causal inference task whether the AV signals originated from common or separate causes. Importantly, we adjusted the AV disparity individually for each participant, such that observers were ~70% correct in their causal decisions both for AV spatially congruent and incongruent trials. This individual adjustment allowed us to dissociate observers' causal inference from physical AV spatial correspondence cues (i.e., spatial congruency). Based on previous research (Noppeney et al., 2010; Gau and Noppeney, 2016) implicating the prefrontal cortex in arbitrating between integration and segregation, we hypothesized that the dorsolateral prefrontal cortex (DLPFC) plays a critical role in causal inference and decisions.

Materials and Methods

Participants

Thirteen right-handed participants (11 females; mean age, 21.4 years; age range, 18–29 years) gave informed consent to take part in the fMRI experiment. Two participants were excluded because their visual regions could not be reliably defined based on the retinotopic localizer scans acquired after the main experiment. One participant took part only in

the retinotopic localizer session but did not progress to the fMRI experiment. The final study thus consisted of 10 participants. The study was approved by the human research ethics committee at the University of Birmingham. We acknowledge that the number of participants in this extensive multiday psychophysics–fMRI study is low compared with other human neuroimaging research, which may limit the sensitivity and reliability of our group results (Thirion et al., 2007). Guided by the results of the current study, future research will be able to design shorter studies for larger cohorts to further substantiate and expand the findings of this report.

Inclusion criteria

All participants were selected before the fMRI experiment based on the following criteria: (1) no history of neurologic or psychiatric illness; (2) normal or corrected-to-normal vision; (3) reported normal hearing; (4) unbiased sound localization performance in the anechoic chamber (day 1), inside the mock scanner (day 2 and 3) and inside the fMRI scanner (day 5); and (5) 60–80% accuracy for the main task at an individually adjusted audiovisual disparity in the mock scanner (day 2 and 3).

Experimental procedure

Typically, participants completed six sessions, each performed on a separate day. On day 1 (~1 h) the sound stimuli were recorded in an anechoic chamber and participants' sound localization performances were assessed. On days 2 and 3 (~2 h in total), participants were trained to determine the subject-specific AV spatial disparities in a mock scanner. On day 4 (~1 h), participants performed a standard retinotopic localizer task for the retinotopic mapping of visual and parietal cortical areas. On days 5 and 6 (~3 h in total), participants performed the main experiment inside the scanner after final adjustment of the spatial disparity. Eye movements were measured in the mock scanner.

Stimuli and sound recording (day 1)

The visual stimuli were clouds of 20 white dots (diameter, 0.4° visual angle) sampled from a bivariate Gaussian presented on a dark gray background (70% contrast) and were presented for 50 ms. The horizontal SD of the Gaussian was set to a 5° visual angle, and the vertical SD was set to a 2° visual angle.

The sound stimuli were bursts of white noise with 5 ms on/off ramp and were presented for 50 ms. They were recorded individually for each participant with Sound Professionals in-ear binaural microphones in an anechoic chamber in the School of Psychology at the University of Birmingham. The process consisted of displaying the sounds with an Apple Pro Speaker (at a distance of 68 cm from the participants) from –8° to 8° visual angle with 0.5° visual angle spacing, and at ±9° and ±12° visual angle along the azimuth. The participant's head was placed on a chin rest with forehead support and controlled by the experimenter to ensure stable positioning during the recording process. Five stimuli were recorded at each location ("recording set") to ensure that sound locations could not be determined based on irrelevant acoustic cues. On each trial, new visual stimuli were generated, and the auditory stimuli were selected from the recording set of five stimuli.

Assessment of sound localization performance—anechoic chamber (day 1)

Participants were presented with the recorded auditory stimuli from ±12°, ±9°, ±7°, ±5°, ±3°, ±2°, ±1°, and 0° visual angle (10 trials/location in pseudorandomized order) in a forced choice left-right classification task. A cumulative Gaussian was fitted to the percentage of "perceived right responses" as a function of stimulus location using maximum-likelihood estimation (Kingdom and Prins, 2010). We estimated the threshold [point of subjective equality (PSE)] and the slope (inverse of the SD) of the psychometric function as free parameters. The guess rate and lapse rate (0 and 0.01, respectively) were fixed parameters. Participants were included in the fMRI study if their sound localization was unbiased as defined by a PSE/SD ratio of <0.3 (i.e., inclusion criterion 4).

Adjustment of spatial disparity and assessment of sound localization—mock scanner (days 2 and 3)

We adjusted the audiovisual spatial disparity inside the mock scanner individually for each subject to obtain an accuracy of $\sim 70\%$ on the main causal inference task (i.e., common vs separate causes). This individual adjustment of AV spatial disparity allowed us to compare the blood oxygenation level-dependent (BOLD) response to physically identical AV signals that were perceived as coming from common or separate causes and thereby dissociate observer's causal inference and decisions from bottom-up spatial correspondence cues (physical spatial congruency). On day 2, we adjusted subject-specific AV spatial disparities in maximally five adaptive staircases, using a 1-up 2-down, procedure (i.e., up after one error and down after two correct responses with equal step size) that targets 70.71% accuracy on the causal inference task. Each staircase was terminated after a minimum number of 30 trials, when 8 reversals occurred within the last 20 trials and the SD of the AV disparity computed over these reversal was $<2^\circ$ of visual angle (Kingdom and Prins, 2010). The spatial disparity thresholds (i.e., the disparities averaged across the final eight reversals within each staircase) were averaged across the five adaptive staircases within each participant (8.1° visual angles ± 1.2 SEMs across participants). These estimates formed the starting estimate for additional manual fine tuning in subsequent runs of 60 trials where the AV disparity was held constant within a run and adjusted across runs in step size of $1\text{--}2^\circ$ visual angles across runs. Participants were included in the fMRI study if their performance accuracy for the individually selected AV disparity (between 4° and 16° of visual angle) was between 60% and 80% (i.e., inclusion criterion 5). This criterion is required to ensure a sufficient number of trials to compare physically identical AV trials that were perceived as emanating from common or separate causes. On day 3, further fine tuning of AV disparities was performed in subsequent runs of 60 trials as before to ensure that participants' performance was stable over days.

On days 2 and 3, the sound localization performance was further assessed based on a left-right classification task with two selected stimulus locations. Typically, 20–60 repetitions per stimulus location were performed in the mock scanner. Unbiased sound localization was defined as $<30\%$ difference in the accuracy for left-side and right-side stimuli (i.e., inclusion criterion 4).

Final assessment of spatial disparity and sound localization—fMRI scanner (day 5)

To account for differences between the mock scanner and the real fMRI scanner, the AV spatial disparity was finally adjusted in additional one to three runs with constant disparity inside the scanner before the main causal inference fMRI experiment. Similar to the mock scanner, the sound localization performance was finally assessed in the scanner using a left-right classification task for two selected stimulus locations (see inclusion criterion 4). Each participant of the main fMRI study completed at least 20 repetitions per stimulus location for the final auditory stimulus locations, resulting in a group mean localization accuracy of 87% (± 0.02 SEMs across participants).

Experimental design (fMRI, day 5)

In the main fMRI experiment, participants were presented with synchronous auditory and visual spatial signals (stimulus duration, 50 ms) independently sampled from two possible visual angles along the azimuth (e.g., -3° or $+3^\circ$ visual angle with respect to a central fixation cross; Fig. 1A). This resulted in the following four trial types: (1) AV spatially

Figure 1. Experimental stimuli and design. **A**, Time course of one physically AV spatially incongruent and congruent trial. On each trial, observers indicate whether they perceived auditory and visual signals as generated by one or two causes (i.e., explicit causal inference or decision). **B**, The experimental design manipulated (1) visual location (left vs right), (2) auditory location (left vs right), and (3) motor response (left vs right hand) as independent variables. The interaction between auditory and visual location defines physical congruency; causal decision (common vs separate causes) was a dependent variable defined based on participants' responses.

congruent left (i.e., A and V at same location); (2) AV spatially congruent right; (3) AV spatially incongruent with A left and V right; and (4) AV spatially incongruent with A right and V left. On each trial, participants reported whether "A and V signals were generated by common or separate causes as accurately as possible" by pressing a key pad with their left or right thumb. Critically, we alternated and counterbalanced the mapping from left/right hand to the decisional outcome of observers (i.e., common vs separate causes) across fMRI runs within each participant to dissociate the participants' motor response from their causal decision. Each fMRI run included 60 trials per trial type $\times$ 4 trial types (i.e., A left/V left, A left/V right, A right/V left, A right/V right) = 240 trials per run. In addition, we included 20 null events ($\sim 8\%$ of trials). To increase the design efficiency, all four trial types and the null events were presented in a pseudorandomized order with a trial onset asynchrony of 2.3 s.

In summary, the experimental design factorially manipulated the following: (1) visual stimulus location (left vs right); (2) auditory stimulus location (left vs right); and (3) motor response (left vs right hand; Fig. 1B). Based on these experimental manipulations, and participants' causal decisions and motor responses, we characterized the functional properties of brain regions according to the following encoding dimensions: (1) visual space (i.e., V left vs right); (2) auditory space (i.e., A left vs right); (3) spatial (i.e., physical) congruency (i.e., AV spatially congruent vs incongruent); (4) observers' causal inference (i.e., causal decision: common vs separate causes); and (5) motor response (i.e., left vs right hand). For the last two dimensions, the "labels" were based on observers' causal decisions (i.e., common cause vs independent cause response) or motor output (i.e., left-hand vs right-hand response).

Eye movement recording and analysis

To address potential concerns that our results may be confounded by eye movements, we evaluated participants' eye movements based on eye-tracking data recorded concurrently during the causal inference task inside the mock scanner. Eye recordings were calibrated ($\sim 35^\circ$ horizontally and $\sim 14^\circ$ vertically) to determine the deviation from the fixation cross. Fixation position was *post hoc* offset corrected. For each position, the number of saccades (radial velocity threshold = $30^\circ/\text{s}$, acceleration threshold = $8000^\circ/\text{s}^2$, motion threshold = 0.15° , radial amplitude $> 1^\circ$) and eye blinks were quantified (0–875 ms after stimulus onset). Critically, the 2 (visual left, right) $\times$ 2 (auditory left, right) repeated-measures ANOVAs on the stimulus conditions performed separately for (1) the percentage of saccades or (2) the percentage of eye blinks revealed no significant main effects or interactions, indicating that differences in BOLD response between conditions are unlikely to be because of eye movement confounds.

Figure 2. Univariate results of the main effect of causal decision and the interaction of causal decision and physical spatial congruency. Activation increases for causal decisional outcome: separate > common cause (green, $p_{FWE} < 0.05$ at the cluster level corrected for multiple comparisons within the entire brain, with an auxiliary uncorrected voxel threshold of $p < 0.001$) and activation increases for causal decision $\times$ physical AV spatial congruency interaction: incorrect > correct (red, $p_{FWE} < 0.05$ at the cluster level corrected for multiple comparisons within the entire brain, with an auxiliary uncorrected voxel threshold of $p < 0.001$) are rendered on an inflated canonical brain. Bar plots (across participants, mean $\pm$ SEM) overlaid with bee swarm plots (for individual participants) show the parameter estimates (averaged across all voxels in the black encircled cluster) in the (1) left inferior frontal sulcus/precentral sulcus, (2) bilateral superior frontal gyrus, (3) right posterior intraparietal sulcus, and (4) right anterior intraparietal sulcus that are displayed on axial slices of a mean image created by averaging the participants' normalized structural images. L, Left; R, right; a.u., arbitrary unit.

Experimental setup

Visual and auditory stimuli were presented using Psychtoolbox version 3.0.11 (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) running under MATLAB R2011b (MathWorks) on a MacBook Pro (Mac OSX version 10.6.8). For the main task, visual stimuli were backprojected to a Plexiglas screen using a D-ILA projector (model DLA-SX21, JVC) visible to the participant through a mirror mounted on the magnetic resonance (MR) head coil. Auditory stimuli were delivered via Sennheiser HD 280 Pro headphones (in the anechoic chamber), Sennheiser HD 219 headphones (in the mock scanner), and MR Confon HP-VS03 headphones (in the scanner). Participants' eye movements were recorded in the mock scanner using an Eyelink Remote system (SR Research) at a sampling rate of 1000 Hz.

MRI data acquisition

A 3 T Philips Achieva scanner was used to acquire both T1-weighted anatomic images (TR, 8.4 ms; TE, 3.8 ms; 175 slices; image matrix, 288×232 ; spatial resolution, $1 \times 1 \times 1 \text{ mm}^3$ voxels) and T2*-weighted echo-planar imaging (EPI) images with BOLD contrast (fast field echo; TR, 2600 ms; TE, 40 ms; 38 axial slices acquired in ascending direction;

image matrix, 80×80 ; spatial resolution, $3 \times 3 \times 3 \text{ mm}^3$ voxels without gap). Typically, there were 10–12 runs with 240 volumes per run over two sessions. The first four volumes were not acquired to allow T1 equilibration effects. In one participant, we repeated a session, since the participant's accuracy was 15% lower than the mean accuracy of the remaining sessions. In another participant, two runs were excluded because of technical problems with the setup. In three participants, one to two runs were removed from further analysis to be able to counterbalance the left and right response hands across runs (see Experimental design).

Statistical analysis

Behavioral data analysis

For the eye movement analysis of the mock scanner data, (1) percentage of saccades and (2) percentage of eye blinks of the participants were entered into separate 2 (visual: left, right) $\times$ 2 (auditory: left, right) repeated-measures ANOVA.

For the reaction time analysis of the main fMRI experiment, participants' response times (i.e., condition specific across trial median) were entered into 2 (physical: congruent, incongruent) $\times$ 2 (perceptual: congruent, incongruent) repeated-measures ANOVA.

Table 1. Univariate results of the main effects of stimulus location and motor response

Brain regions	MNI coordinates, mm			z score, peak	Cluster size, number of voxels	p_{FWE} value, cluster
	x	y	z			
Visual L > visual R						
R calcarine sulcus	12	−72	−2	7.58	935	<0.001
R cuneus	10	−86	20	7.04		
Visual R > visual L						
L middle occipital gyrus	−48	−78	10	7.80	1869	<0.001
L superior occipital gyrus	−20	−86	20	6.96		
L calcarine sulcus	−10	−86	2	5.87		
Auditory R > auditory L						
L planum temporale	−56	−44	14	4.66	274	<0.001
Motor L > motor R						
R postcentral gyrus	54	−16	50	>8	1964	<0.001
R precentral gyrus	40	−16	54	>8		
Motor R > motor L						
L precentral gyrus	−36	−24	52	>8	2153	<0.001
Central sulcus	−44	−24	50	>8		
L postcentral gyrus	−52	−18	50	>8		
L rolandic operculum	−46	−22	18	6.04	346	<0.001

$p_{FWE} < 0.05$ at the cluster level corrected for multiple comparisons within the entire brain, with an auxiliary uncorrected voxel threshold of $p < 0.001$. We also report the z score of the peak voxel (or several peak voxels) with their corresponding MNI coordinates. L, Left; R, right.

Unless stated otherwise, we report effects that are significant at $p < 0.05$.

fMRI data preprocessing

The data were analyzed with statistical parametric mapping [SPM8; Wellcome Trust Center for Neuroimaging, London, UK (<http://www.fil.ion.ucl.ac.uk/spm/>); Friston et al., 1994a] running on MATLAB R2014a. Scans from each participant were realigned using the first as a reference, unwarped, and corrected for slice timing. The time series in each voxel were high-pass filtered to 1/128 Hz. For the conventional univariate analysis, the EPI images were spatially normalized into MNI standard space (Ashburner and Friston, 2005), resampled to $2 \times 2 \times 2$ mm³ voxels, and spatially smoothed with a Gaussian kernel of 6 mm FWHM. For the multivariate decoding analysis, the EPI images were analyzed in native participant space and spatially smoothed with a Gaussian kernel of 3 mm FWHM. For the retinotopic analysis, the data were analyzed in native space and without additional smoothing.

fMRI data analysis

Data were modeled in an event-related fashion with regressors entered into the design matrix after convolving each event-related unit impulse (representing a single trial) with a canonical hemodynamic response function and its first temporal derivative. Realignment parameters were included as nuisance covariates to account for residual motion artifacts.

Univariate fMRI analysis. For the conventional univariate analysis, the general linear model (GLM) modeled the 16 conditions in our 2 (visual: left, right) $\times$ 2 (auditory: left, right) $\times$ 2 (decisional outcome: common, separate causes) $\times$ 2 (hand response: left, right) factorial design. Condition-specific effects for each participant were estimated according to the GLM and passed to a second-level repeated-measures ANOVA as contrasts. Inferences were made at the between-subjects level to allow for random-effects analysis and inferences at the population level (Friston et al., 1999). At the between-subjects level, we tested for the effects of visual signal location (left vs right), auditory signal location (left vs right), hand response (left vs right), physical AV spatial congruency (congruent vs incongruent), and causal inference or decision (decisional outcome: common vs separate causes; Fig. 2, Tables 1, 2).

We report activations at $p < 0.05$ at the cluster level corrected for multiple comparisons within the entire brain, with an auxiliary uncorrected voxel threshold of $p < 0.001$ (Friston et al., 1994b).

Multivariate decoding analysis. To ensure that multivariate decoding is valid and unbiased, it is critical that parameter estimates were estimated with comparable precision (i.e., inverse of variance). Hence, their estimation should be based on the same number of trials. Because the

number of trials may vary across conditions that are defined by observers' causal decisions (e.g., comparing "common cause" vs "independent cause" decisions), we generated design matrices in which we explicitly matched the number of trials per regressor and the number of regressors across conditions. First, each regressor always modeled exactly eight trials from one particular condition. As a result of this subsampling procedure, all parameter estimates that were entered into the multivariate pattern analyses were estimated with comparable precision. Second, we determined the number of regressors (maximally, seven for each condition) such that they were matched across conditions for each comparison (e.g., common cause vs separate cause decision). For instance, to dissociate causal decision (i.e., common vs separate causes) from physical spatial congruency (i.e., congruent vs incongruent), visual (i.e., left vs right), or auditory (i.e., left vs right) location or motor response (i.e., left vs right hand), we defined a GLM that included an equal number of regressors for "common cause" and "separate cause" decisions separately for each condition within the 2 (auditory: left vs right) $\times$ 2 (visual: left vs right) $\times$ 2 (motor: left vs right) design. The remaining trials were entered into one single regressor of no interest to account for general stimulus-related responses. To ensure that the decoding results did not depend on particular subsamples, we repeated this matching and subsampling procedure (with subsequent GLM estimation and multivariate pattern analysis) 10 times and averaged the decoding accuracy across those 10 iterations.

This subsampling and matching procedure ensured that the parameter estimates for common versus separate cause decisions were matched with respect to all other factors (i.e., auditory, visual, physical spatial congruency, and motor responses). This allowed us to identify regions encoding participants' causal decisions unconfounded by physical spatial congruency, auditory or visual location, or motor output. Likewise, we decoded participants' motor responses unconfounded by auditory or visual location, causal decisional outcome, or physical spatial congruency.

For multivariate pattern analyses, we trained a linear support vector classification model as implemented in LIBSVM version 3.20 (Chang and Lin, 2011). More specifically, the voxel response patterns were extracted in a particular region of interest (e.g., A1, see below for definition of region of interest) from the parameter estimate images corresponding to the magnitude of the BOLD response for each condition and run as described above. Each parameter estimate image was based on exactly eight trials (see above). Decoding of experimental factors such as visual location, auditory location, or physical congruency was typically based on 28 parameter estimate images per run $\times$ 10 runs = 280 parameter estimate images in total (for details, see MRI data acquisition). The number of parameter estimate images for decoding "causal decisions" or

Table 2. Univariate results of the main effect of causal decision and the interaction of causal decision and physical spatial congruency

Brain regions	MNI coordinates, mm			z score, peak	Cluster size, number of voxels	p_{FWE} value, cluster
	x	y	z			
Causal decision: separate > common cause						
R posterior intraparietal sulcus	40	−74	34	3.87	229	<0.001
R anterior intraparietal sulcus	38	−46	38	3.50	183	0.001
R inferior frontal sulcus	42	30	18	4.09	179	0.002
R middle frontal gyrus	50	20	8	4.08		
R superior frontal sulcus	26	6	52	3.89	150	0.004
R anterior insula	30	26	−6	4.86	139	0.006
R precuneus	4	−68	46	3.73	112	0.018
Causal decision × physical spatial congruency interaction: correct > incorrect						
R putamen	28	6	0	5.60	757	<0.001
L putamen	−26	2	−8	4.73	388	<0.001
Causal decision × physical spatial congruency interaction: incorrect > correct						
L superior frontal gyrus (medial wall)	−6	14	50	5.63	1589	<0.001
R superior frontal gyrus (medial wall)	6	12	54	5.12		
L anterior cingulate sulcus/gyrus	−2	20	38	4.89		
L inferior frontal sulcus	−50	22	26	5.32	716	<0.001
L precentral sulcus	−40	2	28	4.47		
L anterior insula	−36	18	6	5.47	585	<0.001
R anterior insula	38	16	6	4.27	217	<0.001

$p_{FWE} < 0.05$ at the cluster level corrected for multiple comparisons within the entire brain, with an auxiliary uncorrected voxel threshold of $p < 0.001$. We also report the z score of the peak voxel (or several peak voxels) with their corresponding MNI coordinates. L, Left; R, right.

“motor responses” depended on participants’ choices and hence varied across participants (mean number of parameter estimate images for causal decisions, 116; range across participants, 82–194; mean number of parameter estimate images for motor responses, 225; range across participants, 188–278). To implement a leave-one-run-out cross-validation procedure, parameter estimate images from all but one run were assigned to the training dataset, and images from the “left-out run” were assigned to the test set. Parameter estimate images for training and test datasets were normalized and scaled independently using Euclidean normalization of the images and mean centering of the features. Support vector classification models were trained to learn the mapping from the condition-specific fMRI response patterns to the class labels from all but one run according to the following dimensions: (1) visual signal location (left vs right); (2) auditory signal location (left vs right); (3) physical AV spatial congruency (congruent vs incongruent); (4) causal decisional outcome (common vs separate causes); and (5) motor response (left vs right hand). The model then used this learned mapping to decode the class labels from the voxel response patterns of the remaining run. First, we report decoding accuracies as box plots in Figure 3 to provide insight into intersubject variability. Second, we show the weighted sum of the BOLD parameter estimates for each class in each region of interest (ROI) again as box plots in Figure 4. The weighted sum BOLD parameter estimates illustrate as a summary index the multivariate differences in BOLD responses between class 1 and 2, which form the basis for multivariate pattern decoding.

Nonparametric statistical inference was performed both at the “within-subjects” level and the “between-subjects” (group) level to allow for generalization to the population (Nichols and Holmes, 2002). For the within-subjects level, we generated a null distribution of decoding accuracies for each participant individually by permuting the condition-specific labels of the parameter estimates for each run (i.e., not of individual trials to preserve the autocorrelation structure) and calculating the decoding accuracies for all permutations (500 permutations × 10 GLMs = 5000 repetitions in total). We computed the p value as the fraction of permutations in which the decoding accuracy obtained from the permuted data exceeded the observed decoding accuracy (i.e., directed or one-sided permutation test).

For the between-subjects level permutation test, we first determined the chance decoding accuracy individually for each participant as the average decoding accuracy across all permutations. Next, we subtracted the empirically defined chance accuracy from the corresponding observed decoding accuracy in each participant. Then we generated a

null distribution of decoding accuracies as follows. We randomly assigned the $\pm$ sign to the subject-specific deviations of the observed decoding accuracy from chance decoding accuracy for each participant. We formed the across-participants’ mean. We repeated this procedure for all possible sign assignments ($2^{10} = 1024$ cases for 10 participants). We then compared the original across-participants’ mean of the observed decoding accuracies with the thus generated null distribution. We computed the p value as the fraction of permutations in which the signed decoding accuracy deviation exceeded the observed decoding accuracy difference (i.e., directed or one-sided permutation test).

Likewise, we assessed whether the DLPFC mainly encodes observers’ causal decisional choices (common vs separate sources) rather than the remaining dimensions in our paradigm using nonparametric permutation testing as described above: briefly, we (1) computed the deviations from chance decoding accuracy for each of the five information dimensions individually for each participant, (2) calculated the differences in these relative decoding accuracies between information dimensions for each participant (e.g., causal decision minus physical spatial congruency), and (3) formed the across-participants’ mean of those differences in decoding accuracy. To generate a null distribution for these across-participants’ means, we flipped the sign of these differences randomly for each participant and recomputed the across participants’ mean for each permutation. We computed the p value as the fraction of across-participants’ means (generated via permutation) that exceeded the observed across-participants’ mean.

Unless otherwise stated, we report decoding accuracies at $p < 0.05$ (based on one-sided tests). We apply Bonferroni corrections for multiple comparisons across all 11 ROIs. In Figure 3 and Table 3, we report the uncorrected p values based on between-subjects level permutation test and indicate using a triangle whether these p values are significant when the threshold is adjusted according to Bonferroni correction (i.e., 0.05/11 ROI = 0.0045). In Table 3, we also report the number of subjects that were individually significant (i.e., uncorrected $p < 0.05$) based on a within-subject permutation test (in brackets, we list the number of subjects that were significant after Bonferroni correction for multiple comparisons across the 11 ROI (i.e., uncorrected $p < 0.0045$). Please note because the number of permutations is 500 at the within-subjects level and 1024 at the between-subjects level, the minimal uncorrected p values are $1/500 = 0.002$ and $1/1024 = 0.00098$, respectively. Hence, after Bonferroni correction even the most significant p values will be indicated only by a single triangle to indicate that the Bonferroni-corrected family-wise error (FWE) rate is <0.05 (i.e., $0.02 \times 11 = 0.22$ and $0.00098 \times 11 =$

Figure 3. Multivariate pattern results along the visual and auditory spatial cortical hierarchy. Support vector classification decoding accuracy for: (1) V, visual location: left vs right; (2) A, auditory location: left vs right; (3) S, physical spatial congruency: congruent vs incongruent; (4) D, causal decisional outcome: common versus separate causes; and (5) M, motor response: left versus right hand in the ROIs, as indicated in the figure. Box plots show the accuracies across participants (box for median and interquartile range, whiskers for lowest and highest data points, dots for outside of 1.5 interquartile range). Significance is indicated by $**p < 0.01$, $***p < 0.001$, $\triangle p < 0.0045$; the single triangle indicates that the p value is significant when adjusting the threshold according to Bonferroni correction (i.e., $p < 0.0045 \times 11\text{ROIs} = 0.0495$). The ROIs are delineated on the surface of an inflated single participant brain.

0.01, respectively). Guided by a priori hypotheses, we did not apply Bonferroni correction for testing: visual left/right location in V1 (primary visual cortex), V2 (secondary visual cortex), V3, V3AB (higher-order visual cortices); auditory left/right location in A1, planum temporale (PT); motor left/right hand response in precentral gyrus (PCG); and causal decision (common vs separate causes) in DLPFC. Because we predicted DLPFC to encode mainly causal decisions, we also report the comparisons of decoding accuracy for causal decisions relative to other information dimensions without Bonferroni correction.

Visual retinotopic localizer

Standard phase-encoded polar angle retinotopic mapping (Serenio et al., 1995) was used to define regions of interest along the dorsal visual processing hierarchy (Rohe and Noppeney, 2015b). Participants viewed a checkerboard background flickering at 7.5 Hz through a rotating wedge aperture of 70° width. The periodicity of the apertures was 44.2 s. After the fMRI preprocessing steps (see fMRI data preprocessing), visual responses were modeled by entering a sine and cosine convolved with the hemodynamic response function as regressors into a GLM. The preferred polar angle was determined as the phase lag for each voxel, which is the angle between the parameter estimates for the sine and the cosine.

The preferred phase lags for each voxel were projected on the participants' reconstructed and inflated cortical surface using Freesurfer version 5.3.0 (Dale et al., 1999). Visual regions V1–V3, V3AB, and parietal regions IPS0–4 were defined as phase reversal in angular retinotopic maps. IPS0–4 were defined as contiguous, approximately rectangular regions based on phase reversals along the anatomic IPS (Swisher et al., 2007) and guided by group-level retinotopic probabilistic maps (Wang et al., 2015).

Region of interests used for decoding analysis

For the decoding analyses, all ROIs were combined from the left and right hemispheres.

Occipital, parietal, and frontal eye field regions. Regions in the occipital and parietal cortices were defined based on retinotopic mapping, as described above. The frontal eye field (FEF) was defined by an inverse normalized group-level retinotopic probabilistic map (Wang et al., 2015). The resulting subject-level probabilistic map was thresholded at the 80th percentile, and any overlap with the motor cortex was removed.

Auditory, motor and prefrontal regions. These regions were based on labels of the Destrieux atlas of Freesurfer 5.3.0 (Dale et al., 1999; Destrieux et al., 2010). The primary auditory cortex was defined as the

Figure 4. Characterization of BOLD response patterns. BOLD response parameter estimates for each of the two classes (e.g., left and right visual location) are summed within each region weighted by the support vector classification weights. **A**, Support vector classification for (1) V, visual location: left versus right; (2) A, auditory location: left versus right; (3) S, physical spatial congruency: congruent versus incongruent; (4) D, causal decisional outcome: common versus separate causes; and (5) M, motor response: left versus right hand in the ROIs, as indicated in the figure. Box plots show the weighted sum of parameter estimates across participants (box for median and interquartile range, whiskers for lowest and highest data points, dots for outside of 1.5 interquartile range). **B**, Support vector classification for causal decisional outcome (i.e., DC and DS) trained separately for SC and SI stimuli.

anterior transverse temporal gyrus (Heschl's gyrus, HG). The higher auditory cortex was formed by merging the transverse temporal sulcus and the PT. The motor cortex was based on the precentral gyrus. The DLPFC was defined by combining the superior and middle frontal gyri and sulci as previously described (Yendiki et al., 2010). In line with the study by Rajkowska and Goldman-Rakic (1995), we limited the superior frontal gyrus and sulcus to Talairach coordinates $y = 26$ and $y = 53$, respectively, and the middle frontal gyrus and sulcus to Talairach coordinates $y = 20$ and $y = 50$, respectively.

Results

Behavioral results

Observers' performance accuracy in their causal decisions during the main experiment inside the MRI scanner indicated that the individual adjustment of spatial disparity was adjusted appropriately. As expected participants were $\sim 70\%$ correct when deciding whether auditory and visual signals originated from common or independent causes with a small bias toward

common causes decisions (accuracy_{SC} = $77 \pm 1.7\%$, accuracy_{SI} = $66 \pm 2.2\%$ with the index SC (physically spatially congruent) and SI (physically spatially incongruent; $d' = 1.07 \pm 0.12$; bias: 0.16 ± 0.03 ; and mean $\pm$ SEM in all cases).

A 2 (physical: spatially congruent, incongruent) $\times$ 2 (decision: common, separate causes) repeated-measures ANOVA of response times revealed a significant main effect of causal decisional outcome ($F_{(1,9)} = 8.266$, $p = 0.018$) and a significant physical spatial congruency $\times$ causal decision interaction ($F_{(1,9)} = 15.621$, $p = 0.003$). Overall, participants were slower on trials where they perceived AV signals as caused by separate events (i.e., averaged across physically spatially congruent and incongruent trials). *Post hoc* paired *t* tests of the simple main effects revealed that participants were significantly faster judging physically spatially congruent stimuli as coming from common cause and physically spatially incongruent stimuli as coming from separate causes ($RT_{SC,DC} = 0.89 \pm 0.05$ s; $RT_{SI,DS} = 0.93 \pm 0.06$ s; $RT_{SC,DS} = 1.02 \pm 0.06$ s; $RT_{SI,DC} = 0.96 \pm 0.06$ s; with RT for

Table 3. Multivariate pattern classification results

Brain regions	Visual location left vs right		Auditory location left vs right		Physical spatial congruent vs incongruent		Decision separate vs common cause		Motor response left vs right	
	<i>p</i> value	Subject	<i>p</i> value	Subject	<i>p</i> value	Subject	<i>p</i> value	Subject	<i>p</i> value	Subject
HG	0.409	3 (0)	0.092	5 (5)	0.058	3 (1)	0.826	1 (0)	0.001 ^Δ	10 (10)
PT	0.028	5 (2)	0.001 ^{***}	8 (8)	0.002 ^Δ	8 (4)	0.337	3 (2)	0.001 ^Δ	9 (9)
V1	0.001 ^{***}	10 (10)	0.023	5 (4)	0.021	6 (4)	0.040	5 (2)	0.232	5 (3)
V2	0.001 ^{***}	10 (10)	0.009	7 (5)	0.078	5 (4)	0.003 ^Δ	3 (2)	0.041	6 (3)
V3	0.001 ^{***}	10 (10)	0.033	4 (4)	0.038	4 (3)	0.002 ^Δ	5 (2)	0.004 ^Δ	8 (7)
V3AB	0.001 ^{***}	10 (10)	0.010	8 (7)	0.030	5 (4)	0.013	6 (4)	0.078	3 (2)
IPSO–2	0.001 ^Δ	10 (10)	0.001 ^Δ	8 (6)	0.002 ^Δ	8 (6)	0.004 ^Δ	8 (6)	0.002 ^Δ	7 (5)
IPS3–4	0.001 ^Δ	10 (9)	0.006	7 (7)	0.001 ^Δ	8 (7)	0.001 ^Δ	7 (5)	0.001 ^Δ	10 (10)
FEF	0.001 ^Δ	9 (9)	0.004 ^Δ	8 (7)	0.003 ^Δ	8 (7)	0.003 ^Δ	8 (7)	0.003 ^Δ	8 (7)
PCG	0.006	7 (7)	0.035	6 (5)	0.171	5 (5)	0.704	2 (1)	0.001 ^{***}	10 (10)
DLPFC	0.013	7 (5)	0.001 ^Δ	7 (4)	0.002 ^Δ	6 (4)	0.002 ^{**}	8 (8)	0.295	3 (1)

p values (uncorrected) indicate better than chance decoding accuracy at the group level based on between-subjects permutation test; ^{**}*p* < 0.01, ^{***}*p* < 0.001, ^Δ*p* < 0.0045 (i.e., significant after Bonferroni correction for 11 regions of interest); “subjects” indicates the number of subjects that are associated with better than chance decoding accuracy based on within-subjects permutation test at *p* < 0.05 uncorrected (in brackets: number of subjects with *p* < 0.0045; i.e., significant after Bonferroni correction for 11 regions of interest unless guided by a priori hypothesis).

reaction time and the index DC for common cause decision and DS for separate cause decisions). In other words, observers were faster on their correct responses than on their wrong responses, suggesting that trials with wrong responses were associated with a greater degree of decisional uncertainty. Importantly, we decoded observers’ decisional outcome (i.e., common cause vs separate cause judgements) pooled over correct and incorrect responses (i.e., common cause and separate cause judgements included correct and incorrect trials). Hence, our decoding focused on decisional outcome regardless of decisional uncertainty.

fMRI analysis: univariate results

The current study focused primarily on multivariate pattern analyses to characterize explicit causal inference in audiovisual perception. For completeness, we also provide a brief summary of the results from the conventional univariate analyses (Fig. 2, Tables 1, 2).

Main effects of visual and auditory location and motor response

As expected, the spatially lateralized auditory and visual stimuli elicited stronger activations in the contralateral hemifield (Table 1). Right relative to left visual stimuli increased activations in the left calcarine sulcus, and the middle and superior occipital gyri, while left relative to right visual stimuli increased activations in the right calcarine sulcus and right cuneus. Likewise, right relative to left auditory stimuli increased activations in the left planum temporale (PT).

Moreover, we observed the expected lateralization effects for motor responses: left-hand relative to right-hand responses were associated with greater activations in the right precentral and postcentral gyri, while right-hand relative to left-hand responses were associated with greater activations in the left precentral and postcentral gyri, the central sulcus, and the left rolandic operculum (Table 1).

Main effect of physical AV spatial congruency and observers’ causal decision

We did not observe any significant effects of physical spatial congruency (i.e., interaction between visual and auditory location) most likely because the spatial disparity was too small to elicit the multisensory incongruency effects observed in classical suprathreshold paradigms (Hein et al., 2007; van Atteveldt et al., 2007; Noppeney et al., 2008, 2010; Gau and Noppeney, 2016). However, the outcome of observers’ causal decision influenced

brain activations: stimuli that were judged to come from separate (relative to common) causes increased activations in a widespread right lateralized system including the intraparietal sulcus, the superior and inferior frontal sulci, and the insula (Fig. 2, Table 2). Thus, in our threshold paradigm, observer’s decisional outcome separate causes and hence their perceived AV incongruency increased activations usually observed for physical incongruency. These activation increases for separate causes decisions also dovetail nicely with observers’ longer response times for these trials (see Behavioral results).

Interaction between physical AV spatial congruency and causal decision

To understand the interaction between physical spatial congruency and observers’ causal decision, we note that the interaction is equivalent to correct versus incorrect responses. We found bilateral putamen activations for correct > incorrect responses (Table 2) that is in concordance with previous results showing a role of putamen in audiovisual conditions associated with faster and more accurate responses (von Saldern and Noppeney, 2013). For incorrect > correct responses, we observed increased activations in the prefrontal cortex (e.g., bilateral superior frontal gyri and insulae, inferior frontal sulcus; Fig. 2, Table 2), which have previously been associated with greater executive demands (Noppeney et al., 2008; Werner and Noppeney, 2010a).

fMRI analysis: multivariate results

Using multivariate pattern analyses, we assessed which of our regions of interest encode the key dimensions of our experimental design, as follows: (1) visual signal location (left vs right); (2) auditory signal location (left vs right); (3) physical spatial congruency (congruent vs incongruent); (4) causal decisional outcome (common vs separate causes); and (5) motor response (left vs right hand; Fig. 1B). The multivariate pattern classification results are provided in Table 3, and the decoding accuracies are shown in Figure 3. Further, we show the weighted sum BOLD parameter estimates as summary indices to illustrate the multivariate BOLD response patterns that form the basis for multivariate pattern classification separate for class 1 and 2 in each region in Figure 4.

Decoding of auditory and visual location

Visual location could be decoded significantly better than chance from BOLD response patterns in visual areas including V1, V2,

V3, and V3AB (Fig. 3). In addition, visual location was represented in the parietal cortex (IPS0–4) as well as in the FEF, which is consistent with the well established retinotopic organization of those cortical regions (Swisher et al., 2007; Silver and Kastner, 2009; Wang et al., 2015). Auditory location could be decoded significantly better than chance from the PT as a higher-order auditory area previously implicated in spatial processing (Rauschecker and Tian, 2000; Warren and Griffiths, 2003; Moerel et al., 2014) as well as along the dorsal auditory processing stream including the posterior parietal cortex (IPS0–2), the FEF, and the DLPFC (Rauschecker and Tian, 2000; Arnott et al., 2004; Rauschecker and Scott, 2009; Recanzone and Cohen, 2010; Fig. 3).

Decoding of physical AV spatial congruency and observers' causal decision

By titrating observers' accuracy to ~70% correct, our design allowed us to dissociate observers' causal decision from physical spatial congruency. However, it is important to emphasize that this threshold design will also limit the maximal accuracy with which physical spatial disparity and observers' causal decisions can be decoded from fMRI activation patterns. This is because the small spatial disparity will make observers' commit to a motor response despite a high level of decisional uncertainty.

Physical AV spatial congruency could be decoded from higher-order association cortices encompassing the parietal cortex (IPS0–4), the FEF, and DLPFC as well as the planum temporale (Fig. 3). These results are consistent with the classical view of multisensory processing in which primary auditory and visual cortices are specialized for processing signals of their preferred sensory modality, and higher-order frontoparietal association cortices are involved as convergence zones in combining signals across the senses (Felleman and Van Essen, 1991; Calvert, 2001; Wallace et al., 2004a; Romanski, 2012).

Critically, adjusting spatial disparity individually for each participant to obtain 70% performance accuracy allowed us to compare physically spatially congruent (respectively incongruent) stimuli that were judged as coming from one common cause versus separate causes. In other words, the individual threshold adjustment allowed us to identify regions encoding participants' causal decisions regardless of the physical spatial congruency of the underlying AV signals (see Materials and Methods regarding the subsampling and matching procedures). In line with our predictions, participants' causal decisional outcome could be decoded from DLPFC (Fig. 3). Critically, observers' causal decisions could be decoded from DLPFC better than from any other stimulus feature ($P_{D-V} = 0.0107$, $P_{D-A} = 0.0342$, $P_{D-S} = 0.0078$, $P_{D-M} = 0.0020$; with indexes D – V, D – A, D – S, D – M for comparing the accuracies of causal decision with visual, auditory, physical spatial congruency, and motor response, respectively), suggesting a key role for DLPFC in causal inference. In addition, observers' causal decision could be decoded to a lesser extent from activation patterns in a widespread system encompassing FEF, IPS0–4, and even the early visual areas such as V2 (Fig. 3).

Given the significant interaction between causal decision and spatial disparity in our behavioral and univariate fMRI analyses, we assessed in a subsequent analysis whether observers' causal decisions can be decoded similarly from activation patterns for spatially congruent and disparate audiovisual signals. Indeed, we were able to decode observers' causal decisions similarly for spatially congruent and incongruent audiovisual signals. The decoding accuracy for DLPFC was 60.02 ± 1.78 (group mean $\pm$ SEM; group-level permutation test: $p = 0.001$ uncorrected) for SC

and 58.72 ± 2.06 (group mean $\pm$ SEM, group-level permutation test: $p = 0.003$ uncorrected) for SI signals. These results suggest that the DLPFC encodes observers' decisional choices for both spatially congruent and incongruent signals.

For completeness, we also assessed the decoding accuracies for (1) IPS0–2: 56.40 ± 1.27 for SC (group mean $\pm$ SEM, group-level permutation test: $p = 0.003$ uncorrected) and 55.60 ± 1.35 for SI (group mean $\pm$ SEM, group-level permutation test: $p = 0.002$ uncorrected); (2) IPS3–4: 55.37 ± 1.84 for SC (group mean $\pm$ SEM, group-level permutation test: $p = 0.013$ uncorrected) and 55.09 ± 1.35 for SI (group mean $\pm$ SEM, group-level permutation test: $p = 0.003$ uncorrected); (3) FEF: 58.14 ± 1.35 for SC (group mean $\pm$ SEM, group-level permutation test: $p = 0.002$ uncorrected) and 55.98 ± 0.98 for SI (group mean $\pm$ SEM, group-level permutation test: $p = 0.001$ uncorrected).

Decoding of motor response

We also ensured by experimental design that participants' causal decisions were orthogonal to their motor responses (i.e., left hand vs right hand) by alternating the mapping from participants' causal decisions to the selected hand response across runs. Not surprisingly, the motor response was decoded with a high accuracy from the precentral gyrus (Fig. 3). In addition, we were able to decode observers' motor responses from the FEF, IPS0–4, and V3AB. Further, we were able to decode participants' motor responses from planum temporale (PT) and Heschl's gyrus (HG). The latter decoding of sensory–motor information from activation patterns in Heschl's gyrus may potentially be attributed to activations from the neighboring secondary somatosensory areas (see above for univariate results in the left Rolandic operculum).

Discussion

To form a coherent percept of the world, the brain needs to integrate sensory signals generated by a common cause and segregate those from different causes (Noppeney, 2020). The human brain infers whether or not signals originate from a common cause or event based on multiple correspondence cues such as spatial disparity (Slutsky and Recanzone, 2001; Lewald and Gusk, 2003; Wallace et al., 2004b; Recanzone, 2009), temporal synchrony (Munhall et al., 1996; Noesselt et al., 2007; van Wassenhove et al., 2007; Lewis and Noppeney, 2010; Maier et al., 2011; Lee and Noppeney, 2011b; Parise et al., 2012; Magnotti et al., 2013; Parise and Ernst, 2016), or semantic and other higher-order correspondence cues (Welch, 1999; Parise and Spence, 2009; Sadaghiani et al., 2009; Adam and Noppeney, 2010; Noppeney et al., 2010; Bishop and Miller, 2011; Lee and Noppeney, 2011a). As a result, observers' causal decisions have previously been inherently correlated with the congruency of the audiovisual signals (Rohe and Noppeney, 2015b, 2016; Aller and Noppeney, 2019; Cao et al., 2019; Rohe et al., 2019), making it challenging to dissociate observers' causal decisions from the underlying physical correspondence cues such as audiovisual spatial disparity.

To dissociate the neural processes associated with participants' causal decisions from those driven by the physical AV spatial congruency cues, we adjusted the audiovisual spatial disparity individually for each participant to enable a threshold accuracy of 70%. As a result of external and internal noise (Faisal et al., 2008), spatially congruent audiovisual signals were perceived as coming from the same source in ~70% of cases. Conversely, spatially disparate audiovisual signals were perceived as coming from independent sources in ~70% of cases. This causal

uncertainty allowed us to select and compare physically identical audiovisual signals that were perceived as coming from common or separate causes. Moreover, we dissociated participants' causal decisions from their motor responses by counterbalancing the mapping between causal decision (i.e., common vs separate causes) and motor response (i.e., left vs right hand) over runs. In summary, our experimental design enabled us to characterize a system of brain regions with respect to the following five different "encoding dimensions": (1) visual space (left vs right); (2) auditory space (left vs right); (3) physical spatial congruency (congruent vs incongruent); (4) causal inference and decision (common vs separate causes); and (5) motor response (left vs right hand).

Unsurprisingly, our multivariate decoding results demonstrate that low-level visual areas (V1–3) encode predominantly visual space, PT auditory space, and precentral gyrus participants' motor responses. Physical spatial congruency could be decoded from planum temporale, all parietal areas (IPS0–4), and prefrontal cortices (DLPFC, FEF). This profile of results is consistent with the classical hierarchical organization of multisensory perception, according to which low-level sensory cortices process signals mainly from their preferred sensory modalities and higher-order cortical regions combine signals across the senses (Felleman and Van Essen, 1991; Mesulam, 1998; Calvert, 2001; Kaas and Collins, 2004; Wallace et al., 2004a). This view has been challenged by studies showing multisensory interactions already at the primary cortical level (Molholm et al., 2002; Ghazanfar et al., 2005; Senkowski et al., 2005; Ghazanfar and Schroeder, 2006; Hunt et al., 2006; Kayser and Logothetis, 2007; Lakatos et al., 2007; Driver and Noesselt, 2008; Werner and Noppeney, 2011). However, in primary sensory cortices, stimuli from the nonpreferred sensory modality typically modulated the response magnitude or salience rather than the spatial representation of stimuli from the preferred sensory modality. Likewise, previous multivariate pattern analyses showed that a synchronous yet displaced auditory signal had minimal impact on the spatial representations in primary visual cortices (Rohe and Noppeney, 2015b, 2016). Only later in the processing hierarchy in posterior and anterior parietal cortices were spatial representations formed that integrated auditory and visual signals weighted by their bottom-up reliabilities (IPS0–4) and top-down task relevance (IPS3–4; Rohe and Noppeney, 2015b, 2016, 2018; Aller and Noppeney, 2019). Our current findings thus lend further support for this hierarchical perspective by showing that predominantly higher-order areas (e.g., planum temporale and frontoparietal cortices) encode physical spatial congruency that relies on information from auditory and visual processing streams. Critically, while previous research used spatial localization tasks, in which causal inference is implicit and the spatial location of the signal is explicitly computed and mapped onto a motor response, in the current study spatial representations were not explicitly task relevant but were computed for explicit causal inference (i.e., to determine whether audiovisual signals come from a common cause). Collectively, our research suggests that frontoparietal areas play a key role in integrating auditory and visual signals into spatial representations for both (1) explicit spatial localization that involves implicit causal inference and (2) explicit causal inference (i.e., common source judgments) that requires implicit spatial localization of AV signals.

Previous studies demonstrated that the lateral prefrontal cortex (PFC) is a key convergence zone for multisensory integration (Wallace et al., 2004a; Werner and Noppeney, 2010b; Romanski, 2012); moreover, the lateral PFC has been implicated in

controlling audiovisual integration and segregation (Noppeney et al., 2010; Gau and Noppeney, 2016; Cao et al., 2019) and causal structure learning (Tomov et al., 2018). Critically, our study enabled us to identify brain regions encoding the outcome of participants' causal decisions regardless of the physical spatial correspondence cues. In line with our a priori prediction, the DLPFC was the only region where the decoding accuracy profile peaked for causal judgments. This result indicates that the lateral PFC encodes participants' explicit causal inference regardless of the physical spatial audiovisual correspondence cues or observers' motor responses. A critical question for future research is whether lateral PFC also encodes implicit causal decisions that are required to arbitrate between sensory integration and segregation in multisensory perception. For instance, future studies may use similar threshold designs in an auditory localization task. Guided by previous research showing that the lateral PFC modulates audiovisual binding in McGurk illusion trials, we expect that lateral PFC encodes observers' implicit causal decision that will then in turn influence their auditory spatial percept (Gau and Noppeney, 2016).

Moreover, given the extensive evidence for early integration in low-level sensory cortices discussed earlier, it is rather unlikely that the brain delays multisensory binding until an accumulated causal judgment made by the prefrontal cortex. On the contrary, it is more plausible that the brain integrates or segregates spatial sensory signals already at the primary cortex level and progressively refines the representations via multiple feedback loops across the cortical hierarchy (Rao and Ballard, 1999; Friston, 2005). Recent evidence is in line with such a feedback loop architecture describing (1) top-down control of multisensory representations by the prefrontal cortex (Siegel et al., 2015; Gau and Noppeney, 2016; Rohe and Noppeney, 2018), (2) the hierarchical nature of perceptual inference in the human brain (Rohe and Noppeney, 2015b, 2016), and (3) its temporal evolution involving the dynamic encoding of multiple perceptual estimates in spatial tasks (Aller and Noppeney, 2019) or nonspatial tasks (Cao et al., 2019; Rohe et al., 2019). Therefore, the causal evidence that is accumulated in the prefrontal cortex needs to be projected backward to lower-level sensory areas to inform and update their spatial representation and the binding process. Accordingly, we were able to decode causal decisional outcome also from low-level sensory cortices such as V2–3 and planum temporale, suggesting that the causal inference in the lateral PFC top-down modulates along the sensory processing hierarchy.

Importantly, we were able to decode all dimensions of our design from FEF and IPS0–4, including visual and auditory space, physical AV spatial congruency, and observers' causal decisions and motor responses. Further, our current paradigm enabled us to orthogonalize participants' motor responses with respect to their causal decisions. Even when trials were matched for causal decisions, we were able to decode participants' hand response from IPS0–4 significantly better than chance. These results suggest that IPS0–4 integrates audiovisual signals not only into spatial representations, but also transforms them into motor responses. In concordance with these findings, numerous electrophysiological studies have demonstrated that IPS can transform sensory input into motor output according to learned mappings (Cohen and Andersen, 2004; Gottlieb and Snyder, 2010; Sereno and Huang, 2014).

The sensitivity of the FEF–IPS circuitry to all experimental dimensions suggests that they integrate audiovisual signals into spatial representations informed by the explicit causal inference encoded in the lateral PFC. Our results thus extend previous

findings showing that IPS3–4 arbitrates between audiovisual integration and segregation depending on the physical correspondence cues of the sensory signals for spatial localization (Rohe and Noppeney, 2015b, 2016). They converge with recent findings that parietal cortices (e.g., lateral intraparietal cortex in macaque) might not be directly involved in evidence accumulation per se but rather related to decision formation indirectly as part of a distributed network (Katz et al., 2016). Notably, our ability to decode all information dimensions from activation patterns in frontoparietal cortices aligns well with recent suggestions that parietal cortices represent sensory, motor and potentially decision-related variables via multiplexing (Huk et al., 2017). Future neurophysiology research will need to assess whether these dimensions are encoded in distinct or overlapping neuronal populations.

In conclusion, our study was able to dissociate participants' causal inference from the physical audiovisual correspondence cues and motor responses. Our results suggest that the lateral PFC plays a key role in inferring the causal structure (i.e., the number of sources that generated the noisy audiovisual signals). Moreover, informed by the physical AV spatial congruency cues and the inferred causal structures FEF and IPS form a circuitry that integrates auditory and visual spatial signals into representations to guide behavioral (i.e., motor) response.

References

- Acerbi L, Dokka K, Angelaki DE, Ma WJ (2018) Bayesian comparison of explicit and implicit causal inference strategies in multisensory heading perception. *PLoS Comput Biol* 14:e1006110.
- Adam R, Noppeney U (2010) Prior auditory information shapes visual category-selectivity in ventral occipito-temporal cortex. *Neuroimage* 52:1592–1602.
- Alais D, Burr D (2004) The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol* 14:257–262.
- Aller M, Noppeney U (2019) To integrate or not to integrate: temporal dynamics of hierarchical Bayesian causal inference. *PLoS Biol* 17:e3000210.
- Arnott SR, Binns MA, Grady CL, Alain C (2004) Assessing the auditory dual-pathway model in humans. *Neuroimage* 22:401–408.
- Ashburner J, Friston KJ (2005) Unified segmentation. *Neuroimage* 26:839–851.
- Bertelson P, Radeau M (1981) Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Percept Psychophys* 29:578–584.
- Bishop CW, Miller LM (2011) Speech cues contribute to audiovisual spatial integration. *PLoS One* 6:e24016.
- Bonath B, Noesselt T, Martinez A, Mishra J, Schwiecker K, Heinze H-J, Hillyard SA (2007) Neural basis of the ventriloquist illusion. *Curr Biol* 17:1697–1703.
- Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10:433–436.
- Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex* 11:1110–1123.
- Cao Y, Summerfield C, Park H, Giordano BL, Kayser C (2019) Causal inference in the multisensory brain. *Neuron* 102:1076–1087.e8.
- Chang C-C, Lin C-J (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2:1–27.
- Cohen YE, Andersen RA (2004) Multisensory representations of space in the posterior parietal cortex. In: *The handbook of multisensory processes* (Calvert GA, Spence C, Stein BE, eds), pp 463–482. Cambridge, MA: MIT.
- Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9:179–194.
- Destrieux C, Fischl B, Dale A, Halgren E (2010) Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* 53:1–15.
- Driver J (1996) Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature* 381:66–68.
- Driver J, Noesselt T (2008) Multisensory interplay reveals crossmodal influences on “sensory-specific” brain regions, neural responses, and judgments. *Neuron* 57:11–23.
- Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415:429–433.
- Faisal AA, Selen LPJ, Wolpert DM (2008) Noise in the nervous system. *Nat Rev Neurosci* 9:292–303.
- Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1:1–47.
- Friston K (2005) A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* 360:815–836.
- Friston KJ, Holmes AP, Worsley KJ, Poline J-P, Frith CD, Frackowiak RSJ (1994a) Statistical parametric maps in functional imaging: a general linear approach. *Hum Brain Mapp* 2:189–210.
- Friston KJ, Worsley KJ, Frackowiak RSJ, Mazziotta JC, Evans AC (1994b) Assessing the significance of focal activations using their spatial extent. *Hum Brain Mapp* 1:210–220.
- Friston KJ, Holmes AP, Price CJ, Büchel C, Worsley KJ (1999) Multisubject fMRI studies and conjunction analyses. *Neuroimage* 10:385–396.
- Gau R, Noppeney U (2016) How prior expectations shape multisensory perception. *Neuroimage* 124:876–886.
- Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? *Trends Cogn Sci* 10:278–285.
- Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK (2005) Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci* 25:5004–5012.
- Gottlieb J, Snyder LH (2010) Spatial and non-spatial functions of the parietal cortex. *Curr Opin Neurobiol* 20:731–740.
- Hein G, Doehrmann O, Müller NG, Kaiser J, Muckli L, Naumer MJ (2007) Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *J Neurosci* 27:7881–7887.
- Huk AC, Katz LN, Yates JL (2017) The role of the lateral intraparietal area in (the study of) decision making. *Annu Rev Neurosci* 40:349–372.
- Hunt DL, Yamoah EN, Krubitzer L (2006) Multisensory plasticity in congenitally deaf mice: how are cortical areas functionally specified? *Neuroscience* 139:1507–1524.
- Kaas JH, Collins CE (2004) The resurrection of multisensory cortex in primates: connection patterns that integrates modalities. In: *The handbook of multisensory processes* (Calvert GA, Spence C, Stein BE, eds), pp 285–294. Cambridge, MA: MIT.
- Katz LN, Yates JL, Pillow JW, Huk AC (2016) Dissociated functional significance of decision-related activity in the primate dorsal stream. *Nature* 535:285–288.
- Kayser C, Logothetis NK (2007) Do early sensory cortices integrate cross-modal information? *Brain Struct Funct* 212:121–132.
- Kingdom FAA, Prins N (2010) *Psychophysics: a practical introduction*. Amsterdam: Elsevier.
- Kleiner M, Brainard DH, Pelli DG (2007) What's new in psychtoolbox-3? *Perception* 36:1–16.
- Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L (2007) Causal inference in multisensory perception. *PLoS One* 2:e943.
- Lakatos P, Chen C-M, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53:279–292.
- Lee H, Noppeney U (2011a) Physical and perceptual factors shape the neural mechanisms that integrate audiovisual signals in speech comprehension. *J Neurosci* 31:11338–11350.
- Lee H, Noppeney U (2011b) Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proc Natl Acad Sci U S A* 108:E1441–E1450.
- Lewald J, Guski R (2003) Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Brain Res Cogn Brain Res* 16:468–478.
- Lewis R, Noppeney U (2010) Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. *J Neurosci* 30:12329–12339.
- Magnotti JF, Ma WJ, Beauchamp MS (2013) Causal inference of asynchronous audiovisual speech. *Front Psychol* 4:798.
- Maier JX, Di Luca M, Noppeney U (2011) Audiovisual asynchrony detection in human speech. *J Exp Psychol Hum Percept Perform* 37:245–256.
- Meijer D, Veselić S, Calafiore C, Noppeney U (2019) Integration of audiovisual spatial signals is not consistent with maximum likelihood estimation. *Cortex* 119:74–88.
- Mesulam M (1998) From sensation to cognition. *Brain* 121:1013–1052.

- Moerel M, De Martino F, Formisano E (2014) An anatomical and functional topography of human auditory cortical areas. *Front Neurosci* 8:225.
- Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res* 14:115–128.
- Munhall KG, Gribble P, Sacco L, Ward M (1996) Temporal constraints on the McGurk effect. *Percept Psychophys* 58:351–362.
- Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp* 15:1–25.
- Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze H-J, Driver J (2007) Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *J Neurosci* 27:11431–11441.
- Noppeney U (2020) Multisensory perception: behavior, computations and neural mechanisms. In: *The cognitive neurosciences*, Ed 6 (Poeppel D, Mangun GR, Gazzaniga MS, eds), pp 141–150. Cambridge, MA: MIT.
- Noppeney U, Josephs O, Hocking J, Price CJ, Friston KJ (2008) The effect of prior visual information on recognition of speech and sounds. *Cereb Cortex* 18:598–609.
- Noppeney U, Ostwald D, Werner S (2010) Perceptual decisions formed by accumulation of audiovisual evidence in prefrontal cortex. *J Neurosci* 30:7434–7446.
- Parise CV, Ernst MO (2016) Correlation detection as a general mechanism for multisensory integration. *Nat Commun* 7:11543.
- Parise CV, Spence C (2009) “When birds of a feather flock together”: synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS One* 4:e5664.
- Parise CV, Spence C, Ernst MO (2012) When correlation implies causation in multisensory integration. *Curr Biol* 22:46–49.
- Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10:437–442.
- Rajkowska G, Goldman-Rakic PS (1995) Cytoarchitectonic definition of prefrontal areas in the normal human cortex: II. Variability in locations of areas 9 and 46 and relationship to the Talairach coordinate system. *Cereb Cortex* 5:323–337.
- Rao RPN, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2:79–87.
- Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci* 12:718–724.
- Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc Natl Acad Sci U S A* 97:11800–11806.
- Recanzone GH (2009) Interactions of auditory and visual stimuli in space and time. *Hear Res* 258:89–99.
- Recanzone GH, Cohen YE (2010) Serial and parallel processing in the primate auditory cortex revisited. *Behav Brain Res* 206:1–7.
- Rohe T, Noppeney U (2015a) Sensory reliability shapes perceptual inference via two mechanisms. *J Vis* 15(5):22, 1–16.
- Rohe T, Noppeney U (2015b) Cortical hierarchies perform Bayesian causal inference in multisensory perception. *PLoS Biol* 13:e1002073.
- Rohe T, Noppeney U (2016) Distinct computational principles govern multisensory integration in primary sensory and association cortices. *Curr Biol* 26:509–514.
- Rohe T, Noppeney U (2018) Reliability-weighted integration of audiovisual signals can be modulated by top-down attention. *eNeuro* 5:ENEURO.0315-17.2018.
- Rohe T, Ehlis A-C, Noppeney U (2019) The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. *Nat Commun* 10:1907.
- Romanski LM (2012) Convergence of auditory, visual, and somatosensory information in ventral prefrontal cortex. In: *The neural bases of multisensory processes* (Murray M, Wallace M, eds), pp 667–682. Boca Raton, FL: CRC.
- Sadaghiani S, Maier JX, Noppeney U (2009) Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing. *J Neurosci* 29:6490–6499.
- Senkowski D, Talsma D, Herrmann CS, Woldorff MG (2005) Multisensory processing and oscillatory gamma responses: effects of spatial selective attention. *Exp Brain Res* 166:411–426.
- Sereno MI, Huang R-S (2014) Multisensory maps in parietal cortex. *Curr Opin Neurobiol* 24:39–46.
- Sereno MI, Dale A, Reppas J, Kwong K, Belliveau J, Brady T, Rosen B, Tootell R (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268:889–893.
- Shams L, Beierholm UR (2010) Causal inference in perception. *Trends Cogn Sci* 14:425–432.
- Siegel M, Buschman TJ, Miller EK (2015) Cortical information flow during flexible sensorimotor decisions. *Science* 348:1352–1355.
- Silver MA, Kastner S (2009) Topographic maps in human frontal and parietal cortex. *Trends Cogn Sci* 13:488–495.
- Slutsky DA, Recanzone GH (2001) Temporal and spatial dependency of the ventriloquist effect. *Neuroreport* 12:7–10.
- Swisher JD, Halko MA, Merabet LB, McMains SA, Somers DC (2007) Visual topography of human intraparietal sulcus. *J Neurosci* 27:5326–5337.
- Thirion B, Pinel P, Mériaux S, Roche A, Dehaene S, Poline J-B (2007) Analysis of a large fMRI cohort: statistical and methodological issues for group analyses. *Neuroimage* 35:105–120.
- Tomov MS, Dorfman HM, Gershman SJ (2018) Neural computations underlying causal structure learning. *J Neurosci* 38:7143–7157.
- van Atteveldt NM, Formisano E, Goebel R, Blomert L (2007) Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *Neuroimage* 36:1345–1360.
- van Wassenhove V, Grant KW, Poeppel D (2007) Temporal window of integration in auditory-visual speech perception. *Neuropsychologia* 45:598–607.
- von Saldern S, Noppeney U (2013) Sensory and striatal areas integrate auditory and visual signals into behavioral benefits during motion discrimination. *J Neurosci* 33:8841–8849.
- Wallace MT, Ramachandran R, Stein BE (2004a) A revised view of sensory cortical parcellation. *Proc Natl Acad Sci U S A* 101:2167–2172.
- Wallace MT, Roberson GE, Hairston WD, Stein BE, Vaughan JW, Schirillo JA (2004b) Unifying multisensory signals across time and space. *Exp Brain Res* 158:252–258.
- Wang L, Mruczek REB, Arcaro MJ, Kastner S (2015) Probabilistic maps of visual topography in human cortex. *Cereb Cortex* 25:3911–3931.
- Warren JD, Griffiths TD (2003) Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. *J Neurosci* 23:5799–5804.
- Welch RB (1999) Meaning, attention, and the “unity assumption” in the intersensory bias of spatial and temporal perceptions. In: *Advances in psychology*. (Aschersleben G, Bachmann T, Müsseler J, eds), pp 371–387. Amsterdam: Elsevier.
- Werner S, Noppeney U (2010a) Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cereb Cortex* 20:1829–1842.
- Werner S, Noppeney U (2010b) Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *J Neurosci* 30:2662–2675.
- Werner S, Noppeney U (2011) The contributions of transient and sustained response codes to audiovisual integration. *Cereb Cortex* 21:920–931.
- Yendiki A, Greve DN, Wallace S, Vangel M, Bockholt J, Mueller BA, Magnotta V, Andreasen N, Manoach DS, Gollub RL (2010) Multi-site characterization of an fMRI working memory paradigm: reliability of activation indices. *Neuroimage* 53:119–131.