JNeurosci
THE JOURNAL OF NEUROSCIENCE

Research Report: Regular Manuscript

Causal inference in audiovisual perception

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

Cite as: J. Neurosci 2020; 10.1523/JNEUROSCI.0051-20.2020

Received: 7 January 2020
Revised: 26 June 2020
Accepted: 1 July 2020

Causal inference in audiovisual perception

Agoston Mihalik[1,2] and Uta Noppeney[1,3]

[1]Computational Neuroscience and Cognitive Robotics Centre, University of Birmingham, B15 2TT Birmingham, United Kingdom

[2]Centre for Medical Image Computing, Department of Computer Science, University College London, WC1V 6LJ London, United Kingdom

[3]Donders Institute for Brain, Cognition and Behaviour, Radboud University, 6525 Nijmegen, Netherlands

Abbreviated title: Causal inference in AV perception

Corresponding author: Agoston Mihalik MD, PhD (axm676@alumni.bham.ac.uk)

Number of pages: 45

Number of figures: 4

Number of tables: 3

Number of words for abstract: 246

Number of words for significant statement: 106

Number of words for introduction: 646

Number of words for discussion: 1500

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.

Conflict of interest: The authors declare no competing interests.

Acknowledgements: This work was supported by the European Research Council (ERC-2012-StG_20111109 multsens).

Abstract

In our natural environment the 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 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 same (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 signals' audiovisual disparity 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.

Significance statement

In our natural environment our senses are continuously flooded with myriads 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 (fMRI) study shows that the lateral prefrontal cortex plays a key role in inferring the environment's causal structure. Crucially, informed by the spatial correspondence cues and the inferred causal structure FEF and IPS 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 Guski, 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 towards the visual signal location and vice versa depending on the relative auditory and visual reliabilities (Bertelson and Radeau, 1981; Driver, 1996; Ernst and Banks, 2002; Alais and Burr, 2004; Bonath et al., 2007; Meijer et al., 2019). Crucially, these crossmodal 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, fMRI, MEG and EEG 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 (50–100ms) neural processes in primary sensory areas encoded predominantly

the spatial locations independently for auditory and visual signals, while later processes (100–200ms) in posterior parietal cortices (IPS1–2) formed spatial representations by combining audiovisual signals. Critically, only at the top of the hierarchy in anterior parietal cortices (IPS3–4, 350–450ms) 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 world's causal structure.

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 e.g. 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 approximately 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; 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 multi-day psychophysics-fMRI study is low compared to 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 prior to the fMRI experiment based on the following criteria: i. no history of neurological or psychiatric illness; ii. normal or corrected-to-normal vision; iii. reported normal hearing; iv. 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 v. 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 hour) the sound stimuli were recorded in an anechoic chamber and participants' sound localization performance were assessed. On day 2 and 3 (~2 hours in total), participants were trained to determine the subject-specific AV spatial disparities in a mock scanner. On day 4 (~1 hour) participants performed a standard retinotopic localizer task for the retinotopic mapping of visual and parietal cortical areas. On day 5 and 6 (~3 hours 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 grey background (70% contrast) and were presented for 50 ms. The horizontal standard deviation of the Gaussian was set to a 5° visual angle, and the vertical standard deviation 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™, Inc. (USA) in-ear binaural microphones in an anechoic chamber in the School of Psychology, 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°, 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 '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 standard deviation, STD) 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 $\mathrm{PSE/STD\ ratio} < 0.3$ (i.e. inclusion criterion iv).

Adjustment of spatial disparity and assessment of sound localization – mock scanner (day 2 and 3)

We adjusted the audiovisual spatial disparity inside the mock scanner individually for each subject to obtain an accuracy of ~70% on the main causal inference task (i.e. common vs. separate causes). This individual adjustment of AV spatial disparity allowed us to compare 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 5 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) which 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 standard deviation of the AV disparity computed over these reversal was $< 2°$ visual angles (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 (8.1º visual angles ± 1.2 SEM 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 1–2° visual angles across runs. Participants were included in the fMRI study if their performance accuracy for the individually selected AV disparity (between 4°–16° visual angle) was between 60–80% (i.e. inclusion criterion v). This criterion is required to ensure 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 day 2 and 3, the sound localization performance was further assessed based on a left-right classification task with 2 selected stimulus locations. Typically, 20–60 repetitions per stimulus location were performed in the mock scanner. Unbiased sound localization was defined as less than 30% difference in the accuracy for left and right-side stimuli (i.e. inclusion criterion iv).

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 1–3 runs with constant disparity inside the scanner prior to the main causal inference fMRI experiment. Similarly to the mock scanner, the sound localization performance was finally assessed in the scanner using a left-right classification task for 2 selected stimulus locations (see inclusion criterion iv). 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 SEM 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° visual angle with respect to a central fixation cross; Figure 1A). This resulted in four trial types: i. AV spatially congruent left (i.e. A and V at same location), ii. AV spatially congruent right, iii. AV spatially incongruent with A left and V right and iv. 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 x 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 (~8% of trials). To increase 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: i. visual stimulus location (left vs. right); ii. auditory stimulus location (left vs. right); iii. motor response (left vs. right hand) (Figure 1B). Based on these experimental manipulations, participants' causal decisions and motor responses we characterized the functional properties of brain regions according to the following encoding dimensions: i. visual space (i.e. V left vs. right); ii. auditory space (i.e. A left vs. right); iii. spatial (i.e. physical) congruency (i.e. AV spatially congruent vs. incongruent); iv. observers' causal inference (i.e. causal decision: common vs. separate causes) and v. 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 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 (~35° horizontally and ~14° 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°/s, acceleration threshold = 8000°/s^2, motion threshold = 0.15°, radial amplitude > 1°) and eye blinks were quantified (0–875 ms after stimulus onset). Critically, the 2 (visual left, right) x 2 (auditory left, right) repeated measures ANOVAs on the stimulus conditions performed separately for i. % saccades or ii. % eye blinks revealed no significant main effects or interactions indicating that differences in BOLD-response between conditions are unlikely to be due to eye movement confounds.

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 Inc.) on a MacBook Pro (Mac OSX 10.6.8). For the main task, visual stimuli were back projected to a Plexiglas screen using a D-ILA projector (JVC DLA-SX21) visible to the participant through a mirror mounted on the magnetic resonance (MR) head coil. Auditory stimuli were delivered via Sennheiser HD 280 Pro (in the anechoic chamber), Sennheiser HD 219 (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 Ltd.) at a sampling rate of 1000 Hz.

253 ***MRI data acquisition***

254 A 3T Philips Achieva scanner was used to acquire both T1-weighted anatomical images
255 (TR/TE/TI, 8.4/3.8/min. 540 ms; 175 slices; image matrix, 288 x 232; spatial resolution, 1 x 1
256 x 1 mm^3 voxels) and T2*-weighted echo-planar images (EPI) with blood oxygenation level-
257 dependent (BOLD) contrast (fast field echo; TR/TE, 2600/40 ms; 38 axial slices acquired in
258 ascending direction; image matrix, 80 x 80; spatial resolution, 3 x 3 x 3 mm^3 voxels without
259 gap). Typically, there were 10–12 runs with 240 volumes per run over 2 sessions. The first 4
260 volumes were not acquired to allow T1 equilibration effects. In one participant, we repeated a
261 session, since the participant's accuracy was 15% lower than the mean accuracy of the
262 remaining sessions. In another participant, 2 runs were excluded due to technical problems
263 with the setup. In 3 participants, 1–2 runs were removed from further analysis to be able to
264 counterbalance the left vs. right response hands across runs (see section experimental design).

265 ***Statistical analysis***

266 *Behavioural data analysis*

267 For the eye movement analysis of the mock scanner data, i. % saccades and ii. % eye blinks
268 of the participants were entered into separate 2 (visual: left, right) x 2 (auditory: left, right)
269 repeated-measures ANOVAs.

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

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

274 *fMRI data pre-processing*

275 The data were analysed with statistical parametric mapping (SPM8; Wellcome Trust Centre
276 for Neuroimaging, London, UK; http://www.fil.ion.ucl.ac.uk/spm/; Friston, Holmes,

Worsley, et al., 1995) 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 x 2 x 2 mm^3 voxels, and spatially smoothed with a Gaussian kernel of 6 mm FWHM. For the multivariate decoding analysis, the EPI images were analysed in native participant space and spatially smoothed with a Gaussian kernel of 3 mm FWHM. For the retinotopic analysis, the data were analysed in native space and without additional smoothing.

fMRI data analysis

Data were modelled 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 artefacts.

Univariate fMRI analysis: For the conventional univariate analysis, the general linear model (GLM) modelled the 16 conditions in our 2 (visual: left, right) x 2 (auditory: left, right) x 2 (decisional outcome: common, separate causes) x 2 (hand response: left, right) factorial design. Condition-specific effects for each participant were estimated according to the general linear model 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) (Figure 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 modelled exactly 8 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 7 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 general linear model 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) x 2 (visual: left vs. right) x 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 MVPA) 10 times and averaged the decoding accuracy across those 10 iterations.

This subsampling and matching procedure ensured that the parameter estimates for common vs. 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 response 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 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 8 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 x 10 runs = 280 parameter estimate images in total (see MRI data acquisition for details). The number of parameter estimate images for decoding 'causal decisions' or '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 data set and images from the 'left-out run' were assigned to the test set. Parameter estimate images for training and test data sets 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 responses patterns to the class labels from all but one run according to the following dimensions: i. visual signal location (left vs. right); ii. auditory signal location (left vs. right); iii. physical AV spatial congruency (congruent vs. incongruent); iv. causal decisional outcome (common vs. separate causes); and v. motor response (left vs. right hand). The model then used this learnt 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 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.

Non-parametric 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 auto-correlation structure) and calculating the decoding accuracies for all permutations (500 permutations x 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 +/- 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 non-parametric permutation testing as described above: briefly, we i. computed the deviations from chance decoding accuracy for each of the five information dimensions individually for each participant, ii. calculated the differences in these relative decoding accuracies between information dimensions for each participant (e.g. causal decision minus physical spatial congruency) and iii. 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 re-computed 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 regions of interest. 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. $0.05/11\ \mathrm{ROI} = 0.0045$. In Table 3, we also report the number of subjects that were individually

significant (i.e. uncorrected $p < 0.05$) based on within subject permutation test (in brackets we list the number of subjects which were significant after Bonferroni correction for multiple comparisons across the 11 regions of interest, 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 familywise error rate is < 0.05 (i.e. $0.002 * 11 = 0.022$ and $0.00098 * 11 = 0.01$, respectively). Guided by a priori hypotheses we did not apply Bonferroni correction for testing: visual left/right location in V1, V2, V3, V3AB; auditory left/right location in A1, PT; motor left/right hand response in 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 (Sereno 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 pre-processing steps (see fMRI analysis: data pre-processing), visual responses were modelled by entering a sine and cosine convolved with the hemodynamic response function as regressors into a general linear model. 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 5.3.0 (Dale et al., 1999). Visual

426 regions V1–V3, V3AB, and parietal regions IPS0–4 were defined as phase reversal in angular
427 retinotopic maps. IPS0–4 were defined as contiguous, approximately rectangular regions
428 based on phase reversals along the anatomical IPS (Swisher et al., 2007) and guided by
429 group-level retinotopic probabilistic maps (Wang et al., 2015).

430 *Region of interests used for decoding analysis*

431 For the decoding analyses, all regions of interest (ROI) were combined from the left and right
432 hemispheres.

433 Occipital, parietal and FEF regions: Regions in the occipital and parietal cortices were
434 defined based on retinotopic mapping as described above. The frontal eye-field (FEF) was
435 defined by an inverse normalized group-level retinotopic probabilistic map (Wang et al.,
436 2015). The resulting subject-level probabilistic map was thresholded at the 80 percentile and
437 any overlap with the motor cortex was removed.

438 Auditory, motor and prefrontal regions: These regions were based on labels of the Destrieux
439 atlas of Freesurfer 5.3.0 (Dale et al., 1999; Destrieux et al., 2010). The primary auditory
440 cortex was defined as the anterior transverse temporal gyrus (Heschl's gyrus). The higher
441 auditory cortex was formed by merging the transverse temporal sulcus and the planum
442 temporale (PT). The motor cortex was based on the precentral gyrus. The dorsolateral
443 prefrontal cortex (DLPFC) was defined by combining the superior and middle frontal gyri
444 and sulci as previously described (Yendiki et al., 2010). In line with (Rajkowska and
445 Goldman-Rakic, 1995) we limited the superior frontal gyrus and sulcus to Talairach
446 coordinates $y = 26$ and $y = 53$, respectively, and the middle frontal gyrus and sulcus to
447 Talairach coordinates $y = 20$ and $y = 50$, respectively.

Results

Behavioural 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 about 70% correct when deciding whether auditory and visual signals originated from common or independent causes with a small bias towards common causes decisions ($\text{accuracy}_{\text{SC}} = 77 \pm 1.7\%$, $\text{accuracy}_{\text{SI}} = 66 \pm 2.2\%$ with the index SC and SI for physically spatially congruent and incongruent; d': 1.07 ± 0.12; bias: 0.16 ± 0.03; and mean ± SEM in all cases).

A 2 (physical: spatially congruent, incongruent) x 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 x 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' ($\text{RT}_{\text{SC,DC}} = 0.89 \pm 0.05$ s; $\text{RT}_{\text{SI,DS}} = 0.93 \pm 0.06$ s ; $\text{RT}_{\text{SC,DS}} = 1.02 \pm 0.06$ s; $\text{RT}_{\text{SI,DC}} = 0.96 \pm 0.06$ s; with the index SC and SI for physically spatially congruent and incongruent, DC and DS for common and separate cause decisions, respectively). In other words, observers were faster on their correct than 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' judgments pooled over correct and incorrect responses, i.e. both 'common cause' and 'separate cause' judgments included

correct and incorrect trials. Hence, our decoding focused on decisional outcome irrespective 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 (Figure 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, 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.

Moreover, we observed the expected lateralization effects for motor responses: left relative to right hand responses were associated with greater activations in the right pre- and postcentral gyri, whilst right relative to left hand responses were associated with greater activations in the left pre- 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 (Figure 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 behavioural 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 vs. 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; Figure 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: i. visual signal location (left vs. right); ii. auditory signal location (left vs. right); iii. physical spatial congruency (congruent vs. incongruent); iv. causal decisional outcome (common vs. separate causes); and v. motor response (left vs. right hand) (Figure 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

522 BOLD-response patterns that form the basis for multivariate pattern classification separate
523 for class 1 and 2 in each region in Figure 4.

524 *Decoding of auditory and visual location*

525 Visual location could be decoded significantly better than chance from BOLD-response
526 patterns in visual areas including V1, V2, V3 and V3AB (Figure 3). In addition, visual
527 location was represented in the parietal cortex (IPS0–4) as well as in the frontal eye-fields
528 (FEF) which is consistent with the well-established retinotopic organization of those cortical
529 regions (Swisher et al., 2007; Silver and Kastner, 2009; Wang et al., 2015). Auditory location
530 could be decoded significantly better than chance from the planum temporale (PT) as a higher
531 order auditory area previously implicated in spatial processing (Rauschecker and Tian, 2000;
532 Warren and Griffiths, 2003; Moerel et al., 2014) as well as along the dorsal auditory
533 processing stream including the posterior parietal cortex (IPS0–2), the frontal eye-fields
534 (FEF), and the dorsolateral prefrontal cortex (DLPFC) (Rauschecker and Tian, 2000; Arnott
535 et al., 2004; Rauschecker and Scott, 2009; Recanzone and Cohen, 2010) (Figure 3).

536 *Decoding of physical AV spatial congruency and observers' causal decision*

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

543 Physical AV spatial congruency could be decoded from higher order association
544 cortices encompassing the parietal cortex (IPS0–4), the FEF and DLPFC as well as the
545 planum temporale (Figure 3). These results are consistent with the classical view of
546 multisensory processing in which primary auditory and visual cortices are specialized for

processing signals of their preferred sensory modality and higher order fronto-parietal 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 (resp. incongruent) stimuli that were judged as coming from one common vs. separate causes. In other words, the individual threshold adjustment allowed us to identify regions encoding participants' causal decisions irrespective of the physical spatial congruency of the underlying AV signals (see methods about additional subsampling and matching procedures). In line with our predictions, participants' causal decisional outcome could be decoded from DLPFC (Figure 3). Critically, observers' causal decision 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, auditor, 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 (Figure 3).

Given the significant interaction between causal decision and spatial disparity in our behavioural 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 ± SEM, group-level permutation test: $p = 0.001$ uncorrected) for spatially congruent (SC) signals and 58.72 ±

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

For completeness, we also assessed the decoding accuracies for i. IPS0–2: 56.40 ± 1.27 for SC (group mean ± SEM, group-level permutation test: $p = 0.003$ uncorrected) and 55.60 ± 1.35 for SI (group mean ± SEM, group-level permutation test: $p = 0.002$ uncorrected); ii. IPS3–4: 55.37 ± 1.84 for SC (group mean ± SEM, group-level permutation test: $p = 0.013$ uncorrected) and 55.09 ± 1.35 for SI (group mean ± SEM, group-level permutation test: $p = 0.003$ uncorrected); iii. FEF: 58.14 ± 1.35 for SC (group mean ± SEM, group-level permutation test: $p = 0.002$ uncorrected) and 55.98 ± 0.98 for SI (group mean ± 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 response (i.e. left 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 (Figure 3). In addition, we were able to decode observers' motor response from the FEF, IPS0–4 and V3AB. Further, we were able to decode participants' motor response from planum temporale and Heschl's gyrus. The latter decoding of sensory-motor information from activation patterns in Heschl's gyrus may potentially be attributed to activations from the neighbouring 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

596 brain infers whether or not signals originate from a common cause or event based on multiple
597 correspondence cues such as spatial disparity (Slutsky and Recanzone, 2001; Lewald and
598 Guski, 2003; Wallace et al., 2004b; Recanzone, 2009), temporal synchrony (Munhall et al.,
599 1996; Noesselt et al., 2007; van Wassenhove et al., 2007; Lewis and Noppeney, 2010; Lee
600 and Noppeney, 2011b; Maier et al., 2011; Parise et al., 2012; Magnotti et al., 2013; Parise
601 and Ernst, 2016) or semantic and other higher order correspondence cues (Welch, 1999;
602 Parise and Spence, 2009; Sadaghiani et al., 2009; Adam and Noppeney, 2010; Noppeney et
603 al., 2010; Bishop and Miller, 2011; Lee and Noppeney, 2011a). As a result, observers' causal
604 decisions have previously been inherently correlated with the congruency of the audiovisual
605 signals (Rohe and Noppeney, 2015b, 2016; Aller and Noppeney, 2019; Cao et al., 2019;
606 Rohe et al., 2019) making it challenging to dissociate observers' causal decisions from the
607 underlying physical correspondence cues such as audiovisual spatial disparity.

608 To dissociate the neural processes associated with participants' causal decisions from
609 those driven by the physical AV spatial congruency cues we adjusted the audiovisual spatial
610 disparity individually for each participant to enable a threshold accuracy of 70%. As a result
611 of external and internal noise (Faisal et al., 2008) spatially congruent audiovisual signals
612 were perceived as coming from the same source in ~70% of cases. Conversely, spatially
613 disparate audiovisual signals were perceived as coming from independent sources in ~70% of
614 cases. This causal uncertainty allowed us to select and compare physically identical
615 audiovisual signals that were perceived as coming from common or separate causes.
616 Moreover, we dissociated participants' causal decisions from their motor responses by
617 counterbalancing the mapping between causal decision (i.e. common vs. separate causes) and
618 motor response (i.e. left vs. right hand) over runs. In summary, our experimental design
619 enabled us to characterize a system of brain regions with respect to five different 'encoding
620 dimensions': i. visual space (left vs. right); ii. auditory space (left vs. right); iii. physical

621 spatial congruency (congruent vs. incongruent); iv. causal inference and decision (common
622 vs. separate causes); and v. motor response (left vs. right hand).

623 Unsurprisingly, our multivariate decoding results demonstrate that low level visual
624 areas (V1–3) encode predominantly visual space, planum temporale (PT) auditory space and
625 precentral gyrus participant's motor responses. Physical spatial congruency could be decoded
626 from planum temporale, all parietal areas (IPS0–4) and prefrontal cortices (DLPFC, FEF).
627 This profile of results is consistent with the classical hierarchical organization of
628 multisensory perception, according to which low level sensory cortices process signals
629 mainly from their preferred sensory modalities and higher order cortical regions combine
630 signals across the senses (Felleman and Van Essen, 1991; Mesulam, 1998; Calvert, 2001;
631 Kaas and Collins, 2004; Wallace et al., 2004a). This view has been challenged by studies
632 showing multisensory interactions already at the primary cortical level (Molholm et al., 2002;
633 Ghazanfar, 2005; Senkowski et al., 2005; Ghazanfar and Schroeder, 2006; Hunt et al., 2006;
634 Kayser and Logothetis, 2007; Lakatos et al., 2007; Driver and Noesselt, 2008; Werner and
635 Noppeney, 2011). However, in primary sensory cortices stimuli from the non-preferred
636 sensory modality typically modulated the response magnitude or salience rather than spatial
637 representation of stimuli from the preferred sensory modality. Likewise, previous
638 multivariate pattern analyses showed that a synchronous yet displaced auditory signal had
639 minimal impact on the spatial representations in primary visual cortices (e.g. Rohe and
640 Noppeney, 2015b, 2016). Only later in the processing hierarchy in posterior and anterior
641 parietal cortices were spatial representations formed that integrated auditory and visual
642 signals weighted by their bottom-up reliabilities (ISP0–4) and top-down task-relevance
643 (IPS3–4) (Rohe and Noppeney, 2015b, 2016, 2018; Aller and Noppeney, 2019). Our current
644 findings thus lend further support for this hierarchical perspective by showing that
645 predominantly higher order areas (e.g. planum temporale and frontoparietal cortices) encode

646 physical spatial congruency that relies on information from auditory and visual processing
647 streams. Critically, while previous research used spatial localization tasks, in which causal
648 inference is implicit and the signal's spatial location is explicitly computed and mapped onto
649 a motor response, in the current study spatial representations were not explicitly task-relevant
650 but computed for explicit causal inference, i.e. to determine whether audiovisual signals
651 come from a common cause. Collectively, our research suggests that fronto-parietal areas
652 play a key role in integrating auditory and visual signals into spatial representations for both
653 i. explicit spatial localization that involves implicit causal inference and ii. explicit causal
654 inference (i.e. common source judgments) that requires implicit spatial localization of AV
655 signals.

656 Previous studies demonstrated that the lateral prefrontal cortex (lateral PFC) is a key
657 convergence zone for multisensory integration (Wallace et al., 2004a; Werner and Noppeney,
658 2010b; Romanski, 2012), moreover, the lateral PFC has been implicated in controlling
659 audiovisual integration and segregation (Noppeney et al., 2010; Gau and Noppeney, 2016;
660 Cao et al., 2019) and causal structure learning (Tomov et al., 2018). Critically, our study
661 enabled us to identify brain regions encoding the outcome of participants' causal decisions
662 irrespective of the physical spatial correspondence cues. In line with our a priori prediction,
663 the DLPFC was the only region where the decoding accuracy profile peaked for causal
664 judgements. This result indicates that the lateral PFC encodes participants' explicit causal
665 inference irrespective of the physical spatial audiovisual correspondence cues or observers'
666 motor response. A critical question for future research is whether lateral PFC also encodes
667 implicit causal decisions that are required to arbitrate between sensory integration and
668 segregation in multisensory perception. For instance, future studies may utilise similar
669 threshold designs in an auditory localization task. Guided by previous research showing that
670 the lateral PFC modulates audiovisual binding in McGurk illusion trials we expect that lateral

671 prefrontal cortex encodes observers implicit causal decision that will then in turn influence
672 their auditory spatial percept (Gau and Noppeney, 2016).

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

690 Importantly, we were able to decode all dimensions of our design from the frontal
691 eye-field (FEF) and the intraparietal sulcus (IPS0–4) including visual and auditory space,
692 physical AV spatial congruency, observers' causal decisions and motor responses. Further,
693 our current paradigm enabled us to orthogonalize participants' motor responses with respect
694 to their causal decisions. Even when trials were matched for causal decisions, we were able to
695 decode participants' hand response from IPS0–4 significantly better than chance. These

696 results suggest that IPS0–4 integrates audiovisual signals not only into spatial representations,
697 but it also transforms them into motor responses. In concordance with these findings,
698 numerous electrophysiological studies have demonstrated that IPS can transform sensory
699 input into motor output according to learnt mappings (Cohen and Andersen, 2004; Gottlieb
700 and Snyder, 2010; Sereno and Huang, 2014).

701 The sensitivity of the FEF–IPS circuitry to all experimental dimensions suggests that
702 they integrate audiovisual signals into spatial representations informed by the explicit causal
703 inference encoded in the lateral PFC. Our results thus extend previous findings showing that
704 IPS3–4 arbitrates between audiovisual integration and segregation depending on the physical
705 correspondence cues of the sensory signals for spatial localization (Rohe and Noppeney,
706 2015b, 2016). They converge with recent findings that parietal cortices (e.g. LIP in macaque)
707 might not be directly involved in evidence accumulation per se but rather related to decision
708 formation indirectly as part of a distributed network (Katz et al., 2016). Notably, our ability
709 to decode all information dimensions from activation patterns in fronto-parietal cortices
710 aligns well with recent suggestions that parietal cortices represent sensory, motor and
711 potentially decision-related variables via multiplexing (Huk et al., 2017). Future
712 neurophysiology research will need to assess whether these dimensions are encoded in
713 distinct or overlapping neuronal populations.

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

746 Intell Syst Technol 2:27.
747 Cohen YE, Andersen RA (2004) Multisensory representations of space in the posterior
748 parietal cortex. In: The handbook of mulitsensory processes (Calvert GA, Spence C,
749 Stein BE, eds), pp 463–482. Cambridge, MA: MIT Press.
750 Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and
751 surface reconstruction. Neuroimage 9:179–194.
752 Destrieux C, Fischl B, Dale A, Halgren E (2010) Automatic parcellation of human cortical
753 gyri and sulci using standard anatomical nomenclature. Neuroimage 53:1–15.
754 Driver J (1996) Enhancement of selective listening by illusory mislocation of speech sounds
755 due to lip-reading. Nature 381:66–68.
756 Driver J, Noesselt T (2008) Multisensory interplay reveals crossmodal influences on
757 “sensory-specific” brain regions, neural responses, and judgments. Neuron 57:11–23.
758 Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically
759 optimal fashion. Nature 415:429–433.
760 Faisal AA, Selen LPJ, Wolpert DM (2008) Noise in the nervous system. Nat Rev Neurosci
761 9:292–303.
762 Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate
763 cerebral cortex. Cereb Cortex 1:1–47.
764 Friston K (2005) A theory of cortical responses. Philos Trans R Soc B 360:815–836.
765 Friston KJ, Holmes AP, Price CJ, Büchel C, Worsley KJ (1999) Multisubject fMRI studies
766 and conjunction analyses. Neuroimage 10:385–396.
767 Friston KJ, Holmes AP, Worsley KJ, Poline J-P, Frith CD, Frackowiak RSJ (1994a)
768 Statistical parametric maps in functional imaging: a general linear approach. Hum Brain
769 Mapp 2:189–210.
770 Friston KJ, Worsley KJ, Frackowiak RSJ, Mazziotta JC, Evans AC (1994b) Assessing the

771 significance of focal activations using their spatial extent. Hum Brain Mapp 1:210–220.
772 Gau R, Noppeney U (2016) How prior expectations shape multisensory perception.
773 Neuroimage 124:876–886.
774 Ghazanfar AA (2005) Multisensory integration of dynamic faces and voices in rhesus
775 monkey auditory cortex. J Neurosci 25:5004–5012.
776 Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? Trends Cogn Sci
777 10:278–285.
778 Gottlieb J, Snyder LH (2010) Spatial and non-spatial functions of the parietal cortex. Curr
779 Opin Neurobiol 20:731–740.
780 Hein G, Doehrmann O, Muller NG, Kaiser J, Muckli L, Naumer MJ (2007) Object familiarity
781 and semantic congruency modulate responses in cortical audiovisual integration areas. J
782 Neurosci 27:7881–7887.
783 Huk AC, Katz LN, Yates JL (2017) The role of the lateral intraparietal area in (the study of)
784 decision making. Annu Rev Neurosci 40:349–372.
785 Hunt DL, Yamoah EN, Krubitzer L (2006) Multisensory plasticity in congenitally deaf mice:
786 how are cortical areas functionally specified? Neuroscience 139:1507–1524.
787 Kaas JH, Collins CE (2004) The resurrection of multisensory cortex in primates: connection
788 patterns that integrates modalities. In: The handbook of multisensory processes (Calvert
789 GA, Spence C, Stein BE, eds), pp 285–294. Cambridge, MA: MIT Press.
790 Katz LN, Yates JL, Pillow JW, Huk AC (2016) Dissociated functional significance of
791 decision-related activity in the primate dorsal stream. Nature 535:285–288.
792 Kayser C, Logothetis NK (2007) Do early sensory cortices integrate cross-modal
793 information? Brain Struct Funct 212:121–132.
794 Kingdom FAA, Prins N (2010) Psychophysics: a practical introduction. Academic Press.
795 Kleiner M, Brainard DH, Pelli DG (2007) What's new in psychtoolbox-3? Perception 36:1–

796 16.
797 Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L (2007) Causal
798 inference in multisensory perception. PLoS One 2:e943.
799 Lakatos P, Chen C-M, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations
800 and multisensory interaction in primary auditory cortex. Neuron 53:279–292.
801 Lee H, Noppeney U (2011a) Physical and perceptual factors shape the neural mechanisms
802 that integrate audiovisual signals in speech comprehension. J Neurosci 31:11338–11350.
803 Lee H, Noppeney U (2011b) Long-term music training tunes how the brain temporally binds
804 signals from multiple senses. Proc Natl Acad Sci 108:E1441–E1450.
805 Lewald J, Guski R (2003) Cross-modal perceptual integration of spatially and temporally
806 disparate auditory and visual stimuli. Cogn Brain Res 16:468–478.
807 Lewis R, Noppeney U (2010) Audiovisual synchrony improves motion discrimination via
808 enhanced connectivity between early visual and auditory areas. J Neurosci 30:12329–
809 12339.
810 Magnotti JF, Ma WJ, Beauchamp MS (2013) Causal inference of asynchronous audiovisual
811 speech. Front Psychol 4:798.
812 Maier JX, Di Luca M, Noppeney U (2011) Audiovisual asynchrony detection in human
813 speech. J Exp Psychol Hum Percept Perform 37:245–256.
814 Meijer D, Veselič S, Calafiore C, Noppeney U (2019) Integration of audiovisual spatial
815 signals is not consistent with maximum likelihood estimation. Cortex 119:74–88.
816 Mesulam M (1998) From sensation to cognition. Brain 121:1013–1052.
817 Moerel M, De Martino F, Formisano E (2014) An anatomical and functional topography of
818 human auditory cortical areas. Front Neurosci 8:225.
819 Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory
820 auditory–visual interactions during early sensory processing in humans: a high-density

electrical mapping study. 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.

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, 6th Edition. (Poeppel D, Mangun GR, Gazzaniga MS, eds), pp 141–150. MIT Press.

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, Spence C (2009) 'When birds of a feather flock together': synesthetic correspondences modulate audiovisual integration in non-synesthetes. PLoS One 4:e5664.

Parise C V., Spence C, Ernst MO (2012) When correlation implies causation in multisensory integration. Curr Biol 22:46–49.

Parise C V, Ernst MO (2016) Correlation detection as a general mechanism for multisensory integration. Nat Commun 7:11543.

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.

846 Rao RPN, Ballard DH (1999) Predictive coding in the visual cortex: a functional
847 interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79–87.
848 Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman
849 primates illuminate human speech processing. Nat Neurosci 12:718–724.
850 Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of “what” and
851 “where” in auditory cortex. Proc Natl Acad Sci 97:11800–11806.
852 Recanzone GH (2009) Interactions of auditory and visual stimuli in space and time. Hear Res
853 258:89–99.
854 Recanzone GH, Cohen YE (2010) Serial and parallel processing in the primate auditory
855 cortex revisited. Behav Brain Res 206:1–7.
856 Rohe T, Ehlis A-C, Noppeney U (2019) The neural dynamics of hierarchical Bayesian causal
857 inference in multisensory perception. Nat Commun 10:1907.
858 Rohe T, Noppeney U (2015a) Sensory reliability shapes perceptual inference via two
859 mechanisms. J Vis 15:22.
860 Rohe T, Noppeney U (2015b) Cortical hierarchies perform Bayesian causal inference in
861 multisensory perception. PLoS Biol 13:e1002073.
862 Rohe T, Noppeney U (2016) Distinct computational principles govern multisensory
863 integration in primary sensory and asociation cortices. Curr Biol 26:509–514.
864 Rohe T, Noppeney U (2018) Reliability-weighted integration of audiovisual signals can be
865 modulated by top-down attention. eNeuro 5:ENEURO.0315-17.2018.
866 Romanski LM (2012) Convergence of auditory, visual, and somatosensory information in
867 ventral prefrontal cortex. In: The neural bases of multisensory processes (Murray M,
868 Wallace M, eds), pp 667–682. Boca Raton (FL): CRC Press.
869 Sadaghiani S, Maier JX, Noppeney U (2009) Natural, metaphoric, and linguistic auditory
870 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 M, 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 (80) 268:889–893.

Sereno MI, Huang R-S (2014) Multisensory maps in parietal cortex. Curr Opin Neurobiol 24:39–46.

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 (80) 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 ventriloquism 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 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: North-Holland/Elsevier Science Publishers.

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.

921 Yendiki A, Greve DN, Wallace S, Vangel M, Bockholt J, Mueller BA, Magnotta V,
922 Andreasen N, Manoach DS, Gollub RL (2010) Multi-site characterization of an fMRI
923 working memory paradigm: reliability of activation indices. Neuroimage 53:119–131.
924

Figure legends

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: i. visual location (left vs. right), ii. auditory location (left vs. right), iii. 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.

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 x 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 i. left inferior frontal sulcus/precentral sulcus; ii. bilateral superior frontal gyrus; iii. right posterior intraparietal sulcus; and iv. right anterior intraparietal sulcus that are displayed on axial slices of a mean image created by averaging the participants' normalized structural images. L, Left;

949 R, right; SC, physically spatially congruent; SI, physically spatially incongruent; DC,
950 common cause decision; DS, separate cause decision; a.u., arbitrary unit.
951
952 **Figure 3. Multivariate pattern results along the visual and auditory spatial cortical**
953 **hierarchy.** Support vector classification decoding accuracy for: i. V = visual location: left vs.
954 right; ii. A = auditory location: left vs. right; iii. S = physical spatial congruency: congruent
955 vs. incongruent; iv. D = causal decisional outcome: common vs. separate causes; and v. M =
956 motor response: left vs. right hand in the regions of interest (ROI) as indicated in the figure.
957 Box plots show the accuracies across participants (box for median and interquartile range,
958 whiskers for lowest and highest data points, dots for outside of 1.5 interquartile range).
959 Significance is indicated by $^{**}p < 0.01$, $^{***}p < 0.001$, $^{\Delta}p < 0.0045$; the single triangle
960 indicates that the p-value is significant when adjusting the threshold according to Bonferroni
961 correction i.e. $p < 0.0045 * 11\text{ROIs} = 0.0495$. The ROIs are delineated on the surface of an
962 inflated single participant brain. V1, primary visual cortex; V2, secondary visual cortex; V3,
963 V3AB, higher order visual cortices; HG, Heschl's gyrus; PT, planum temporale; IPS0–2,
964 posterior intraparietal sulcus; IPS3–4, anterior intraparietal sulcus; PCG, precentral gyrus;
965 FEF, frontal eye-fields; DLPFC, dorsolateral prefrontal cortex.
966
967 **Figure 4. Characterization of BOLD-response patterns.** BOLD-response parameter
968 estimates for each of the two classes (e.g. left and right visual location) are summed within
969 each region weighted by the support vector classification weights. **(A)** Support vector
970 classification for i. V = visual location: left vs. right; ii. A = auditory location: left vs. right;
971 iii. S = physical spatial congruency: congruent vs. incongruent; iv. D = causal decisional
972 outcome: common vs. separate causes; and v. M = motor response: left vs. right hand in the
973 regions of interest as indicated in the figure. Box plots show the weighted sum of parameter

974 estimates across participants (box for median and interquartile range, whiskers for lowest and
975 highest data points, dots for outside of 1.5 interquartile range). **(B)** Support vector
976 classification for causal decisional outcome (i.e. common (DC) and separate causes (DS))
977 trained separately for physically spatial congruent (SC) and incongruent (SI) stimuli. V1,
978 primary visual cortex; V2, secondary visual cortex; V3, V3AB, higher order visual cortices;
979 HG, Heschl's gyrus; PT, planum temporale; IPS0–2, posterior intraparietal sulcus; IPS3–4,
980 anterior intraparietal sulcus; PCG, precentral gyrus; FEF, frontal eye-fields; DLPFC,
981 dorsolateral prefrontal cortex.

Tables

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.

989 **Table 2. Univariate results of the main effect of causal decision and the interaction of**
990 **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 x 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 x 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

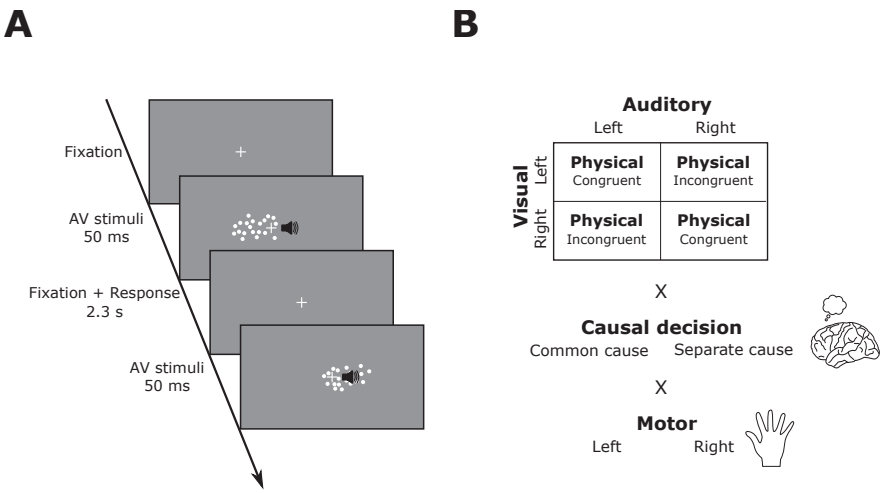
991 $p_{FWE} < 0.05$ at the cluster level corrected for multiple comparisons within the entire brain,
992 with an auxiliary uncorrected voxel threshold of $p < 0.001$. We also report the z-score of the
993 peak-voxel (or several peak voxels) with their corresponding MNI coordinates. L, Left; R,
994 right.

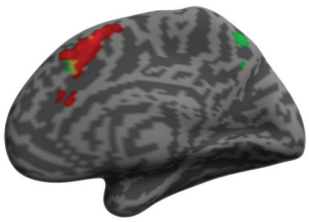
995

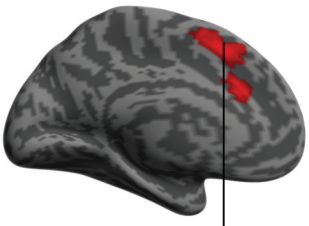
996 **Table 3. Multivariate pattern classification results.**

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	
	p-value	subject	p-value	subject	p-value	subject	p-value	subject	p-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)
IPS0–2	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)

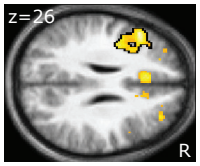
997 p-values (uncorrected) indicate better than chance decoding accuracy at the group level based
998 on between-subjects permutation test, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{\Delta}p < 0.0045$ (i.e. significant
999 after Bonferroni correction for 11 regions of interest); 'subjects' indicate the number of
1000 subjects that are associated with better than chance decoding accuracy based on within-
1001 subjects permutation test at $p < 0.05$ uncorrected (in brackets: number of subjects with
1002 $p < 0.0045$, i.e. significant after Bonferroni correction for 11 regions of interest unless
1003 guided by priori hypothesis); V1, primary visual cortex; V2, secondary visual cortex; V3,
1004 V3AB, higher order visual cortices; HG, Heschl's gyrus; PT, planum temporale; IPS0–2,
1005 posterior intraparietal sulcus; IPS3–4, anterior intraparietal sulcus; PCG, precentral gyrus;
1006 FEF, frontal eye-fields; DLPFC, dorsolateral prefrontal cortex.

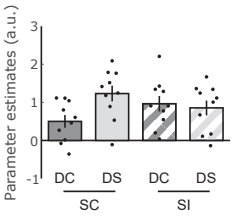


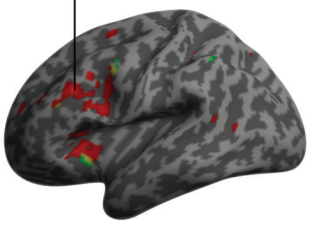




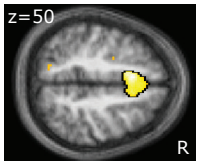
(i) causal decision x physical spatial congruency (incorrect > correct): left inferior frontal sulcus/precentral sulcus

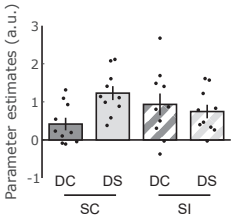


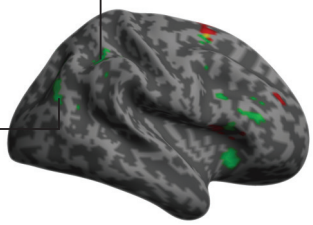




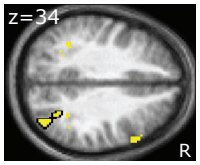
(ii) causal decision x physical spatial congruency (incorrect > correct): bilateral superior frontal gyrus

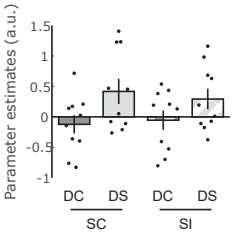




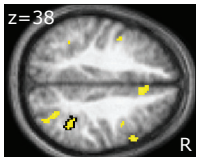


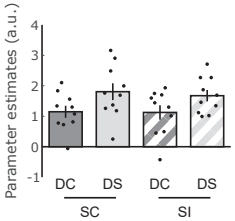
(iii) causal decision (separate > common cause): right posterior intraparietal sulcus



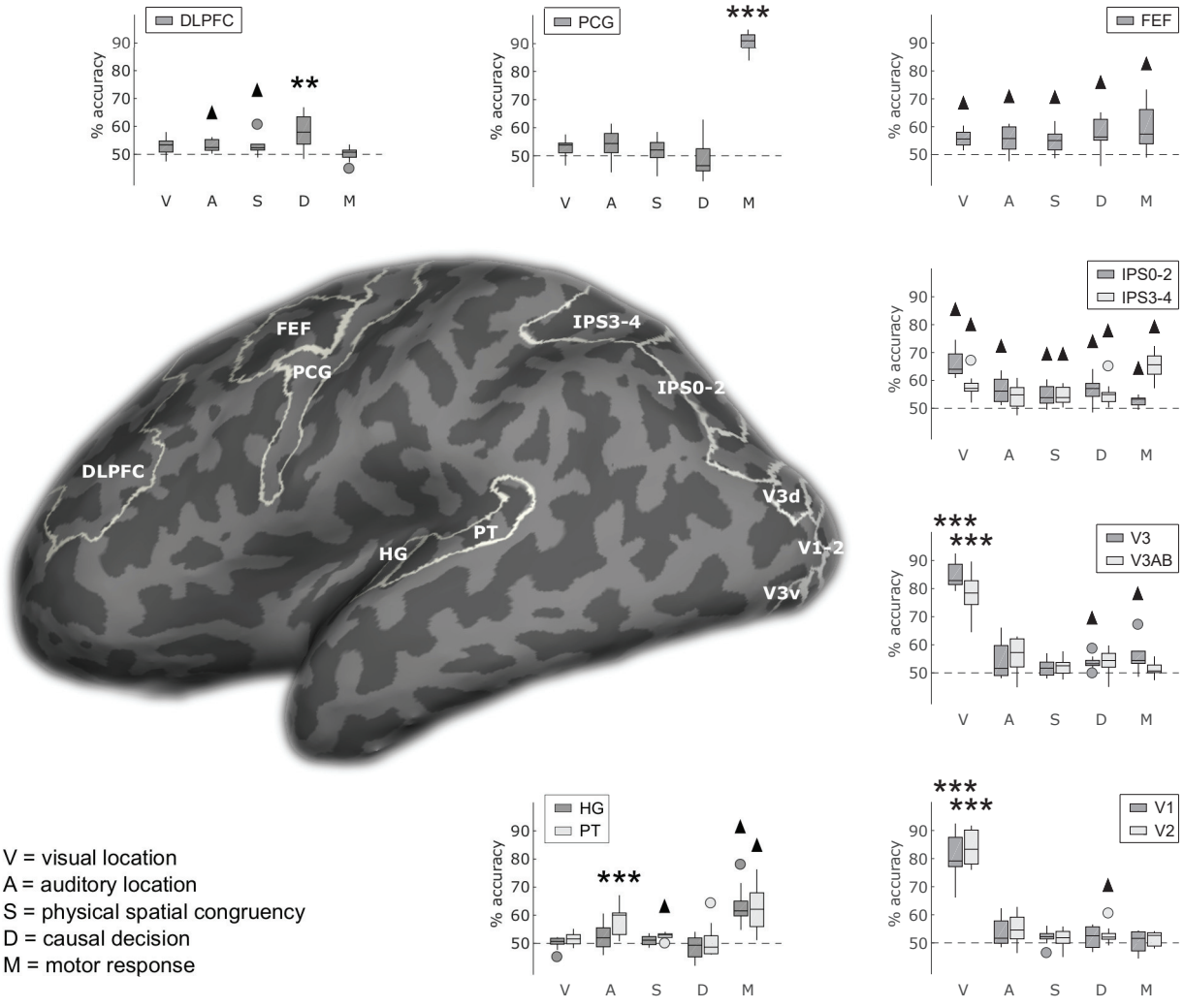


(iv) causal decision (separate > common cause): right anterior intraparietal sulcus





causal decision x physical spatial congruency (incorrect > correct)
causal decision (separate > common cause)



V = visual location
A = auditory location
S = physical spatial congruency
D = causal decision
M = motor response