Spatio-temporal dynamics of visual information representation and transformation in brain networks resolving algorithmic functions

Katarzyna Jaworska[1], Nicola J. van Rijsbergen[2], Robin A.A. Ince[1], & Philippe G. Schyns[1*]

[1] Institute of Neuroscience and Psychology, University of Glasgow, 62 Hillhead Street, G12 8QB, Glasgow

[2] Department of Psychology, Edge Hill University, St Helens Road, L39 1QP, Ormskirk

[*] Corresponding author: Philippe.Schyns@glasgow.ac.uk

Abstract

A key challenge in systems neuroscience remains to understand where, when and how mass brain signals that reflect network activity dynamically represent, transmit and transform sensory information for task behavior. Here, we used the classic XOR, OR and AND that imply a different computation on the same inputs for correct task behavior. We disentangled MEG source activity into three distinct information processes that linearly represents each input before nonlinearly integrating them for task behavior.

Experiment 1 was a visual XOR, a nonlinear computation that returns "true" when only one of its two inputs is true (see Figure 1A). In 10 observers, we simultaneously projected each input (the dark vs. clear lenses of a pair of glasses on a face) in the left and right visual hemifields, to enable, across their initial contra-lateral representation separately the left and right hemispheres, their cross-hemispheric transfer and subsequent integration for explicit XOR decisions—i.e. press the "yes" key whenever only one of the inputs was dark; the "no" key otherwise (see Figure 1A).

Independently in each observer (i.e. in 9 replications), we disentangled concurrently measured magnetoencephalography (MEG) activity (source localized with an LCMV beamformer from 248-sensors) into different systems-level processes that (a) linearly represent the left and the right visual inputs (i.e. dark vs. clear, see Figure 1A), (b) nonlinearly integrate them for XOR decisions and (c) behavioral responses (e.g. reaction times, RT). Figure 1A summarizes the logic of these analyses in analogy to an XOR network. We controlled XOR with 10 additional observers who performed an AND task and 10 others who performed an OR task, importantly from the same inputs.

Experiment 2 repeated the design, but with a 1 s delay inserted between presentation of the first and second inputs (see Figure 2A), to enable task-specific strategies and study their impact on the three system-level processes. These strategies depended on the state of the first presented input. In AND, a clear first input logically discloses a "no" response, rendering the representation, transfer and integration of the second input in principle unnecessary, possibly speeding up RTs. We present the results of these two experiments in turn.

Figure 1B-C summarizes the time courses of information processes that represent, transfer and transform the two inputs for behavior (see Supplementary Video for animations). First, we note the dynamic, contra-lateral linear representation of each input in V1-4 regions (left input color-coded in blue; right in orange) from ~60 ms post-stimulus. Following cross-hemispheric transfer, ventral and dorsal pathways voxels linearly represent both inputs ~100 ms (color-coded in magenta) before their nonlinear integration for decision in frontal-parietal regions ~260 ms (color-coded in green), itself followed by representation of behavioral RTs in left motor cortex (color-coded in yellow). Figure 1B therefore reveals marked transition in states of information processing (i.e. network reorganizations), from linear representations over most of occipital, temporal and parietal voxels, to nonlinear integration in the right occipital-temporal and parietal cortices preceding behavior in left motor cortex. Adjacent color-coded distributions indicate that these effects were all replicated across all observers (except one integration effect in one participant, prevalence =

[61.3% 99.3%]). Finally, Supplementary Figure S1 details how individual voxels represented each state of information processing at its peak time point. Supplementary Figure S2 replicates all these effects in all individual participants of the AND and OR tasks (prevalence = [74.5% 100%]).

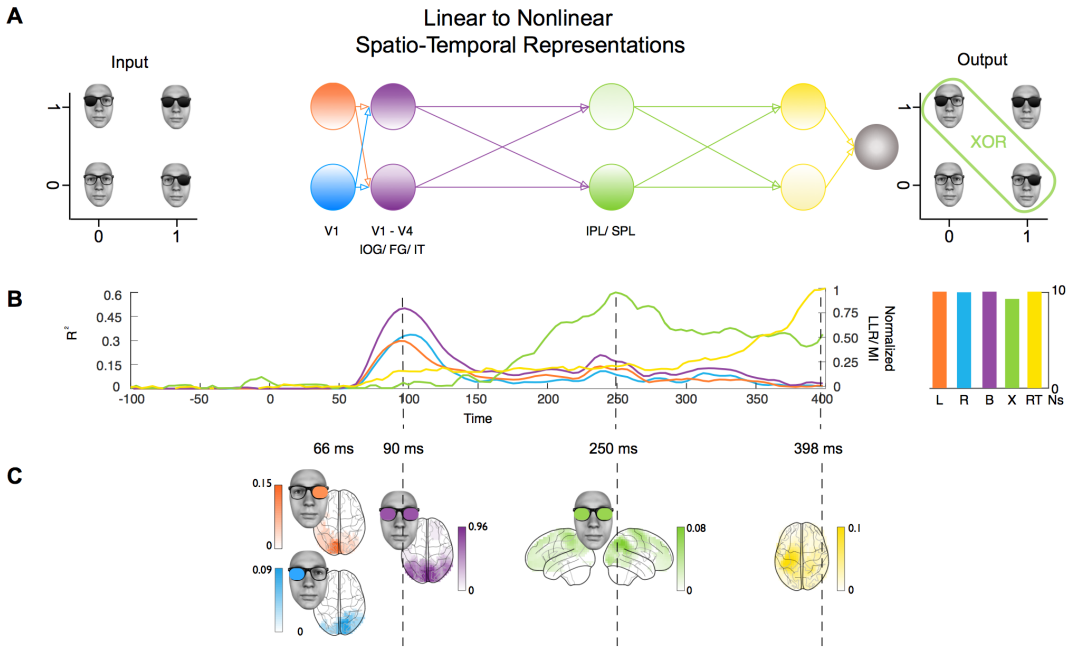

Figure 1. (A) Schematic brain network and hypotheses. Stimuli comprised a face wearing glasses, with left and right inputs being dark or clear, for a total of 4 input conditions for behavioral decisions (illustrated for XOR). Note: the face stimulus was averaged from several artificially synthesized faces that did not belong to any real person. **(B) Time course of maximum linear and nonlinear input representations.** To determine the linear vs. nonlinear input representations at each stage, we applied a data-driven analysis at each voxel and time point and computed a log-likelihood ratio (LLR) between two linear models of input representation—i.e. with and without an interaction term (see *Methods*). Average (across 10 XOR participants) time courses of the maximum (across voxels) R^2 for left (blue), right (orange), and both inputs (magenta) linear representation and (2) LLR indicating nonlinear representations of the two inputs. Histogram to the right indicates how many participants showed significant representation or integration at the times of interest -- N = 10/10 for left, right, and both lenses representation; and for nonlinear integration, N = 9/10 in time windows 214-230 and 261-277 ms in parietal cortex. **(C) Summary of localized states of information representation and transformation.** Glass brains depict average (across participants) Mahalanobis distances (see *Methods, Mahalanobis Distances*) at specific time points: At stage 1 (~60 ms post-stimulus), blue and orange brains localize the group average onset of contra-lateral linear representation of the left and right inputs. Around 100 ms post-stimulus, a purple brain localizes the median peak-time average linear representation of both inputs in midline occipital cortex. At stage 2 (250 ms post-stimulus), the green brain localizes the average nonlinear (integrated) XOR representation of both inputs in right parietal cortex at median peak time. At stage 3 (398 ms post-stimulus), the yellow brain localizes the average representation of reaction times in left motor cortex.

Experiment 1 revealed three information processes that represent, transfer and integrate two spatially distinct visual inputs for task behavior. Experiment 2 includes an additional temporal separation that enabled the intervention of different strategies that should affect the information processes and behavioral RTs. To explain, each trial started with one lateralized input (i.e. dark or clear) while the other input was greyed out for 1s. Following this, the second input became dark or clear (see Figure 2A). We instructed participants (N = 5/task) to wait until both inputs were available to respond. Note that the first input logically resolves AND and OR (on trials when it is clear, in AND, leading to "no" and dark, in OR, leading to "yes"), whereas XOR requires nonlinear integration of the two inputs on each trial.

First, we replicated the information processes of Experiment 1 in each participant and task—i.e. linear representation of the first input, followed by linear representation of the second input, this time with a 1 s gap, then their nonlinear integration and behavioral response, see Figure 2B. Then, we turned to the analyzing the impact of task-specific strategies on the information processes and behavioral RTs.

We found the expected shorter RTs when the clear input logically resolves AND and when a dark first input resolved OR. Table 1 shows the average RTs (computed over the median RTs of each participant), where AND (vs. OR) participants responded faster when the first input resolved the task, whereas the first input had no effect on XOR RTs (see Table 1).

Table 1. Reaction times. Median RTs per participant and condition were averaged across observers. The first row reports the group average and the corresponding standard deviation. The last column reports group difference between conditions (First clear minus First dark) quantified with a t-test. Square brackets indicate 95% confidence intervals for the difference.

	AND	OR	XOR
Average	398 ± 120 ms	408 ± 88 ms	542 ± 82 ms
First clear	368 ms	439 ms	541 ms
First dark	434 ms	379 ms	549 ms
First clear – First dark	-65 ms [-116, -14], p = 0.024	60 ms [-6, 126], p = 0.066	-8 ms [-40, 23], p = 0.506

Such an effect of strategies on behavioral RTs should have systems level correlate. Here, our methodology focusing on information processing could ascribe RT differences to the integration stage (primarily in parietal cortex) and to the motor preparation stage (primarily in pre-motor cortex). Specifically, we found that representation of behavioral RTs into MEG activity was earlier in AND and OR than in XOR (see Figure 2B). To better understand the impact of task-specific strategies on the information processes, we split the trials according to the state of the first input (ie. clear or dark, irrespective of left or right presentation). Stronger parietal synergistic interactions reflected stronger input integration in the second

time window, when the first input did not disclose the solution to the task on its own—i.e. dark in AND and clear in OR (see Supplementary Figure S5; SMI, see *Methods, Information Theoretic Analysis*). In fact, a linear-mixed effects model (with random effect of observer) could predict single-trial RTs from the interaction between first lens representation (dark vs. clear) and SMI peak in AND ($t(4519) = -2.02$, $p=0.04$) and OR ($t(4647) = 4.99$, $p<0.0001$), but not in XOR participants ($t(4867) = -0.18$, $p=0.85$) when trial-wise synergy was taken at the voxel in the right hemisphere showing the strongest average synergy. These models predicted that when the first lens was clear and synergy was higher, single-trial reaction times would be ~60 ms slower in OR participants, but ~64 ms faster in AND participants. Importantly, these strategic effects started before the slow and sustained representation of behavioral RTs into MEG--on average, by 70 ms [51, 89], $p < 0.001$ (Figure 5E), peak to peak, 87 ms [48, 126] ($p < 0.001$).

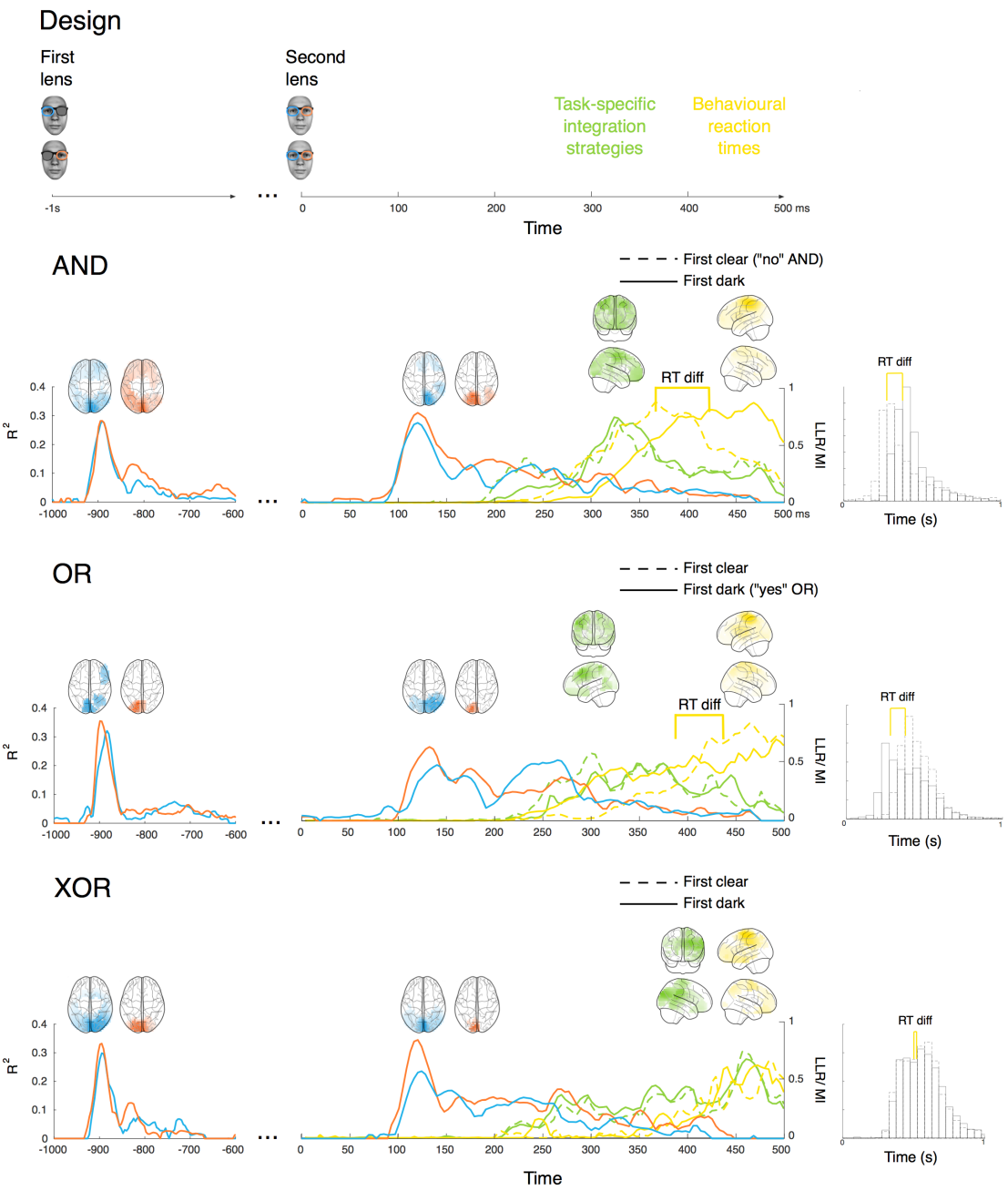

Figure 2. Using the same stimuli as in experiment 1, we presented either the right or the left input as grey for the first 1000 ms following stimulus onset. The other input was either clear or dark. After 1000 ms, the grey input changed into either clear or dark. Participants waited until the grey input changed, and then responded according to the logical function task: AND (N=5, top), OR (N=5, middle), and XOR (N=5, bottom). Linear representation of the first input started around 80 ms post-stimulus (with a ~100 ms peak; or around -900 ms with respect to second lens presentation). The second input was linearly represented from ~100 ms with a peak at ~120 ms (blue and orange time courses). As in Experiment 1, input representations were again located in occipital, temporal and parietal regions (blue and orange brains). Nonlinear input integration started around 190 ms in the right fronto-temporal cortex (with a 277 ms peak when the left input was first available; 433 ms when the right input was first available; green time courses). As in Experiment 1, nonlinear integration spanned a widespread network of temporal, parietal and frontal regions past 200 ms. Similar results characterized AND and OR participants. We also plotted task-relevant nonlinear integration of the two inputs (green) according to whether the first lens was clear (dashed lines) or dark (solid lines; see

also *Methods*). We then plotted, in yellow, the average MI(MEG; RT) depending on first lens input – dashed for clear, solid for dark. Insets on the right present histograms of reaction times across participants. Note: the face stimulus was averaged from several artificially synthesized faces that did not belong to any real person.

Summary

Here, we conducted two experiments that tested the nonlinear integration of features (with the classic XOR, OR and AND tasks). We summarize the main findings as follows. First, we revealed three information-processing stages that represent, transfer and integrate two spatially distinct visual inputs for task behavior (XOR). Specifically, after initial contra-lateral linear representation of the left and the right lens as early as ~60 ms post-stimulus, we observed linear representation both inputs ~100 ms before their nonlinear integration for decision ~260 ms, itself followed by representation of behavioral RTs. As such, we revealed marked transition in states of information processing (i.e. network reorganizations), from linear representations over most of occipital, temporal and parietal voxels, to nonlinear integration in the right occipital-temporal and parietal cortices preceding behavior in left motor cortex.

Second, we replicated these findings in two separate groups of participants (AND and OR), by showing the same linear to nonlinear transition of information processing states, although with a different task representation at the granularity of single voxels.

Third, we also replicated the linear to nonlinear transition of information processes of Experiment 1 in a follow up study with the same stimuli and tasks, this time with a 1 s gap inserted between the presentation of the first and the second input. This temporal gap enabled us to uncover task-specific strategies on the information processes and behavioral RTs. Specifically, these strategies depended on the state of the first presented input. In AND, a clear first input logically disclosed a “no” response (whereas a dark first input disclosed a “yes” response on OR), rendering the representation, transfer and integration of the second input in principle unnecessary, and speeding up RTs. Both inputs were required to resolve XOR, therefore there was no differentiation of RTs depending on first input state.

Altogether, with mass brains signals, our results provide a systems-level understanding of where, when and how brain networks dynamically represent, transfer and nonlinearly integrate features for behavior.

Methods

Observers

Thirty-five observers participated in Experiment 1 (all right-handed; 24 females), and sixteen observers participated in Experiment 2 (all right-handed, 10 females). All observers reported normal or corrected-to-normal vision, and gave written informed consent. The study was conducted in accordance with the British Psychological Society ethics guidelines, and approved by the ethics committee at the College of Medical, Veterinary and Life Sciences, University of Glasgow.

Stimuli

We used an average face taken from the GFG database at the Institute of Neuroscience and Psychology, University of Glasgow (Yu, Garrod, & Schyns, 2012), and added an image of glasses using Photoshop. We then manipulated visibility of the lenses in Experiment 1 by adding a black patch to the left, the right, or both lenses, so that the resulting ("original") stimulus classes were: 1) left Clear/ right Dark; 2) left Dark/ right Clear; 3) both Dark; and 4) both Clear. In Experiment 2, we made additional stimuli where the right or the left lens was grey, i.e. 1) left Grey/ right Clear; 2) left Grey, right Dark; 3) left Clear, right Grey; 4) left Dark, right Grey. The edges of the left and the right lens were 0.5 deg away from the centrally presented fixation cross.

Task Procedure

Experiment 1. Each trial began with a fixation cross displayed for a random duration of 500-1000 ms, immediately followed by one of four "original" stimulus classes as explained above, and displayed for 150 ms. We instructed participants to maintain fixation on each trial and to pay attention to the opacity of the left and the right lens, and respond as quickly and accurately as possible by pressing one of two keys ascribed to each response choice with the index and middle fingers of their right hand. The responses were "same" vs. "different" in the XOR task; "both black" vs. "otherwise" in the AND task; or "at least one black" vs. "otherwise" in the OR task.

Experiment 2. Each trial began with a fixation cross displayed for random duration of 500-1000 ms. Next, one of four stimulus classes containing a grey lens either on the left or on the right hemifield was presented for 1000 ms, immediately followed by one of four "original" stimulus classes that was displayed for 150 ms. This presentation regime resulted in 8 trial conditions: Grey/Clear – Clear/Clear; Grey/Clear – Dark/Clear; Grey/Dark – Clear/Dark; Grey/Dark – Dark/Dark; Clear/Grey – Clear/Dark; Clear/Grey – Clear/Clear; Dark/Grey – Dark/Clear; Dark/Grey – Dark/Dark. The fixation cross remained on the face image throughout the duration of the trial. Participants were instructed to wait until the grey lens changed into clear or dark, and then respond as quickly and accurately as possible by pressing one of the keys with their index and middle fingers of the right hand, as per the tasks described above.

In both Experiments 1 and 2, stimuli were presented in blocks of 80 trials; with randomized inter-trial interval of 800-1300 ms. The order of stimulus presentation was randomized in each block. Participants completed a total of 20-24 blocks split across 2-3 single day sessions, with short breaks between blocks. Each session lasted 2.5-3 hours.

MEG Data Acquisition and Pre-processing

We recorded the participants' MEG activity using a 248-magnetometer whole-head system (MAGNES 3600 WH, 4-D Neuroimaing) at a 1017 Hz sampling rate. For each participant, we discarded runs with more than 0.6 cm head movement measured with pre- vs. post-run head position recording. We excluded five participants in Experiment 1, and one participant in Experiment 2 for too many runs with excessive movement, resulting in a final sample sizes of N = 30 (Experiment 1; 10 per task) and N = 15 (Experiment 2; 5 per task). Mean head movement (averaged across blocks) across participants was 0.3 cm (min = 0.12, max = 0.44).

We performed analyses using the Fieldtrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011) and in-house MATLAB code, according to recommended guidelines (Gross et al., 2013). We high-pass filtered the data at 0.5 Hz (5th order two-pass Butterworth IIR filter), filtered for line noise (notch filter in frequency space) and de-noised via a PCA projection of the reference channels. We identified noisy channels, jumps, muscle and other signal artifacts using a combination of automated techniques and visual inspection. The median number of trials entered into subsequent analyses was 1064 (min = 701, max = 1361).

Next, we epoched the data into trial windows (-500 to 1000 ms around stimulus onset in Experiment 1; -500 to 1500 ms in Experiment 2), low-pass filtered the data at 45 Hz (3rd order two-pass Butterworth IIR filter), resampled to 256 Hz, and decomposed using ICA, separately for each observer. We identified and projected out of the data the ICA sources corresponding to heartbeat, and eye blinks or movements (2 – 4 components per participant).

Source reconstruction

For each participant, we co-registered their structural MRI scan with their head shape recorded on the first session, and warped to standardized MNI coordinate space. Using brain surfaces segmented from individual warped MRI, we then prepared a realistic single-shell head model. Next, we low-pass filtered the clean dataset at 40 Hz, re-epoched the data in Experiment 1 between -100 and 400 ms (and -500 to 1500 ms in Experiment 2) around stimulus onset, demeaned using a pre-stimulus baseline, and computed covariance across the entire epoch. Using average sensor positions across good runs (i.e. where head movement was <0.6 cm, see above), and a 6mm uniform grid warped to standardized MNI space, we then computed the forward model, keeping the two orientations of MEG activity. We proceeded to compute the Linearly Constrained Minimum Variance (LCMV) beamformer (Hillebrand & Barnes, 2005) solution with parameters "lambda = 6%" and "fixedori = no". The resulting inverse filter applied to the sensor-space MEG activity then allowed us to reconstruct single-trial 2D MEG time courses on 12,773 grid points. Using a Talaraich-Daemon atlas, we excluded all cerebellar and non-cortical sources, and performed statistical analyses on 5,848 cortical grid points.

Statistical Analyses

Linear Representation

We computed multivariate linear regressions to model the dependence between lens opacity (clear vs. dark, corresponding to logical 0 and 1 in Figure 1, respectively) and 2D MEG source responses on each voxel and 4 ms time point between -100 and 400 ms post-stimulus. We consider linear regression models for each lens opacity separately (left and right), and for the additive combination of both lenses (3 models).

$$\mathbf{y} = \beta_0 + \beta_1 L$$

$$\mathbf{y} = \beta_0 + \beta_1 R$$

$$\mathbf{y} = \beta_0 + \beta_1 L + \beta_2 R$$

We fit each model by ordinary least-squares obtaining beta coefficients for the intercept and slope. To quantify the fit in the 2D response space we used a multivariate R^2 which quantifies multivariate variance as the determinant of the covariance matrix:

$$R^2 = 1 - \frac{|(\boldsymbol{y} - \widehat{\boldsymbol{y}})^T(\boldsymbol{y} - \widehat{\boldsymbol{y}})|}{|(\boldsymbol{y} - \overline{\boldsymbol{y}})^T(\boldsymbol{y} - \overline{\boldsymbol{y}})|}$$

where $\boldsymbol{y}, \overline{\boldsymbol{y}}, \widehat{\boldsymbol{y}}$ are the 2D data, mean and model predictions respectively. This linear modelling produced a time course of R^2 values per voxel, with one R^2 value every 4 ms.

We obtained a non-parametric statistical threshold controlling the Familywise Error Rate (FWER) over all considered time points and sources using the method of maximum statistic (Groppe, Urbach, & Kutas, 2011). Specifically, on each of 100 permutations we randomly shuffled lens opacity across the experimental trials, and repeated the linear modelling and R^2 computation explained above, and extracted the maximum R^2 across all voxels and time points. This produced a distribution of 100 maximum R^2 values, of which we used the 95th percentile as statistical threshold (FWER $p < 0.05$).

Nonlinear Representation

On each source voxel, we also extended our linear regression model of dynamic MEG activity with a fourth model considering an interaction term between the two eye features.

$$\boldsymbol{y} = \beta_0 + \beta_1 L + \beta_2 R$$

$$\boldsymbol{y} = \beta_0 + \beta_1 L + \beta_2 R + \beta_3 L \; x \; R$$

We used a log-likelihood ratio (LLR) test to test whether the addition of the interaction term significantly improved model fit ($p < 0.05$, FDR-corrected over time points and voxels, Benjamini & Yekutieli, 2001; Groppe et al., 2011).

Mahalanobis Distances

Inspection of the MEG source data revealed clusters in the representation of the four input conditions of lens opacity, corresponding to different information processing states. For each participant, we applied k-means clustering across all voxels and time points to isolate six states of information-processing.

To rigorously quantify each information processing state (i.e. ensure that linear and nonlinear representations relate to tested hypotheses about linear and nonlinear input representations), we applied two computations on each voxel and time point, independently for each participant. First, we computed six pairwise Mahalanobis distances between the color-coded 2D distributions of single trial MEG activity corresponding to each input condition (see Figure 3, Supplementary Figure 1). That is, we averaged the covariance matrices of each pair of input conditions and multiplied the inverse average covariance by the difference of the condition means:

$$dist = (\mu 1 - \mu 2)^T * C^{-1} * (\mu 1 - \mu 2)$$

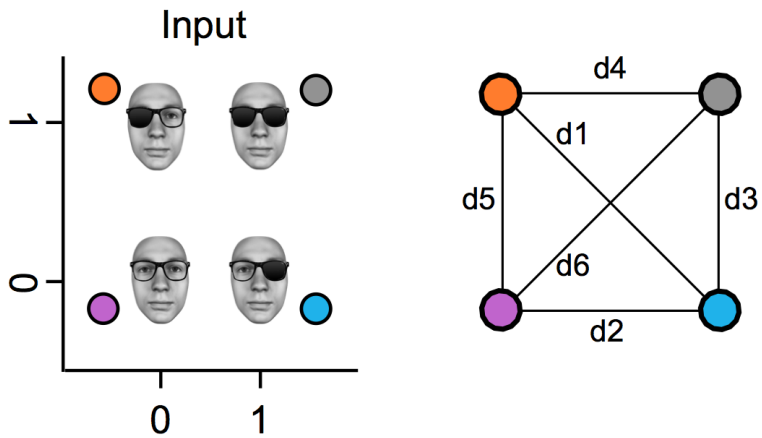

Figure 3. Note: the face stimulus was averaged from several artificially synthesized faces that did not belong to any real person.

Second, to quantify the geometric relationships between the centroids of each input distribution, we computed the differences between the six resulting pairwise distances as follows (see Figure 3, right):

- left lens representation (LL): mean([d1, d3, d5, d6]) – mean([d2, d4])
- right lens representation (RL): mean([d1, d2, d4, d6]) – mean([d3, d5])
- both lenses representation (BL): mean(all) – std(all)
- XOR representation: mean([d2, d3, d4, d5]) – mean([d1, d6])
- AND representation: mean([d4, d3, d6]) – mean([d1, d2, d5])
- OR representation: mean([d6, d2, d5]) – mean([d4, d3, d1])

We tested the statistical significance of each information processing state with permutation testing, using the method of maximum statistics. Over 25 permutations, at each voxel and time point we randomly shuffled lens opacity, repeated the distance calculations, computed the maximum differences and used the 95th percentiles of these maxima as thresholds (FWER $p < 0.05$, corrected).

Finally, we applied the dual thresholding of the significant R^2 mask for the left or right lens linear representation (see *Methods, Linear Representation*) with significant LL, RL and BL information processing state, and the dual thresholding of the significance mask for the LLR test statistic (see *Methods, Nonlinear Representation*) with significant XOR, AND, and OR information processing state.

To visualise the magnitude of group-level Mahalanobis differences per information processing state, we averaged thresholded voxel x time point matrices across observers.

Localization of Mahalanobis distances

We quantified the temporal dynamics of three stages of information processing in individual participants as follows. For early representation of the left and right lenses, we computed onsets of representation as the first significant time point of R^2. For early representation of both lenses, we computed peak time of its R^2. Finally, for late representation of the nonlinear XOR representation, we also selected peak latency of the LLR test statistic. For each state, we then computed time stamps as median across observes and extracted the voxelwise representation averaged across observers at the corresponding time stamp. We plotted these voxels back onto glass brains in Figure 1C using *Nilearn* (Abraham et al., 2014).

Information Theoretic Analysis

As an alternative analysis to the linear modelling we use an information theoretic approach to quantify the interactions between the representations of the left and right lenses (Jaworska et al., 2020; Schyns, Zhan, Jack, & Ince, 2020). We use Gaussian-Copula Mutual Information (GCMI; Ince et al., 2017), which is a semi-parametric estimator of information theoretic quantities. We used this to quantify the representation of each (binary) lens in the

2d MEG source responses: MI(L; MEG), MI(R; MEG) (analogous to the R2 from the individual models). We can also calculate the MI conveyed about the joint state of both lenses on each trial (4 possible states): MI(L,R; MEG), (analogous to the R^2 from the full model). A particular advantage of the information theoretic approach is that these quantities have a meaningful additive effect size in bits, and so they can be directly compared. By considering MI(L,R; MEG) – (MI(L; MEG) + MI(R; MEG)) we can quantify the difference in the true joint MI and that which would occur if the lenses were coded independently. If positive, this value reflects synergy: simultaneous knowledge of both lenses together increases the ability to predict the MEG signal, compared to the prediction that would be made if each lens was observed independently. In this sense, synergy is measuring a statistical interaction in the representation of the two lenses, as does the interaction term in our full regression model. However it is important to note that they are measuring different types of interaction, the regression approach quantifies a non-linear interaction, whereas synergy is a statistical interaction that can also arise in linear systems.

A further advantage of the information theoretic approach is that we can quantify the contribution of each individual trial to the overall dependence. By substituting the pointwise or local values of mutual information (Bouma, 2009; Lizier, 2014) corresponding to the variable values on each trial into the above expression for synergy we can obtain a samplewise synergy (SMI). We have one samplewise synergy value for each trial, and these values are such that the mean over trials is equal to the overall synergy value. Trials with high sample-wise synergy values represent trials contributing to the interaction. We split trials into two sets: those in which one lens was clear in the first presentation vs. those on which one lens was dark in the first presentation. This split would have a different effect on the AND and OR tasks – when dark lens was presented first, the AND task could not be resolved until the second lens is shown, whereas when clear lens was presented first, the task could already be resolved as “not AND” response. To the contrary, the OR task could be resolved as soon as the first lens was dark, but not when the first lens was clear. Both lenses were needed in order to resolve the XOR task. We calculate the mean sample-wise synergy separately for these two sets of trials. We performed these computations only on voxels and time points that displayed significant LLR test statistic (See *Methods, Nonlinear Representation*).

Acknowledgments

P.G.S. received support from the Wellcome Trust (Senior Investigator Award, UK; 107802) and the Multidisciplinary University Research Initiative/Engineering and Physical Sciences Research Council (USA, UK; 172046-01). R.A.A.I. was supported by the Wellcome Trust (Seed Awards in Science; 214120/Z/18/Z). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., … Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*(FEB), 14. https://doi.org/10.3389/fninf.2014.00014

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, *29*(4), 1165–1188. https://doi.org/10.1214/aos/1013699998

Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. *Undefined*.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*. Blackwell Publishing Inc. https://doi.org/10.1111/j.1469-8986.2011.01273.x

Gross, J., Baillet, S., Barnes, G. R., Henson, R. N., Hillebrand, A., Jensen, O., … Schoffelen, J. M. (2013, January 15). Good practice for conducting and reporting MEG research. *NeuroImage*. Neuroimage. https://doi.org/10.1016/j.neuroimage.2012.10.001

Hillebrand, A., & Barnes, G. R. (2005). Beamformer Analysis of MEG Data. *International Review of Neurobiology*, *68*(05), 149–171. https://doi.org/10.1016/S0074-7742(05)68006-3

Ince, R. A. A., Giordano, B. L., Kayser, C., Rousselet, G. A., Gross, J., & Schyns, P. G. (2017). A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. *Human Brain Mapping*, *38*(3), 1541–1573. https://doi.org/10.1002/hbm.23471

Ince, R. A. A., Jaworska, K., Gross, J., Panzeri, S., van Rijsbergen, N. J., Rousselet, G. A., & Schyns, P. G. (2016). The Deceptively Simple N170 Reflects Network Information Processing Mechanisms Involving Visual Feature Coding and Transfer Across Hemispheres. *Cerebral Cortex*, *26*(11), 4123–4135. https://doi.org/10.1093/cercor/bhw196

Ince, R. A. A., van Rijsbergen, N. J., Thut, G., Rousselet, G. A., Gross, J., Panzeri, S., & Schyns, P. G. (2015). Tracing the Flow of Perceptual Features in an Algorithmic Brain Network. *Scientific Reports*, *5*(1), 17681. https://doi.org/10.1038/srep17681

Jaworska, K., Yi, F., Ince, R. A. A., Rijsbergen, N. J., Schyns, P. G., & Rousselet, G. A. (2020). Healthy aging delays the neural processing of face features relevant for behavior by 40 ms. *Human Brain Mapping*, *41*(5), 1212–1225. https://doi.org/10.1002/hbm.24869

Lizier, J. T. (2014). Measuring the Dynamics of Information Processing on a Local Scale in Time and Space (pp. 161–193). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-54474-3_7

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*. https://doi.org/10.1155/2011/156869

Schyns, P. G., Zhan, J., Jack, R. E., & Ince, R. A. A. (2020). Revealing the information contents of memory within the stimulus information representation framework. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1799), 20190705. https://doi.org/10.1098/rstb.2019.0705

Yu, H., Garrod, O. G. B., & Schyns, P. G. (2012). Perception-driven facial expression synthesis. *Computers and Graphics (Pergamon)*, *36*(3), 152–162. https://doi.org/10.1016/j.cag.2011.12.002

Zhan, J., Ince, R. A. A., Van Rijsbergen, N. J., & Schyns, P. G. (2018). Dynamic construction of reduced representations in the brain for perceptual decision behavior. *Current Biology*, *29*(2), 319–326.

Supplementary Results

Voxel-level transition of information processing states in XOR

We detail the linear and nonlinear representations of the inputs in MEG activity at the granularity of individual voxels. Specifically, in each participant we analyzed the 2D MEG single-trial response distributions of individual voxels in the four color-coded input conditions (see Figure S1 A). These revealed four representations corresponding to distinct information processing states—specifically, linear representations of the left, right, and both inputs, and their nonlinear integration in XOR.

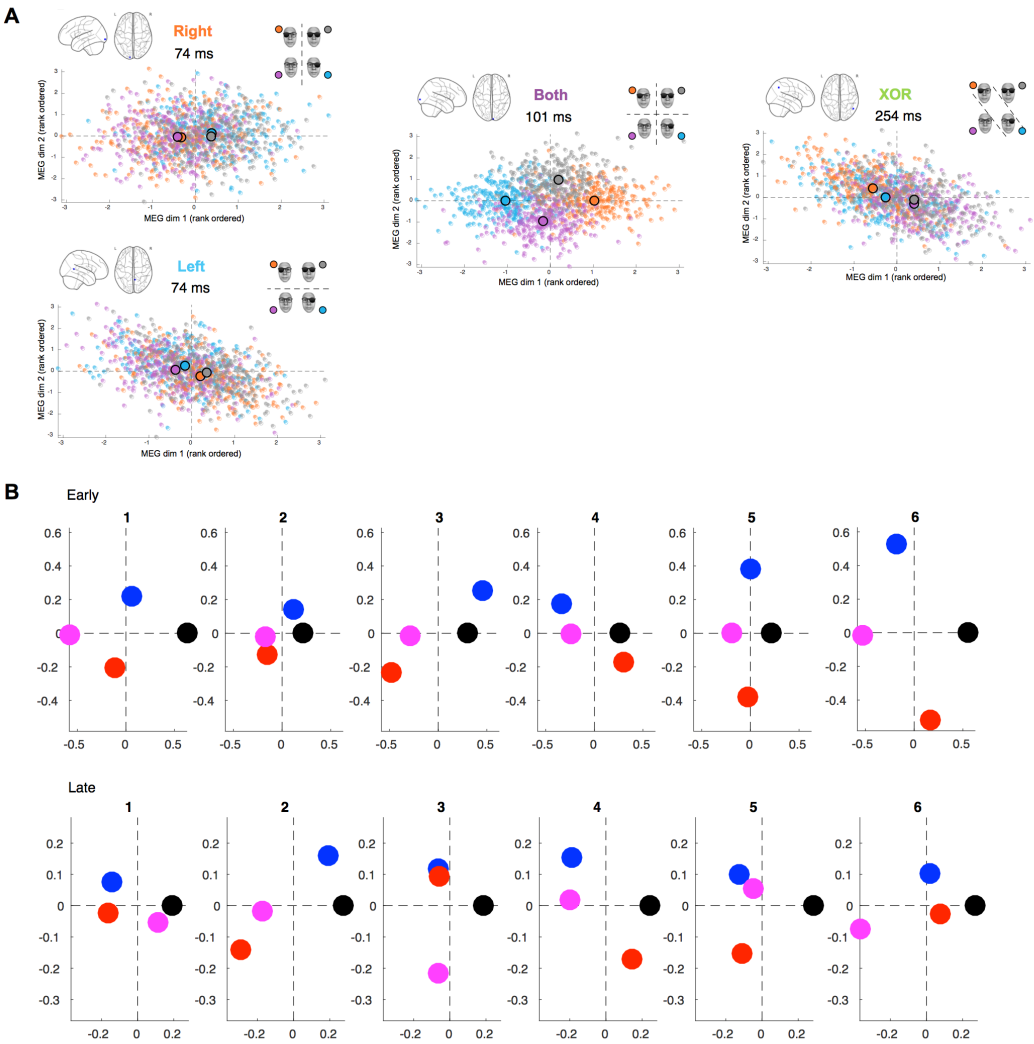

Figure S1. (A) MEG voxel activity representation of information processing states. Scatterplots show, color coded by input conditions the single-trial MEG activity (small dots; large dots – input condition averages) visualized in two-dimensional reconstruction of single voxels representative of each information processing state (from a representative XOR participant), at a single time point specified above each scatterplot. Note: the face stimulus was averaged from several artificially synthesized faces that did not belong to any real person. **(B) K-means results.**

AND and OR tasks

In AND and OR participants, we found similar patterns of results as in XOR participants, whereby both AND and OR comprised three information processing stages, starting with a linear representation of the contra-lateral input, followed with a linear representation of both inputs across ventral and dorsal regions around 100 ms, and from ~140 ms a nonlinear integration of both inputs for decision. Figure S2 B-D summarizes these results. Finally, response distributions in AND and OR observers partitioned the single trials following the AND or OR task function past ~140 ms post-stimulus (Figure S2 E), with strongest representation of each function in the right occipital-temporal and parietal cortices (see Figure S2 D, green glass brains).

In sum, Experiment 1 analyses revealed, across observers, three stages of input representation and transformation in XOR, AND and OR. At stage 1 (~60 ms post-stimulus), linear representation of each input in contra-lateral occipital cortex --i.e. representation of the left (vs. right) input in the right (vs. left) cuneus and lingual gyrus (orange and blue brains in Figure S2 C). At stage 2 (~100 ms post-stimulus), linear representation of the two inputs on voxels in occipital, parietal and temporal cortices (magenta brain in Figure S2 C). This implies cross-hemispheric transfer of the input contra-laterally represented at stage 1. At stage 3 (~140 ms post-stimulus), nonlinear representation (i.e. integration) of the two inputs on occipital, parietal, temporal, and frontal voxels (with a peak ~265-289 ms on right occipital-temporal and parietal voxels in each task; green brains in Figure S2 C). Finally, around 400 ms the representation of reaction times peaked in the left motor cortex (yellow brains in Figure S2 C).

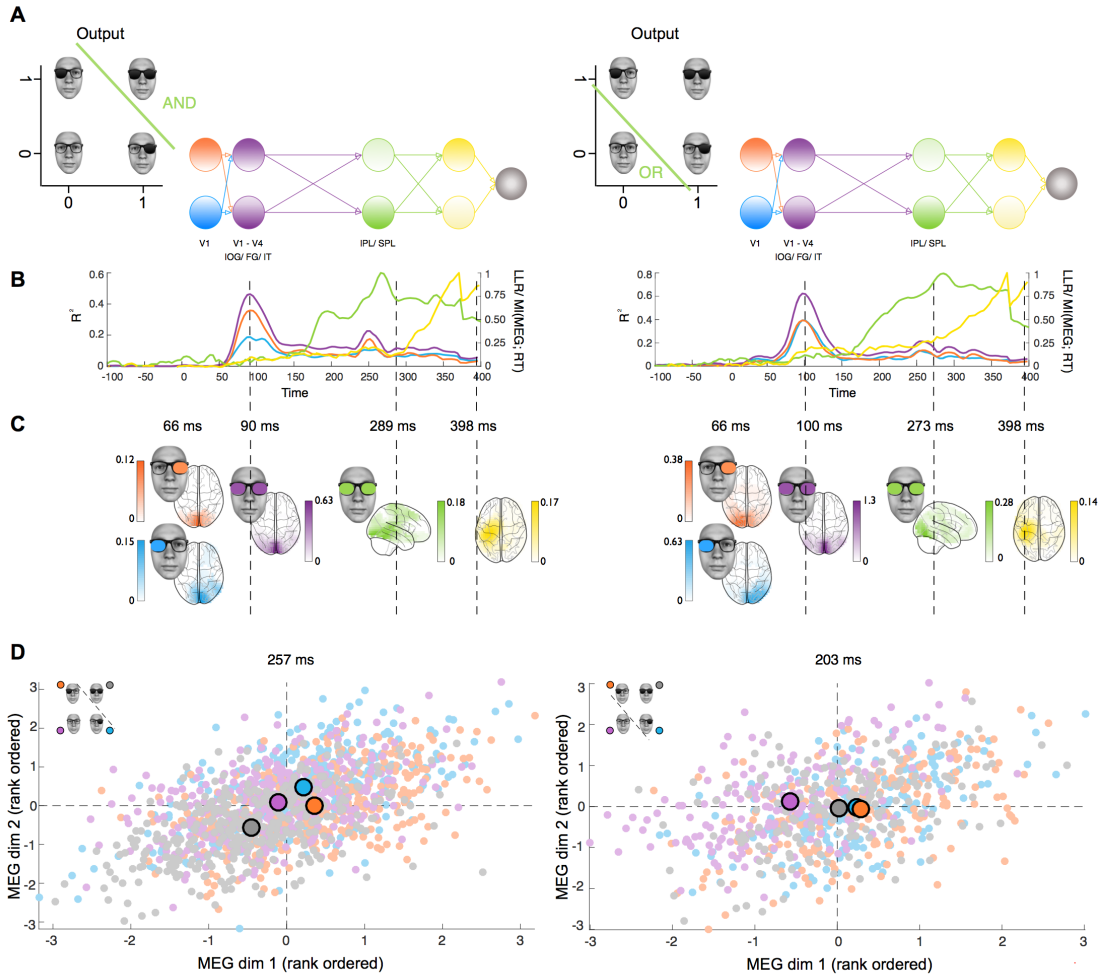

Figure S2. (A) Schematic brain network and hypotheses. Our stimuli comprised a face wearing glasses, with left and right inputs that were either dark or clear, for a total of 4 input conditions for behavioral decisions (illustrated for AND – left, and OR – right). We found a three-stage process of input representation and transformation in brain networks, with specific spatio-temporal dynamics at each stage. At stage 1, V1-4 regions in the hemisphere contra-lateral to input presentation should separately represent each input--i.e. right (vs. left) hemisphere input representation the left (vs. right) color-coded in blue (vs. orange) input. At stage 2, following cross-hemispheric input transfer, ventral and dorsal pathway voxels should represent both inputs. At stage, 3, representations should nonlinearly transform (color-coded in green) to integrate the inputs for behavior. To determine the linear vs. nonlinear input representations at each stage, we applied a data-driven analysis at each voxel and time point and computed a log-likelihood ratio (LLR) between two linear models of input representation—i.e. with and without an interaction term (see Equation 1). **(B) Time courses of maximum linear and nonlinear input representations.** Average (across 10 AND and 10 OR participants) time courses of the maximum (across voxels) R^2 for left (blue), right (orange), and both inputs (magenta) linear representation and LLR indicating nonlinear representations of the two inputs (green); and finally, a representation of reaction times in MEG activity (yellow). Timecourses showing LLR and RT representation were first normalized to 1 within participants. **(C) Summary of localized states of information representation and transformation.** Glass brains depict average (across participants) Mahalanobis distances (see panel D and *Methods, Mahalanobis Distances*) at specific time points: At stage 1 (~60 ms post-stimulus), blue and orange brains localize the group average onset of contra-lateral linear representation of the left and right inputs. At stage 2 (~ 100 ms post-stimulus), purple brains localize the median peak-time average linear representation of both inputs in midline occipital cortex. At stage 3 (289/ 273 ms post-stimulus), the green brains localize the average

nonlinear (integrated) AND/OR representation of both inputs in right occipital-temporal cortex at median peak time. The yellow brains localize representation of reaction times in the left motor cortex ~400 ms. **(D) MEG voxel activity representation of information processing states.** Scatterplots show, color coded by input conditions the single-trial MEG activity (small dots; large dots – input condition average) visualized in two-dimensional reconstruction of single voxels representative of each information processing state (from a representative AND/ OR participant). Note: the face stimulus was averaged from several artificially synthesized faces that did not belong to any real person.

Experiment 2

Experiment 1 demonstrated nonlinear integration in a brain network that received two spatially separated inputs. Experiment 2 included an additional temporal separation to the design. Specifically, each trial started with one lateralized input (i.e. dark or clear), but the other one was greyed out for a 1s duration, following which it became dark or clear (Figure S3 A). Thus, for 1 s, the information of only one of the two inputs was available to participants, following which both were available. We instructed participants (N = 5 in each of the XOR, AND and OR tasks) to wait until both inputs were available to perform their task. However, we were aware that AND and OR could be performed if one of the first inputs was clear (for AND participants responding "not AND") or dark (for OR participants responding "yes OR"), whereas XOR could not, because comparisons of the two sequentially presented input is mandatory in this task.

Consequently, we arranged our analyses in two steps. In the first step, we analyzed the data similarly to Experiment 1, this time splitting the trials according to the single input (i.e. left or right lens) available in the first time window. We then modelled the initial representation of the available left (vs. right) input into MEG activity (on all 5,848 voxels) with independent linear regressions. Next, in the second time window, we modelled the representation of the second lens (i.e. the one that became available after 1s). To establish where, when and how nonlinear integration occurred when the two inputs became available, LLR modelling resumed as in Experiment 1.

Split-trial representation and integration

Over the first 1s, linear representation of the single available left or right input started around 80 ms post-stimulus (with a ~100 ms peak), returning to baseline between 300-400 ms. Following its availability 1s after the first input, the second input was represented for ~1.1s (with a ~1.12s peak, see Figure 3A blue and orange time courses). As in Experiment 1, input representations were again located in occipital, temporal and parietal regions (Figure S3 B, blue and orange).

Nonlinear input integration started ~1.19s in right fronto-temporal cortex (with a ~1.277s peak when the left input was first available; ~1.433s when the right input was first available, Figure S3 A, green time courses). As in Experiment 1, nonlinear integration spanned a widespread network of temporal, parietal and frontal regions past 1.3s (Figure S3 B). Similar results characterized AND and OR participants (Figure S4), with less task-related nonlinear integration in the late time window for AND participants (Figure S4 B).

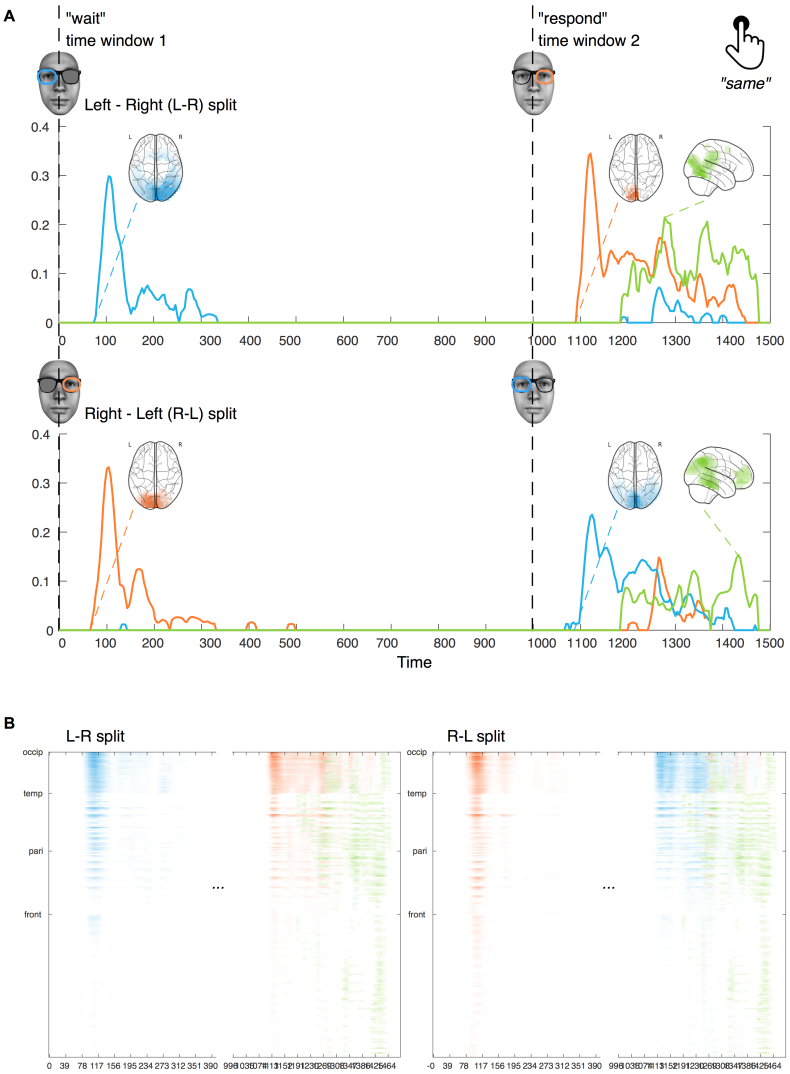

Figure S3. (A) Time courses of linear and nonlinear input representations in Experiment 2. Using the same stimuli as in experiment 1, we presented either the right (top) or the left (bottom) input as grey for the first 1000 ms following stimulus onset. The other input was either clear or dark. After 1000 ms, the grey input changed into either clear or dark. Participants (N = 5) waited until the grey input changed, and then responded "same" or "different" depending on whether the left and the right input matched or mismatched on left/right input opacity. For each split (see also main text), we plotted the maximum (across voxels) of the averaged (across 5 participants) time courses of linear representation of the left (blue) and the right (orange) input, measured with R^2. We also plotted task-relevant nonlinear integration of the two inputs (green) as follows: we computed significant Mahalanobis distances indicating an XOR pattern from single trial MEG activity corresponding to each input condition (see Figure S1 and *Methods, Mahalanobis Distances*) on all voxels and time points in the second time window, and selected those voxels and time points that also showed significant LLR. We then plotted, in green, the maximum (across voxels) of the averaged (across participants) Mahalanobis distances (thresholded with both significant Mahalanobis distances and LLR thresholds). Note: the face stimulus was averaged from several artificially synthesized faces that did not belong to any real person. **(B) Spatiotemporal overview of linear and nonlinear input representations.** Blue and orange colors detail where (i.e. which voxels ordered by region of interest, Y axis) and when (post-stimulus in ms, X axis) voxels linearly represented the left and right inputs (average R^2 across N = 5 participants), and green colors indicate their nonlinear transformations (intersection between LLR and Mahalanobis distances, see panel A; average across N = 5 participants).

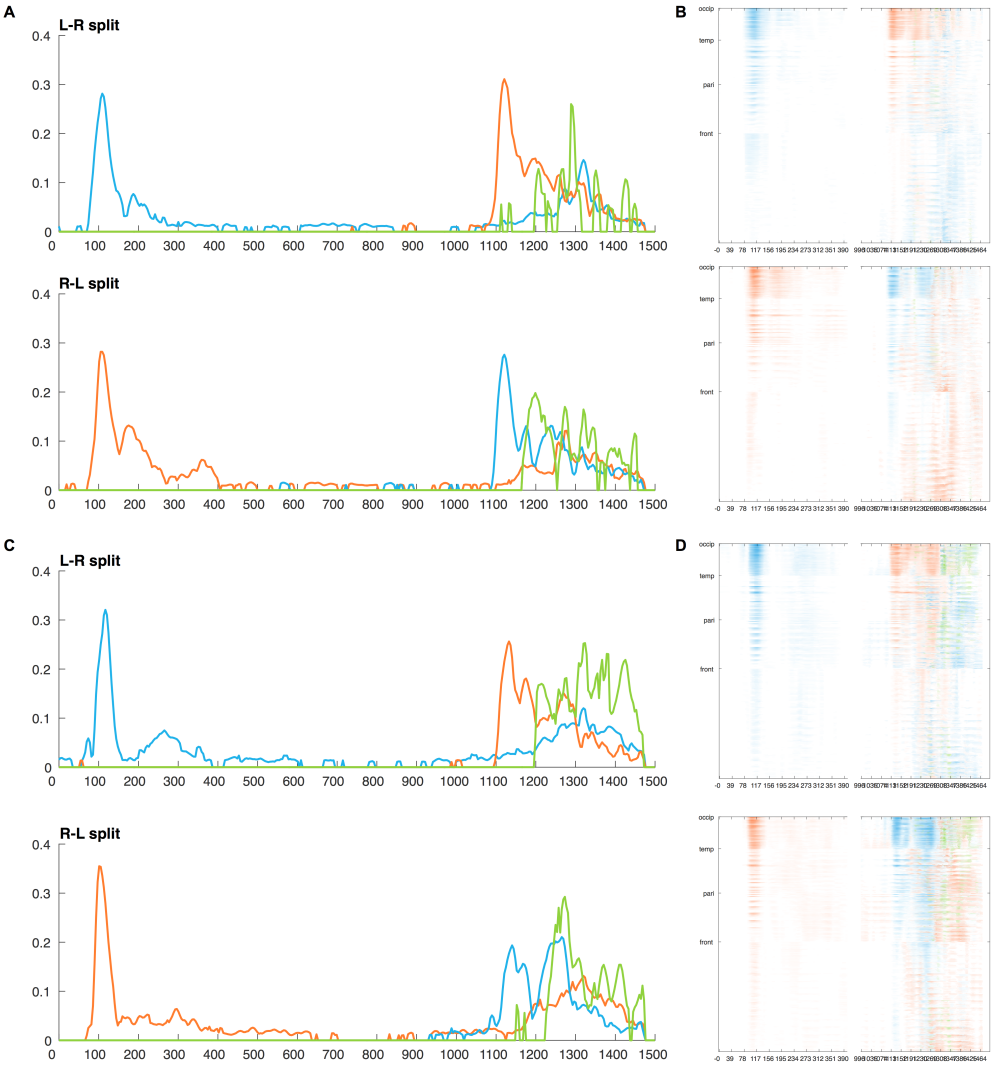

Figure S4. Results from AND (A-B) and OR (C-D) participants. For details please refer to Figure 3 caption.

Integration in the brain, when required by the available input evidence

When the first input enables task-specific strategies, the trials concerned should not require integration of the two inputs. On integrative voxels, we therefore measured the synergistic interactions between the information of the left and right inputs into the MEG activity when both inputs became available (SMI, see *Methods, Information Theoretic Analysis*). An advantage of the information theoretic approach is that we can quantify the contribution of each individual trial to the overall dependence and, as such, investigate whether trials requiring integration would be associated with higher synergy. To do so, we split the trials according to whether the first input was clear or dark (on the left or right, see Figure S5 A). The grey lines (one per participant) in Figure S5 A indicate that AND participants had higher synergy when the first input was dark and still had to be integrated to resolve AND. Relatedly, synergy was higher in OR participants when the first input was clear and still had to be integrated with the second input to resolve OR. As predicted, XOR participants did not show these patterns. The integrative voxels participating in these task-specific strategies were primarily located in the fronto-parietal regions (Figure S5 B).

Further evidence of task-specific strategies involved differential representations of the inputs in the second time window when integration was required. In 4/5 AND participants (grey lines, Figure S5 C), MI representation of the second input was stronger when the first one was dark. In contrast, In OR participants, MI representation of the second input (black lines, Figure S5 C) was stronger when the first input was clear (left input, N = 3/5; right input, N = 5/5). MI representation of the second lens was also stronger in XOR participants (left, 4/5; right, N = 5/5) (red lines, Figure S5 C).

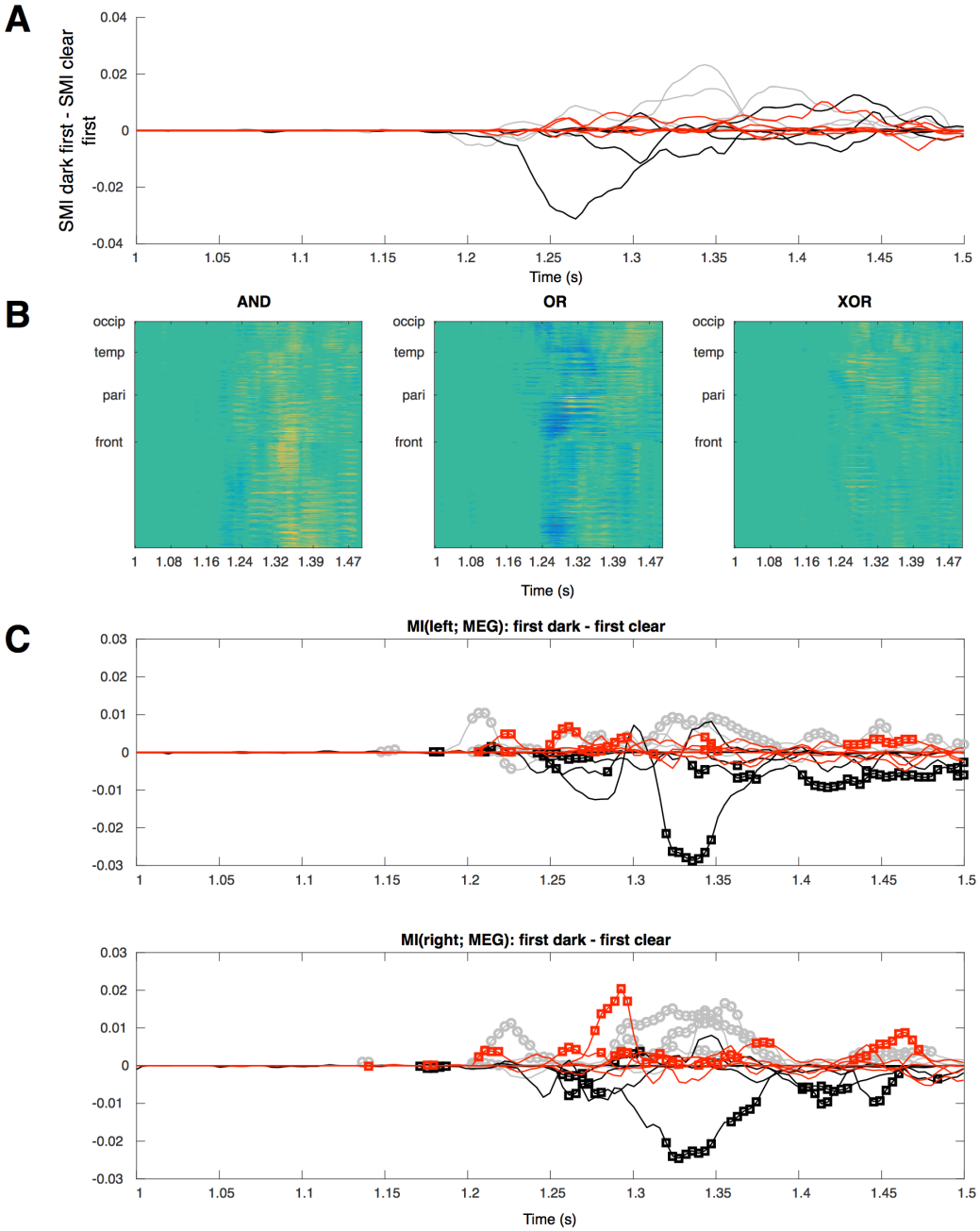

Figure S5. (A) Time courses of SMI difference. Each line corresponds to the mean across voxels per participant (AND – grey; OR – black; XOR – red). Positive values indicate SMI was higher on trials starting with black input presentation, whereas negative values indicate higher SMI on trials starting with clear input presentation. **(B) Spatiotemporal decomposition of SMI differences.** Each plot shows where (i.e. which voxels ordered by region of interest, Y axis) and when (post-stimulus in ms, X axis) the average (across AND, OR, and XOR participants) difference in SMI_{first_dark} - SMI_{first_clear} was strongest. **(C) MI between lens opacity and MEG.** For each AND (grey lines), OR (black lines), and XOR (red lines) participant, we plotted the difference between average MI when the first lens was dark vs. when it was clear, separately for the left (top) and the right (bottom) lens. Each line corresponds to the mean across voxels. Squares overlaid on each line show time points of significant differences between MI preceded by dark lens, and MI preceded by clear lens, computed with a t-test.

Unfolding of representation and integration in the temporal cortex

To better characterize the temporal unfolding of input representation and integration, we now move to region of interest analysis. We picked voxels in the left and right temporal cortex (FG, ITG, MTG, STG) because of their anatomical separation, which overcomes the problem of source mixing in the early occipital cortex. First, we looked at onsets of input coding in contralateral (right for the left input; left for the right input) and ipsilateral (left for the left input; right for the right input) temporal cortices. Previously, using EEG we have shown coding of the eye features first on the hemisphere contralateral to the represented eye (Ince et al., 2016), followed by transfer of the same feature to the ipsilateral hemisphere about 15 ms later and suggestive of network behaviour (Ince et al., 2015; Zhan, Ince, Van Rijsbergen, & Schyns, 2018). Here, we show a similar temporal precedence of input coding in contralateral over ipsilateral cortex (Table S1) that ranged between 4-12 ms regardless of the input or time window, implying inter-hemispheric transfer of input representation that is a prerequisite for integration.

Table S1. Median (across N = 5 participants) onsets of significant R^2 for left or right features in early (0-1000 ms) or late (> 1000 ms) time window, and median of their pairwise differences

		Contralateral	Ipsilateral	Difference (ipsi – contra)
Left	*Time window 1*	82 ms	97 ms	4 ms
	Time window 2	1101 ms	1113 ms	12 ms
Right	*Time window 1*	78 ms	86 ms	4 ms
	Time window 2	1093 ms	1101 ms	4 ms

Generalisation

To test whether the effects observed in Experiments 1 and 2 would generalize to stimuli other than faces, we performed a pilot experiment using spatially separated Gabor patches. Three participants performed an XOR task, responding whether the two patches were the same or different in luminance (i.e. dark or light). Using the same analyses as before we were able to track linear representations of each Gabor patch (on the left or the right; blue and orange time courses in Figure S7). These representations occurred between 50 – 400 ms and were more cyclical than in Experiments 1 and 2. Similarly to previous experiments, the nonlinear integration occurred after 150 ms and peaked between 250 – 400 ms (green time courses in Figure S7) and was followed by representation of reaction times in MEG activity (yellow time courses in Figure S7).

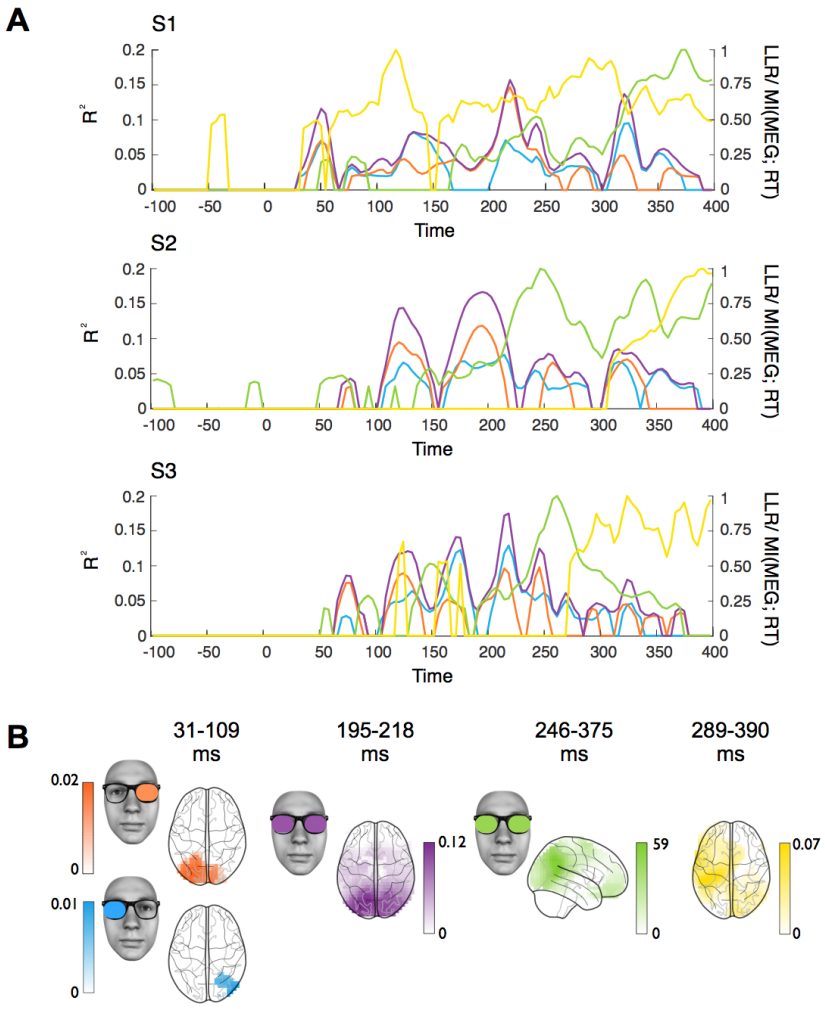

Figure S7. (A) Time courses of representation and integration. For each participant (top, middle and bottom) we plot the maximum across voxels of the representation of the left (blue), right (orange) or both Gabor patches (magenta), followed by their nonlinear integration (green) and representation of reaction times (yellow). **(B) Localization of representation and integration.** Brain topologies of the left (blue) or right (orange) Gabor patch representations were taken at the onset of each time course per participant (ranging between 31 and 109 ms), and averaged across participants. Brain topologies for the representation of both patches (magenta), their nonlinear integration (green) and reaction times (yellow) were computed at the peak of their respective timecourses (numbers above each

brain), and averaged across participants. Note: the face stimulus was averaged from several artificially synthesized faces that did not belong to any real person.