Article:
Delis, I orcid.org/0000-0001-8940-5036, Ince, RAA, Sajda, P et al. (1 more author) (2022) Neural Encoding of Active Multi-Sensing Enhances Perceptual Decision-Making via a Synergistic Cross-Modal Interaction. The Journal of Neuroscience, 42 (11). pp. 2344-2355. ISSN 0270-6474

https://doi.org/10.1523/JNEUROSCI.0861-21.2022

Neural encoding of active multi-sensing enhances perceptual decision-making via a synergistic cross-modal interaction

Ioannis Delis[1*], Robin A.A. Ince[2], Paul Sajda[3,4], Qi Wang[3]

[1]School of Biomedical Sciences, University of Leeds, Leeds, LS2 9JT, UK

[2]School of Psychology and Neuroscience, University of Glasgow, G12 8QQ, UK

[3]Department of Biomedical Engineering, Columbia University, New York, NY, 10027, USA

[4]Data Science Institute, Columbia University, New York, NY, 10027, USA

*Please address correspondence to i.delis@leeds.ac.uk

Abbreviated tile: Neural mechanisms of active multisensory decisions

Competing Interest Statement: The authors declare no competing financial interests.

Acknowledgments

This work was supported by the the European Commission (H2020-MSCA-IF-2018/845884, “NeuCoDe” to I.D.), the Physiological Society (2018 Research Grant Scheme to I.D.), the National Institutes of Health (R01-MH085092 to P.S.), the U.S. Army Research Laboratory (W911NF-10-2-0022 to P.S.), the Wellcome Trust ([214120/Z/18/Z] to R.A.A.I.) and the UK Economic and Social Research Council (ES/L012995/1 to P.S.), and a NARSAD Young Investigator award to Q.W.

No of pages: 32

No of figures: 4

Abstract: 250 words

Significance statement: 120 words

Discussion: 1331 words

Abstract

Most perceptual decisions rely on the active acquisition of evidence from the environment involving stimulation from multiple senses. However, our understanding of the neural mechanisms underlying this process is limited. Crucially, it remains elusive how different sensory representations interact in the formation of perceptual decisions. To answer these questions, we employed an active sensing paradigm coupled with neuroimaging, multivariate analysis and computational modeling to probe how the human brain processes multisensory information to make perceptual judgments. Participants of both sexes actively sensed to discriminate two texture stimuli using visual (V) or haptic (H) information or the two sensory cues together (VH). Crucially, information acquisition was under the participants' control, who could choose where to sample information from and for how long on each trial. To understand the neural underpinnings of this process, we first characterized where and when active sensory experience (movement patterns) is encoded in human brain activity (electroencephalography - EEG) in the three sensory conditions. Then, to offer a neurocomputational account of active multisensory decision formation, we used these neural representations of active sensing to inform a drift diffusion model of decision-making behavior. This revealed a multisensory enhancement of the neural representation of active sensing which led to faster and more accurate multisensory decisions. We then dissected the interactions between the V, H and VH representations using a novel information-theoretic methodology. Ultimately, we identified a synergistic neural interaction between the two unisensory (V, H) representations over contralateral somatosensory and motor locations that predicted multisensory (VH) decision-making performance.

Significance Statement

In real-world settings, perceptual decisions are made during active behaviors, such as crossing the road on a rainy night, and include information from different senses (e.g. car lights, slippery ground). Critically, it remains largely unknown how sensory evidence is combined and translated into perceptual decisions in such active scenarios. Here we address this knowledge gap. First, we show that the simultaneous exploration of information across senses (multi-sensing) enhances the neural encoding of active sensing movements. Second, the neural representation of active sensing modulates the evidence available for decision and, importantly, multi-sensing yields faster evidence accumulation. Finally, we identify a crossmodal interaction in the human brain that correlates with multisensory performance, constituting a putative neural mechanism for forging active multisensory perception.

Introduction

In our daily lives, we make judgments based on noisy or incomplete information that we gather from our environment (Heekeren et al., 2004; Juavinett et al., 2018; Najafi and Churchland, 2018), usually including stimuli from multiple senses (Angelaki et al., 2009; Chandrasekaran, 2017). The acquired sensory information crucially depends on our actions - what we see, hear and touch is influenced by our movements - a process known as active sensing (Schroeder et al., 2010; Yang et al., 2016b). For example, imagine attempting to cross the road on a rainy night. You need to interact with the environment, i.e. turn your head and move your eyes, and process the incoming stimuli (e.g. car lights, slippery ground) to decide whether and when it is safe to do so. If you feel the road is slippery, you may need to monitor your steps and at the same time you may have to walk faster or step back if a car is approaching.

This example indicates that in real-world settings most perceptual decisions are made during active behaviors (Musall et al., 2019). The quality of the acquired evidence is driven by such active behaviors, which, in turn, affect the efficiency of the perceptual decisions that we make as a result of this active sensing process (Yang et al., 2016a; Gottlieb and Oudeyer, 2018). A first crucial element of fast and accurate perceptual decisions is the combination of evidence from different sensory streams (e.g. sight and touch) to form a unified percept and reduce uncertainty about the stimulus (Ernst and Banks, 2002). However, while there is extensive evidence that the integration of information from different sensory modalities improves perceptual choice accuracy (Lewis and Noppeney, 2010; Raposo et al., 2012) and response time (Drugowitsch et al., 2014), multisensory information processing has not been studied in an active scenario, where human participants are allowed to implement their own strategy for gathering evidence - as is the case in real life settings.

Here we addressed this gap in the literature aiming to uncover the neural mechanisms underlying the formation of perceptual decisions via the active

acquisition and processing of multisensory information. To achieve this, we capitalised on our previous work probing the neural correlates of active tactile decisions (Delis et al., 2018) and extended it to a multisensory setting that includes visual and haptic information presented simultaneously or separately. We hypothesised that the neural encoding of active sensory experience would be enhanced when multisensory information was available and that this neural multisensory gain would lead to improvements in decision-making performance.

An important aspect of our study is that the participants had full control of the evolution and duration of each trial. In other words, they could choose how much information to sample, where to sample this information from and for how long. Thus, we first aimed to characterise cortical coupling to continuous active sensing and then combined this with a popular sequential-sampling model of decision-making, the drift diffusion model (DDM) (Ratcliff and McKoon, 2008), to understand how the identified representations of active sensing behaviors influence decisions in the human brain. Here, to bridge the gap between active evidence acquisition and decision formation, we used the neural correlates of active (multi-)sensing to constrain the DDM.

Finally, to quantify cross-modal interactions in the brain, we applied a novel information-theoretic framework named Partial Information Decomposition (PID) (Williams and Beer, 2010; Timme et al., 2014; Ince, 2017). PID quantifies the contribution of a) each sensory modality and b) cross-modal representational interactions ("redundant" or "synergistic") to the multisensory neural representation (Park et al., 2018). Redundancy measures the similarity of the neural representation of the two modalities, while synergy indicates a better prediction of the neural response from both modalities simultaneously. Ultimately, this approach revealed the interactions between representations of different sensing modalities in the brain and shed light onto their role in decision-making behavior.

Materials and Methods

Experimental design and paradigm. Fourteen healthy right-handed participants (8 female, aged 24±2 years) performed a two-alternative forced choice (2AFC) discrimination task during which they had to compare the amplitudes of two sinusoidal stimuli of the same frequency. All experimental procedures have been reviewed and approved by the Institutional Review Board (IRB) at Columbia University.

To generate visual and tactile stimuli that can be actively sensed, we employed a haptic device called a Pantograph (Campion et al., 2005), which can be controlled to generate the sensation of exploring real surfaces (Fig. 1A). The Pantograph is a 2-dimensional force-feedback device, that is, a) it produces a 2D tactile output and b) it simultaneously measures 2D information about the finger position and applied force. Here we used its first property to generate stimulation and the second property to record the kinematics of the movements performed by the participants while they actively explored the presented stimuli. In particular, we split the workspace of the Pantograph (of dimensions 110 mm x 60 mm) into two subspaces (left - L and right - R, 55 mm x 60mm each) and generated continuous sinusoidal stimuli of different amplitudes (but same wavelength of 10 mm) in the two subspaces (Fig. 1B). Then, we instructed the participants to discriminate the amplitude of the two subspaces as quickly and as accurately as possible a) using only visual (V) information, b) using only haptic (H) information and c) combining the two sensory cues (VH). Crucially for our investigation here, participants were free to choose how to explore this virtual texture, i.e., where and how fast to move their fingers and how long to explore each one of the two sides for before making their perceptual choice. Participants placed their right index finger on the interface plate of the Pantograph (see Fig. 1A) and moved it freely to explore the textures of both subspaces (Fig. 1C) before reporting their choice (i.e., which amplitude is higher) by pressing one of two buttons on a keyboard (left or right arrow) using their left hand.

Specifically, in the H condition, the Pantograph produced sinusoidal forces of different intensity between L and R. When the participants placed their index fingers on the plate (interface) of the Pantograph, these forces at the interface had the effect of causing fingertip deformations and thus tactile sensations that resembled exploring real surfaces. Thus, when moving their finger on the Pantograph, participants had the sensation of touching a rough surface (with different amplitudes between L and R – see Fig.1B middle panel). In the V condition, stimuli matching the tactile stimuli were presented on a screen of the same dimensions. More precisely, amplitudes of the sinusoidal virtual texture in H were translated into contrast levels of sinusoidal gratings in V, i.e. the participants were seeing black and white stripes of different intensity/contrast between L and R. Presentation of visual stimuli was generated using Psychtoolbox and visual contrast varied between 0.5 and 1.5 around the default contrast value. The visual angle was $12^{o}\pm6^{o}$. Stimulus presentation was controlled by a real-time hardware system (Matlab XPCTarget) to minimize asynchrony which was <1ms. Importantly, to match the sense of touch, only the part of the workspace corresponding to the participant's finger location was revealed on the screen (i.e. a moving dot following the participant's finger - see Fig. 1B left panel). Thus, in the V condition, grayscale visual textures (of different contrast between L and R) were shown wherever the participants moved their fingers while no forces were applied to the participants' fingers (i.e. no H stimulation). Hence, in both sensory domains, participants could only sense the presented stimulus via active exploration, i.e. finger movements on the x-axis. Accordingly, in the VH condition, both the visual and haptic textures were congruently presented and sensed by the participants using finger movements (Fig. 1B right panel). Overall, participants had to decide whether L or R had higher amplitude based on their haptic (in H trials), visual (in V trials) or visuo-haptic (in VH trials) perception of this virtual surface. Participants reported that they perceived the V and H signals as one stimulus in the VH condition.

The amplitude difference between L and R (representing the difficulty of the task) varied from trial to trial. On each trial, participants compared between the reference amplitude 1 (presented either on the left or right subspace) and one of six other amplitude levels (0.5, 0.75, 0.9, 1.1, 1.25, 1.5). Each trial was initiated by the participant. Trial onset was considered the time point at which horizontal finger velocity exceeded 0. Trial duration was determined by the participant and lasted for the whole period during which the participant made exploratory movements to sense the surface. The trial ended when the participant pressed the < or > key on the keyboard with their left hand to indicate their L or R choice. Each participant performed 20 trials for each amplitude level and for each sensory condition (V, H, VH), resulting $K = 20$ trials x 6 amplitudes x 3 conditions = 360 trials in total. One participant showed poor behavioral performance (accuracy was not significantly different from chance level) and another participant's EEG recordings were significantly contaminated with eye movement artifacts, thus data from these two participants were removed from any subsequent analyses. We report results from the remaining $N = 12$ participants. We also discarded trials in which participants did not respond within 10secs from trial onset or their RTs were shorter than 0.3 seconds. This resulted in the rejection of 4.9% of the trials.

Data recording and pre-processing. During performance of the task, we measured a)_the choice accuracy and response time (RT) of participants' responses, b) movement kinematics (x, y coordinates of finger position recorded by the Pantograph) at a sampling frequency of 1000Hz and c) EEG signals at 2048 sampling frequency using a Biosemi EEG system (ActiveTwo AD-box, 64 Ag-AgCl active electrodes, 10-10 montage).

To compare accuracies and RTs across sensory conditions, we used two-way ANOVAs with factors condition and stimulus difference followed by Bonferroni-corrected post-hoc t-tests. We also fit psychometric curves to the accuracy data

of each participant using a cumulative Gaussian distribution and computed the point of subjective equality (PSE) and slope of the curve at the PSE.

Single-trial movement velocity waveforms were computed using the derivatives of the recorded position. EEG recordings were preprocessed using EEGLab (Delorme and Makeig, 2004) as follows. EEG signals were first down-sampled to 1000Hz to match movement kinematics and dynamics. Then, they were bandpass filtered to 1-50Hz using a Hamming windowed FIR filter. To isolate the purely neural component of the EEG data, we used the following procedure: we first reduced the dimensionality of the EEG data by reconstituting the data using only the top 32 principal components derived from Principal Component Analysis (PCA). Although we record from 64 channels, we expect our recordings to span a considerably lower-dimensional space (as a result of correlations, crosstalk and common sources), thus to enhance the ability of ICA to identify truly independent components, we reduce the data dimensions to half using PCA. Thereafter, an Independent Component Analysis (ICA) decomposition of the data was performed using the Infomax algorithm (Bell and Sejnowski, 1995). We then used an ICA-based artifact removal algorithm called MARA (Winkler et al., 2011) to remove ICs attributed to blinks, horizontal eye movements (HEOG), muscular activity (EMG), and any loose or highly noisy electrodes. MARA assigned each IC a probability of being an artifact; we removed components with probabilities above 0.5.

Decoding finger kinematics from EEG signals. To assess the neural encoding of the participants' active sensory experience in the three sensory conditions, we used a multivariate linear regression analysis introduced in (Di Liberto et al., 2015) and shown in Equation 1 below. As in our previous work (Delis et al., 2018), we hypothesized that the sensorimotor strategy employed by the participant can be represented by the velocity profiles of the participant's exploratory movements which capture changes of movement direction as well as speed changes. Thus, as kinematic feature representing the active sensing behavior, we used 1-d finger velocity on the x-axis (capturing L-R finger movements) but also finger position (on the x-axis) yielded qualitatively very

similar results. Finger movement in the y-axis (which did not provide any sensory information) did not show any significant correlation with the EEG signals and was not considered further. We thus performed a multivariate ridge regression (Crosse et al., 2016) predicting the 1-d finger velocity (on the x-axis) from the EEG data. Specifically, our decoding analysis aimed to reconstruct the movement velocity from a linear combination of the EEG recordings with time lags ranging between [-200ms, 400ms] with respect to the instantaneous velocity values. Specifically, we aimed to decode the velocity profile $s(t)$ of the participants' scanning movements from the simultaneously recorded EEG signals $m(i,t)$, as follows:

$$\hat{s}(t) \cong \sum_{\tau} \sum_{i} g(\tau, i) m(t+\tau, i) \qquad (1)$$

where $\hat{s}(t)$ is the reconstructed finger velocity and $g(i,\tau)$ is a filter that integrates information spatially across EEG channels i and temporally across time lags τ to decode the velocity profile from the EEG recordings. Here we used $\tau \in [-200ms, 400ms]$, i.e. we examined the EEG information about the finger velocity at time t from $t-200ms$ ($200ms$ earlier) up to $t+400ms$ ($400ms$ later). Varying these lags did not improve reconstruction performance and yielded qualitatively similar results with the main effects always in the $[-200ms, 400ms]$ temporal window, so we used this window for all our further analyses. To learn the decoding filters and compute the velocity approximation accuracy (r^2) between the original and the reconstructed velocity profiles, we used the multivariate temporal response function (mTRF) Matlab Toolbox implementing regularised linear (ridge) regression (Crosse et al., 2016). In all our filter estimations, we used a cross-validation procedure. We first randomly split our data into two sets, a training set (80% of the trials) to learn the filters and a test set (the remaining 20% of the trials) to apply the filters to and compute the reported r^2 values. In the training set, we performed 5-fold cross-validation to identify the optimal value of the ridge parameter λ (varying $\lambda = 2^0, \dots, 2^{20}$) that maximizes r^2 between the

293 estimated and the measured velocity. These investigations revealed that values
294 of λ between 2^0 and 2^4 yielded almost identical r^2 across all models, thus we
295 used $\lambda = 2^2$ for all models for consistency.

296 Since the weights of the decoding filters are not interpretable in terms of the
297 neural origins of the underlying processes (Haufe et al., 2014), we transformed
298 them into encoding filters $f(\tau, i)$ using the "forward model" formalism (Parra et
299 al., 2002; Haufe et al., 2014), as follows:

$$f(\tau, i) = \frac{m(t, i)^T m(t, i) g(i, \tau)}{\hat{s}(t)^T \hat{s}(t)} \qquad (2)$$

300 We then plotted the weights of the forward models $f(\tau, i)$ at specific time lags τ
301 as scalp maps to visualise the relationship between sensorimotor behavior and
302 neural activity in each one of the three sensory conditions (V, H, VH). *Statistical*
303 *analysis of EEG-behavior couplings.* To determine statistical significance of the
304 learned EEG-velocity mappings, we randomized the phase spectrum of the EEG
305 signals, which disrupted the temporal relationship between the EEG activity and
306 the kinematics while preserving the autocorrelation structure of the signals
307 (Theiler et al., 1992). We generated 1000 phase-randomized surrogates of the
308 EEG data and computed correlations with the kinematics to define the null
309 distribution from which we estimated p-values. This phase-randomization
310 procedure maintains the magnitude spectrum of the EEG signals, thus
311 conserving their autocorrelation structure, which is a fundamental feature of the
312 original signals when the significance of cross-correlation is assessed. Hence,
313 using this procedure, the obtained surrogates that define the null distribution are
314 a more plausible comparison (resulting in a stricter statistical test) than randomly
315 shuffled surrogates.

316 *Informed modeling of decision-making performance*. Having characterised
317 the cortical coupling to the sensorimotor strategies in the three sensory
318 conditions, we then probed the relationship between the identified EEG-velocity

couplings and decision-making performance. To provide this missing link between active sensing and decision formation, we implemented a Hierarchical Drift Diffusion Model (HDDM), a well-known cognitive model of decision-making behaviour, and informed it with the results of our previous decoding analysis.

We fit the participants' decision-making performance, i.e. accuracy and response time (RT), with a hierarchical drift diffusion model (HDDM) (Wabersich and Vandekerckhove, 2014) which assumes a stochastic accumulation of sensory evidence over time, toward one of two decision boundaries corresponding to correct and incorrect choices (Ratcliff and McKoon, 2008). The model returns estimates of internal components of processing such as the rate of evidence accumulation (drift rate), the distance between decision boundaries controlling the amount of evidence required for a decision (decision boundary), a possible bias towards one of the two choices (starting point) and the duration of non-decision processes (non-decision time), which include stimulus encoding and response production. As per common practice, we assumed that stimulus differences affected the drift rate (Palmer et al., 2005).

In short, the model iteratively adjusts the above parameters to maximize the summed log likelihood of the predicted mean response time (RT) and accuracy. The DDM parameters were estimated in a hierarchical Bayesian framework, in which prior distributions of the model parameters were updated on the basis of the likelihood of the data given the model, to yield posterior distributions (Wiecki et al., 2013; Wabersich and Vandekerckhove, 2014). The use of Bayesian analysis, and specifically the hierarchical drift diffusion model has several benefits relative to traditional DDM analysis. First, this framework supports the use of other variables as regressors of the model parameters to assess relations of the model parameters with other physiological or behavioral data (Frank et al., 2015; Turner et al., 2015; Nunez et al., 2017). This regression model, which is included in HDDM, allows estimation of trial-by-trial influences of a covariate (e.g. a brain measure) onto DDM parameters. In other words, trial-by-

trial fluctuations of the estimated HDDM parameters can be approximated as a linear combination of other trial-by-trial measures of cognitive function (Wiecki et al., 2013; Forstmann et al., 2016). This property of the HDDM enabled us to establish the link between the results of the EEG-velocity coupling analysis and the decision parameters of the model, by using the EEG-velocity couplings as predictors of the HDDM parameters, as explained below (also see Eq. 3 for an example of such a linear regression of the drift rate parameter). Second, the model estimates posterior distributions of the main parameters (instead of deterministic values), which directly convey the uncertainty associated with parameter estimates (Kruschke, 2010). Third, as a result of the above, the hierarchical structure of the model allows estimation of the HDDM parameters across participants and conditions, thus yielding distributions at different levels of the model hierarchy (e.g. the population level and the participant level respectively). In this way, the HDDM capitalizes on the statistical power offered by pooling data across participants (population-level parameters) but at the same time accounts for differences across participants (represented by the variance of the population-level distribution and the individual participant-level estimates). Fourth, the Bayesian hierarchical framework has been shown to be especially effective when the number of observations is low (Ratcliff and Childers, 2015).

To implement the hierarchical DDM, we used the JAGS Wiener module (Wabersich and Vandekerckhove, 2014) in JAGS (Plummer, 2003), via the Matjags interface in Matlab to estimate posterior distributions. For each trial, the likelihood of accuracy and RT was assessed by providing the Wiener first-passage time (WFPT) distribution with the four model parameters (boundary separation, starting point, non-decision time, and drift rate). Capitalizing on the advantages of HDDM, we ran the model pooling data across all participants and conditions and estimated both population-level and participant-level distributions. Parameters were drawn from uniformly distributed priors and were estimated with non-informative mean and standard deviation group priors. As per standard practice for accuracy-coded data, the starting point was set as the midpoint

between the two decision boundaries as participants could not develop a bias towards correct or incorrect choices.. For each model, we ran 3 separate Markov chains with 5500 samples of the posterior parameters each; the first 500 were discarded (as “burn-in”) and the rest were subsampled (“thinned”) by a factor of 50 following the conventional approach to MCMC sampling whereby initial samples are likely to be unreliable due to the selection of a random starting point and neighboring samples are likely to be highly correlated (Wabersich and Vandekerckhove, 2014). The remaining samples constituted the probability distributions of each estimated parameter. To ensure convergence of the chains, we computed the Gelman-Rubin R^2 statistic (which compares within-chain and between-chain variance) and verified that all group-level parameters had an R2 close to 1 and always lower than 1.01.

Here, to obtain a mechanistic account of the effect of EEG-velocity coupling on decision-making behaviour, we incorporated the single-trial measures of these couplings (r^2 values defined above) into the HDDM parameter estimation (Fig. 3B).Specifically, as part of the model fitting within the HDDM framework, we used the single-trial velocity reconstruction accuracies r^2 as regressors of the decision parameters to assess the relationship between trial-to-trial variations in EEG-velocity couplings and each model parameter. Furthermore, to characterise the effect of active sensing movements on decision formation, we also incorporated movement parameters in the HDDM framework. Specifically, we computed the following movement parameters: a) the average finger velocity (v_m) on each trial, b) the number of crossings (n_{cr}) between L and R which is an indicator of the time it took participants to switch between the two stimuli and c) the time participants spent exploring one of the two stimuli (here we arbitrarily selected the low-amplitude stimulus on each trial, t_{low}) as an indicator of exploration time. To understand how these movement parameters affect the decision-making process and specifically whether they relate to a) sensory processing and movement planning/execution (i.e. non-decision

processes) and/or b) evidence accumulation (i.e. decision processes) and/or c) the speed-accuracy trade-off adopted by the participants, we used these parameters as regressors for non-decision time, drift rate and decision boundary, as follows:

$$\tau = \beta_0 + \beta_1 * r^2 + \beta_v * v_m + \beta_{sw} * n_{cr} + \beta_{exp} * t_{low} \quad (3)$$

$$\delta = \gamma_0 + \gamma_1 * r^2 * s + +\gamma_v * v_m + \gamma_{sw} * n_{cr} + \gamma_{exp} * t_{low} \quad (4)$$

$$\alpha = \theta_0 + \vartheta_1 * r^2 + +\vartheta_v * v_m + \vartheta_{sw} * n_{cr} + \vartheta_{exp} * t_{low} \quad (5)$$

where τ, δ, α represent the single-trial non-decision time, drift rate and decision boundary respectively., Velocity reconstruction accuracy r^2, mean finger velocity v_m, number of crossings n_{cr} and time spent exploring the lower amplitude stimulus t_{low} are the single-trial predictor variables with regression coefficients $\beta_i, \gamma_i, \delta_i$ respectively and $s = 0.1{,}0.25{,}0.5$ is the stimulus difference on each trial $k = 1, \dots, K$ of each participant $n = 1, \dots, N$. As per common practice, we modelled a linear relationship between drift rates and stimulus differences reflecting the dependence of the speed of information integration on the amount of evidence available (Palmer et al., 2005; Ratcliff and McKoon, 2008).

By using the above regression approach we were able to test the influence of the above EEG and movement parameters on each of the HDDM parameters. Thus, we tested different models in which the single-trial values of the above parameters were used as predictors for all combinations of the HDDM parameters (drift rate, non-decision time and decision boundary). To select the best-fitting model, we used the Deviance Information Criterion (DIC), a measure widely used for fit assessment and comparison of hierarchical models (Spiegelhalter et al., 2002). DIC selects the model that achieves the best trade-off between goodness-of-fit and model complexity. Lower DIC values favor models with the highest likelihood and least degrees of freedom.

Statistical analysis of modeling results. Posterior probability densities of each regression coefficient were estimated using the sampling procedure described above. Significantly positive (negative) effects were determined when >95% of the posterior density was higher (lower) than 0. To take into account the hierarchical structure of the model which estimated both population-level distributions and participant-level distributions of the parameters, all statistical tests at the population level were performed by contrasting the group-level distributions (not the individual participant means) across sensory conditions. This hierarchical statistical testing has been shown to reduce biases and actually yield conservative effect sizes (Boehm et al., 2018).

Partial Information Decomposition. We then aimed to uncover whether the visual (V) and haptic (H) neural representations of active sensing contained the same information (redundancy) that is present in the multisensory (VH) representation or to what extent their contributions are distinct (unique information) or complementary (synergy). To achieve this, we used the Partial Information Decomposition (PID) (Williams and Beer, 2010; Timme et al., 2014) applied to the predictions of the finger velocity encoding models learned in the different experimental conditions. PID provides an information theoretic approach to compare the outputs of different predictive models that goes beyond simply comparing accuracy to determine whether the different models share or convey unique predictive information content (Daube et al., 2019b). PID extends the concept of co-information (McGill, 1954), which is defined as follows:

$$I(VH;V;H) = I(VH;V) + I(VH;H) - I(VH;[V,H]) \qquad (6)$$

where $I(X;Y)$ denotes the mutual information (MI) between variables X and Y. MI is a nonparametric measure of dependence between two variables which has the unique property that its effect size is additive (Shannon, 1948). Hence, co-information (also called interaction information when defined with opposite sign)

quantifies the difference between the sum of the MI when each modality is considered alone and the MI when the two modalities are observed together (Park et al., 2018).

Positive values of this difference indicate that some information about the predictions of the multisensory VH model is shared between the predictions obtained from the models trained in the unisensory V and H conditions. I.e. there are common or redundant representations of finger velocity in both V and H conditions. Negative values of the interaction information indicate a super-additive or synergistic interaction between the predictions of the V and H models, i.e. the two models provide more information about the multisensory (VH) prediction when observed together than would be expected from observing each individually. However, interaction information measures the net difference between synergy and redundancy in the system, thus it is possible to have zero interaction information, even in the presence of redundant and synergistic interactions that cancel out in the net value (Williams and Beer, 2010; Ince, 2017). This occurs because classic Shannon quantities cannot separate redundant and synergistic contributions, which has led to a growing field developing Partial Information Decomposition measures to address this shortcoming.

To give a simple example of such a case, let us consider 3 variables, each consisting of two bits (i.e. binary (0/1) variables with p(0)=p(1)=0.5). Let also assume that the first bit is shared between all 3 variables and the second bit follows the XOR distribution across the three variables. In this case, there is clear redundancy and synergistic structure, but co-information / interaction information is zero (Griffith and Koch, 2014).

More precisely, PID addresses this methodological problem by decomposing MI into unique redundant and synergistic components, as follows:

$$I(VH;V;H) = I_{uni}(VH;V) + I_{uni}(VH;H) + I_{red}(VH;V,H) + I_{syn}(VH;V,H) \quad (7)$$

where $I_{uni}(VH;V)$ is the part of the VH model predictions that can be explained only from the V model predictions, $I_{uni}(VH;H)$ is the part of the VH model predictions that can be explained only from the H model predctions, $I_{red}(VH;V,H)$ is the part of the VH model predictions that is common (redundant) to both the V and H model predictions and $I_{syn}(VH;V,H)$ is the extra (synergistic) information about the VH model predictions that arises when both V and H predictions are considered together. PID decomposes the joint mutual information between two predictor signals (here the EEG activity predicted from an encoding model trained in the unisensory V, H conditions) and a target signal (here the EEG activity predicted from an encoding model trained in the multisensory VH condition) into four terms: redundancy, the unique information in each predictor, and synergy. Redundancy quantifies the information in the target signal that is shared between the two predictor signals. Synergy quantifies improvement in prediction of the target when both predictors are observed together and represents information about the target signal which cannot be obtained from the individual predictors separately.

To perform PID here, we used a recent implementation based on common change in surprisal for Gaussian variables (Ince, 2017) which has been shown to be effective when applied to neuroimaging data (Park et al., 2018; Daube et al., 2019a).

To implement the above approach on our data, we used the recordings of the VH condition where the two unisensory representations of active sensory experience could be directly compared with the multisensory representation. We took the velocity-encoding models obtained in each condition (V, H, VH) and applied them to the VH data (see Eq. 3) to obtain the V, H and VH predictions of each EEG sensor activity for all VH trials. Since the unisensory models (V, H) were fit in the corresponding unisensory condition, they could only have learned a unisensory representation, whereas the VH model learned a multisensory representation of

active sensing velocity. Thus, we applied PID for each participant separately to predict the VH model predictions from the two unisensory V and H model predictions, which enabled us to quantify the cross-modal interactions between the two unisensory representations across all EEG sensors.

Statistical analysis of PID results. We performed this decomposition independently for each EEG channel and obtained scalp maps for the four PID terms (redundant information, unique information of A, unique information of V, synergistic information) for each participant. To avoid over-fitting, we implemented a 5-fold cross-validation procedure. We randomly split the VH data into 5 subsets used 4 of them to learn the VH, V and H models and the held-out set to perform the PID on. We repeated this process 5 times to obtain PID values for all the VH data. To assess statistical significance of the obtained values, we performed a permutation test. Specifically, we shuffled the target signal, i.e. the VH model of active sensing, 1000 times while keeping the two predictor signals (V and H models respectively) unchanged and applied PID to predict the VH model surrogate data. Output values of the original PID decomposition were considered significant if the exceeded the 99-th percentile of the distribution of the surrogate data. Multiple comparisons were corrected for using FDR (Genovese et al., 2002).

Results

We collected behavioral and EEG data while 14 participants actively interrogated a 2-dimensional texture stimulus that differed in its amplitude in one dimension (left - L vs right – R). Participants used visual information (V), haptic information (H) or both types of sensory information simultaneously (VH) to make a 2-alternative forced perceptual choice (2AFC), i.e., report (via a key press) as quickly and as accurately as possible on which side (L or R) the texture stimulus had higher amplitude (Fig. 1B). To sample information from both sides, participants performed finger movements scanning the workspace of the Pantograph before reaching a decision (Fig. 1C).

In the H condition, the Pantograph (see Materials and Methods for more details on the device used to generate the stimuli) was programmed to produce sinusoidal forces, which yielded the sensation of exploring a rough texture surface (with different amplitudes between L and R (when participants moved their index finger on the workspace of the Pantograph (see Fig.1B middle panel). In the visual domain, participants were moving their fingers to reveal greyscale stripes of different intensity/contrast between L and R (see Fig. 1B left panel).In the VH condition, both the visual and haptic textures were congruently presented wherever the participants moved their fingers (Fig. 1B right panel). Overall, participants had to decide whether L or R had higher amplitude based on their haptic (in H trials), visual (in V trials) or visuo-haptic (in VH trials) perception of this virtual surface.

Multisensory gain in behavioral performance

Multisensory stimulation resulted in significantly higher discrimination accuracy (91.5%±2.1% in VH vs 85.8%±2.2% in V and 86.3%±2.2% in H, two-way ANOVA with factors condition and stimulus difference, $F(2,99)=5.64$, $p<0.005$, see also slopes in the corresponding psychometric curves in Fig. 1D,

$PSE_v = 0.034 \pm 0.013, PSE_h = -0.001 \pm 0.009\ PSE_{vh} = -0.019 \pm 0.007, slope_v = 2.397\ \pm 0.2964, slope_h = 1.826\ \pm 0.147, slope_{vh} = 3.001\ \pm 0.2514$) compared to the unisensory conditions (post-hoc t-tests, Bonferroni corrected, p=0.009 for V-VH and p=0.019 for H-VH). Response times also reduced in VH (4.11±0.30s vs 4.41±0.31s in V and 4.25±0.29s in H, two-way ANOVA with factors condition and stimulus difference, F(2,99)=3.19, *p*=0.045, see also corresponding cumulative distribution functions in the three conditions, Fig. 1E). This result was significant at the population level for VH vs V differences (post-hoc t–test, *p*=0.021, Bonferroni corrected) but not VH vs H differences (post-hoc t-test, *p*=0.066, Bonferroni corrected) in response times. As expected, we also found a main effect of stimulus differences, with accuracy increasing ($F(2) = 91.82$, $p < 0.0001$) and reaction times decreasing ($F(2) = 4.56$, $p < 0.02$) with larger stimulus differences, respectively. There was no interaction between the sensory condition and stimulus difference on either measure (accuracy: $F(4) = 0.66$, $p = 0.62$; reaction times: $F(4) = 0.05$, $p = 0.99$). Taken together, these results indicate that multisensory information increased decision-making performance.

Reconstruction of active sensing velocity from EEG recordings

We then aimed to establish a relationship between brain activity and the active sensory experience of the participants in each one of the three sensory conditions. To this end, we performed a multivariate ridge regression (Crosse et al., 2016) between the EEG data and the 1-d finger velocity data (on the x-axis) to quantify neural encoding of sensorimotor behavior.

This analysis yielded the optimal linear combination of EEG channel activations with time lags ranging between [-200ms, 400ms] that approximated the measured movement velocities. We found that reconstruction accuracy r^2 was above chance level in all sensory conditions (all *p-values*<0.01, Fig. 2B). To obtain interpretable topographies of the neural activity underlying these EEG-velocity couplings, we inverted the obtained velocity-decoding (backward)

models into velocity-encoding (forward) models (Parra et al., 2005; Haufe et al., 2014). This revealed that centro-frontal locations (with positive weights) and occipital locations (with negative weights) contributed most to velocity reconstruction in the three sensory conditions with time lags ranging from 20 to 160 ms - see Fig. 2A showing the scalp topographies of the forward models and Fig. 2C-D showing the corresponding temporal response functions (averaged across frontal and occipital channels respectively) in the three sensory conditions.

Impact of active multi-sensing on the quality of perceptual evidence

To characterise the relationship between the identified EEG-velocity couplings and decision-making performance,, we employed a Hierarchical Drift Diffusion Model (HDDM). In brief, the HDDM decomposes task performance (i.e. accuracy and RT), into internal components of processing representing the rate of evidence integration (drift rate, δ), the amount of evidence required to make a choice (decision boundary separation, α), and the duration of other processes, such as stimulus encoding and response production (non-decision time, τ). Ultimately, by comparing the obtained values of all three core HDDM parameters across the V, H and VH trials, we could associate any behavioral differences resulting from the deployment of multisensory information (more accurate and faster perceptual choices as in Fig. 1) to the constituent internal process reflected by each model parameter.

Here, to obtain a mechanistic account of the formation of perceptual decisions via the active sampling of (multi-)sensory information, we incorporated the single-trial measures of brain-sensing- couplings (r^2 values) into the HDDM parameter estimation (Fig. 3B). Specifically, we applied the obtained decoding filters to the single-trial EEG data and computed velocity reconstruction accuracies for each trial of each sensory condition (using a nested cross-

validation process, see Materials and Methods for more details). Then, as part of the HDDM fitting process, we integrated these single-trial r^2 values in the HDDM framework by using them as regressors of the three core HDDM parameters (drift rate, non-decision time and decision boundary – see Materials and Methods). The corresponding regression coefficients were estimated together with the HDDM parameters thus enabling the assessment of the relationship between trial-to-trial variations in EEG-velocity couplings and each model parameter. We also used as regressors three movement parameters (average velocity v_m, number of crossings between L and R n_{cr} and time spent on the lower amplitude stimulus t_{low}) which served to dissociate the effect of the exploratory movements (captured by these parameters) on decision formation from the effect of the neural encoding of these active sensing movements (captured by r^2).

We found that the best-fitting model (achieving the best complexity-approximation trade-off as evaluated by the Deviance Information Criterion, Fig. 3A) was the one using r^2 as regressor of the drift rate only and n_{cr}, t_{low} as regressors of non-decision time only (Figure 3B shows a graphical illustration of the best-fitting model and Fig. 3C shows the model fitting of the accuracy and RT data where bars represent actual data and lines represent model fits). The means and confidence intervals of the estimated values of the three core HDDM parameters are reported in Table 1. Crucially for our investigation here, the EEG-velocity couplings r^2 were predictive of drift rates in single trials (regression coefficients β_1 were larger than zero for all three sensory conditions, $Prob(\gamma_1(V) > 0) > 0.97$,$Prob(\gamma_1(H) > 0) > 0.99$, $Prob(\gamma_1(VH) > 0) > 0.999$; Fig. 3D). Furthermore, the contribution of r^2 to drift rate was higher in VH trials compared to V and H trials ($Prob\big(\gamma_1(VH) > \gamma_1(V)\big) > 0.95$ and $Prob\big(\gamma_1(VH) > \gamma_1(H)\big) > 0.99$; Fig. 3D) indicating a multisensory enhancement of evidence accumulation rates via an increased weighting of the EEG-velocity couplings in the VH condition.

We then examined whether this multisensory gain could explain the observed improvements in behavioral performance when multisensory information is available. Indeed, this enhanced contribution of r^2 to drift rate was predictive of multisensory improvements in behavioral performance. Specifically, cross-participant differences in β_1's across conditions correlated with the reported increases in accuracy ($r = 0.58,\ p = 0.049$ for VH vs V and $r = 0.75,\ p = 0.005$ for VH vs H; Fig. 3F), suggesting that differences in accuracies across participants were accounted for by the contributions of EEG-velocity couplings to evidence accumulation. Thus, participants with greater drift rate amplification achieved stronger enhancements in their behavioral performance as a result of multisensory information available.

We also found that both switching time between the two stimuli as captured by n_{cr} and exploration time spent on one of the two stimuli as captured by t_{low} were predictive of non-decision time ($Prob(\beta_{sw} > 0) > 0.999$, $Prob\big(\beta_{exp} > 0\big) > 0.999$ for all V,H,VH; Fig. 3G-H) in single trials indicating that non-decision processes (i.e. related to sensory processing and movement planning/execution) are dependent on switching and exploration times. There was a positive cross-participant correlation ($r = 0.695, p = 0.0121$) between β_{exp} and RT (averaged across trials and sensory conditions) suggesting that participants with larger contributions of exploration time to their non-decision times took longer to respond (Fig 3I). However, we found no reliable difference in the corresponding regression coefficients (β_{sw}, β_{exp}) between the three sensory conditions ($Prob\big(\beta_{sw}(VH) > \beta_{sw}(V)\big) = 0.632$, $Prob\big(\beta_{sw}(VH) > \beta_{sw}(H)\big) = 0.843$, $Prob\left(\beta_{exp}(VH) > \beta_{exp}(V)\right) = 0.107$, $Prob\left(\beta_{exp}(VH) > \beta_{exp}(H)\right) = 0.210$;Fig. 3G-H). There was also no difference in the decision boundaries in the three sensory conditions ($Prob\big(\alpha(VH) > \alpha(V)\big) = 0.731$, $Prob\big(\alpha(VH) > \alpha(H)\big) > 0.804$; Fig. 3E). These results indicate that neither the switching and exploration times

nor the amount of evidence required to make a decision were dependent on the sensory condition.

Quantifying multisensory interactions

Having established that the neural encoding of the behavioral kinematics is related to the multisensory gain in decision evidence, we then aimed to assess how the neural representations of the two unisensory stimuli (V, H) interact to form a multisensory representation. To this end, we employed Partial Information Decomposition, which enables the quantification of cross-modal representational interactions in the human brain (see Methods for details). Specifically, the PID information theoretic framework quantifies the degree to which a) each unisensory (V,H) representation contributes uniquely to the encoding of active sensing behavior (unique V or H information), b) the two unisensory (V,H) representations share information about active sensing (redundancy) and c) the two unisensory (V,H) representations convey more information when observed simultaneously (synergy). Here, we used PID to predict the forward (velocity-encoding) VH model (target signal) from the two unisensory forward models V and H (predictor signals). The decomposition revealed that the V model provided unique information in right parieto-temporal locations whereas the H model contributed uniquely in left prefrontal and parieto-occipital locations (Fig. 4A, all *p-values*<0.01, FDR corrected). Crucially, we also found multisensory interactions in the form of a) redundant effects in left prefrontal and parieto-occipital electrodes and b) synergistic effects over left centro-parietal scalp (Fig. 4A, all *p*s<0.01, FDR corrected). Here, a redundant interaction means that the representation of velocity is common to both the V and H modalities (Ince et al., 2017; Park et al., 2018). A synergistic interaction means a better prediction of the modelled multisensory response can be made when considering both the V and the H representations together (rather than independently). That is,

knowledge of the simultaneous combination of the EEG signal predicted by V and H models gives more information about the VH EEG signal.

Multisensory accuracy scales with synergistic interactions

Next we investigated the behavioral relevance of the identified cross-modal interactions. In particular, we asked whether the identified synergistic representation of the two modalities was predictive of behavioral performance across participants. Indeed, we found a significant positive correlation (Pearson's R = 0.75 and 0.72, all $p < 0.01$) between synergy in both significant channels (CP3 and C5) and accuracy in VH, suggesting that participants with more synergistic representations at left centro-parietal electrodes achieved better multisensory performance (Fig. 4B). This result suggests that synergy in contralateral centro-parietal EEG signals modulates multisensory decision-making behavior. Due to small sample size we cannot be sure this finding will generalise, but nonetheless report it as an interesting exploratory finding.

Discussion

In this work, we coupled neural decoding of continuous sensorimotor behavior with modeling of decision-making performance and a quantitative assessment of crossmodal neural interactions to understand how the human brain forms perceptual decisions via the active acquisition of multisensory evidence. We showed that the neural encoding of active sensing modulates the decision

749 evidence regardless of the sensing modality. We further demonstrated that the
750 simultaneous sensing of different modalities enhances this neural coupling and
751 this enhancement drives the dynamics of active multisensory decisions. We
752 finally dissected the neural information conveyed by cross-modal interactions and
753 identified a potential neural mechanism supporting multisensory decisions.

754 Recent research on active sensing uncovered the strategies implemented by
755 humans to sample sensory information (Yang et al., 2016b). Here we
756 investigated this active sensing approach in a decision-making task using a
757 computational approach which decodes the neural activity that encodes
758 movement kinematics. Crucially, we made a first step in broadening this line of
759 research to a) include sensory information from multiple modalities and b) reveal
760 its neural underpinnings. These two developments enabled us to uncover the
761 different sensory representations of active sampling behavior in the human brain.

762 To achieve this, we implemented an informed cognitive modeling approach that
763 linked the neural correlates and the movement characteristics of active sensing
764 behavior with the cognitive processes involved in decision-making. Specifically,
765 we asked if decision-making depends on the neural representations of active
766 (multi-)sensing. To answer this question, we used a single-trial measure of the
767 neural encoding of active sensing behavior as predictor of decision-making
768 performance and found that, indeed, trial-to-trial fluctuations of the neural
769 representations of active sensing are predictive of the rate of evidence
770 accumulation for all three sensory conditions (V, H, VH). Crucially, we showed
771 that the multisensory (VH) representation of active sensing was a stronger
772 predictor of drift rate (Figure 3D) thus offering a neural link between active multi-
773 sensing and perceptual decision-making. We also split the motion profile into its
774 two main components, i.e. a) switching between the two alternative stimuli and b)
775 exploration within one particular stimulus and demonstrated that both
776 components were predictive of the duration of non-decision processes (Figure
777 3G-H), thus simply reflecting the time spent for movement planning and

execution and the consequent acquisition and encoding of sensory information. These novel findings were only made possible by the use of an active multi-sensing paradigm in a decision-making task and the joint cognitive modeling of behavioral, neural and sensorimotor signals.

We then capitalized on the identified neural representations of active (multi-sensing), to dissect cross-modal interactions in the human brain. To this end, we employed PID, a recently developed rigorous methodology for the quantification of information conveyed uniquely or jointly by different neural representations (Williams and Beer, 2010; Timme et al., 2014; Ince, 2017). PID further distinguishes between two types of interactions between the neural representations of the two sensory modalities (V, H). A synergistic interaction indicates that a better prediction of the multisensory neural response can be made when the predicted values of the unimodal forward models for V and H are considered jointly rather than independently. Our results suggest that this synergistic interaction of the two neural representations correlates with multisensory behavioural performance (Figure 4B). Instead, a redundant interaction indicates that the two unimodal models provide the same information about the multisensory condition, thus the multisensory response there is common to both modalities (Park et al., 2018; Daube et al., 2019a). This suggests that the underlying neural signals reflect a modality-invariant representation.

As a result of this analysis, we were able to identify neural signals representing these two types of interactions. Specifically, we found that EEG channels in (parieto-)occipital and prefrontal areas carried redundant representations of the two sensory streams, perhaps reflecting supramodal coding mechanisms of active sensing (Figure 4A, redundancy). This finding is in line with previous research assigning a multimodal role to occipital cortex (Lacey et al., 2007; Murray et al., 2016) and suggesting that multisensory enhancements originate from the sensory cortices (Kayser and Logothetis, 2007; Lakatos et al., 2007;

Lewis and Noppeney, 2010). Specifically, recent research involved the visual cortex in audio-visual interactions (Mishra et al., 2007; Cao et al., 2019; Rohe et al., 2019) as well as tactile perception and visuo-haptic interactions (Lucan et al., 2010; Sathian, 2016; Gaglianese et al., 2020). In agreement with the above, here we also found unique H information in parieto-occipital electrodes. Concerning the prefrontal cortex (PFC), recent evidence assigned to it a modality-general role in arbitrating between segregation or fusion of sensory evidence from different modalities (Cao et al., 2019). Thus, the involvement of the PFC in the regulation of adaptive multisensory behaviors in general (Koechlin and Summerfield, 2007; Donoso et al., 2014; Tomov et al., 2018) and perceptual decisions in particular (Heekeren et al., 2006; Philiastides et al., 2011; Rahnev et al., 2016; Sterzer, 2016) makes it a likely contributor to the formation of the most appropriate sensory representation that drives decision-making behavior. In other words, the PFC may support a mechanism gauging candidate (multisensory or unisensory) representations for selecting among multiple strategies to solve the task at hand (Calvert, 2001; Hein et al., 2007; Noppeney et al., 2010; Cao et al., 2019). Our active multi-sensing task requires participants to continuously weigh different sensing strategies and refine their scanning patterns to maximize information gain. Hence, the PFC may capitalise on multisensory information (when of benefit) to support such flexible behavior striking a balance between sampling more evidence and committing to a choice.

The above findings are consistent with our previous study focusing on the tactile modality, which attributed a sensory processing function to occipital cortex (specifically localized to the lateral occipital complex) and a decision formation function to right prefrontal cortex (middle frontal gyrus) (Delis et al., 2018). Taken together with the current results, our findings suggest these two brain areas may play a crossmodal role in supporting active perception and decision-making. Overall, our work adds to the existing literature on multi-sensory interactions by quantifying how sensory representations interact to encode active sensing behaviors.

More importantly, here we revealed a novel functional role for contralateral centro-parietal signals in active visuo-haptic decisions. We found that brain signals over left centro-parietal scalp locations showed stronger encoding of active sensing when the two sensory streams were available (Figure 4A, synergy), thus possibly representing a neural mechanism of multisensory integration. In line with the ongoing debate on the multisensory nature of primary sensory cortices (Ghazanfar and Schroeder, 2006; Liang et al., 2013), cross-modal visuo-haptic interactions leading to enhanced neural representations have been found in the primary somatosensory cortex (S1) (Zhou and Fuster, 2000; Dionne et al., 2010). Here we further characterised these interactions as carrying super-additive/synergistic representations of the active multi-sensory experience and demonstrated that they are related to the accuracy of active multisensory judgments.

It is also worth noting that our results do not rule out the possibility that other brain areas – not directly related to active sensing - may contribute to regulating the speed and accuracy of active multisensory decisions. In fact, recent research breakthroughs have explained the development of multisensory representations from different sensory streams in the human brain (Aller and Noppeney, 2019; Cao et al., 2019; Rohe et al., 2019). Furthermore, recent studies have started to investigate how the interactions between sensory representations shape decision formation (Bizley et al., 2016; Franzen et al., 2020; Mercier and Cappe, 2020).

Our primary aim here was to provide the missing link between the active acquisition of multisensory evidence and its transformation to choice. Overall, our findings validated the hypotheses that a) active sensing guides decision formation via evidence sampling and accumulation and b) multisensory information spurs perceptual decisions by enhancing the neural encoding of active behaviors. Our information-theoretic analysis also revealed the neural substrates of multisensory interactions in the human brain that support active multisensory perception. Ultimately, we identified and characterised a set of

human brain signals that underpin multisensory judgements by subserving an enhancement of the neural encoding of active perception when multisensory information is available.

References

Aller M, Noppeney U (2019) To integrate or not to integrate: Temporal dynamics of hierarchical Bayesian causal inference. PLoS Biol 17:e3000210.

Angelaki DE, Gu Y, DeAngelis GC (2009) Multisensory integration: psychophysics, neurophysiology, and computation. Curr Opin Neurobiol 19:452-458.

Bell AJ, Sejnowski TJ (1995) An information-maximization approach to blind separation and blind deconvolution. Neural Comput 7:1129-1159.

Bizley JK, Jones GP, Town SM (2016) Where are multisensory signals combined for perceptual decision-making? Curr Opin Neurobiol 40:31-37.

908 Boehm U, Marsman M, Matzke D, Wagenmakers EJ (2018) On the importance of avoiding
909 shortcuts in applying cognitive models to hierarchical data. Behav Res Methods
910 50:1614-1631.
911 Calvert GA (2001) Crossmodal processing in the human brain: insights from functional
912 neuroimaging studies. Cereb Cortex 11:1110-1123.
913 Campion G, Wang Q, Hayward V (2005) The Pantograph Mk-II: A haptic instrument. 2005
914 Ieee/Rsj International Conference on Intelligent Robots and Systems, Vols 1-4:723-728.
915 Cao Y, Summerfield C, Park H, Giordano BL, Kayser C (2019) Causal Inference in the Multisensory
916 Brain. Neuron 102:1076-1087 e1078.
917 Chandrasekaran C (2017) Computational principles and models of multisensory integration. Curr
918 Opin Neurobiol 43:25-34.
919 Crosse MJ, Di Liberto GM, Bednar A, Lalor EC (2016) The Multivariate Temporal Response
920 Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous
921 Stimuli. Frontiers in Human Neuroscience 10.
922 Daube C, Ince RAA, Gross J (2019a) Simple Acoustic Features Can Explain Phoneme-Based
923 Predictions of Cortical Responses to Speech. Curr Biol 29:1924-1937 e1929.
924 Daube C, Giordano BL, Ince RAA, Gross J (2019b) Quantitatively Comparing Predictive Models
925 with the Partial Information Decomposition. In: 2019 Conference on Cognitive
926 Computational Neuroscience. Berlin, Germany.
927 Delis I, Dmochowski JP, Sajda P, Wang Q (2018) Correlation of neural activity with behavioral
928 kinematics reveals distinct sensory encoding and evidence accumulation processes
929 during active tactile sensing. Neuroimage 175:12-21.
930 Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG
931 dynamics including independent component analysis. Journal of Neuroscience Methods
932 134:9-21.
933 Di Liberto GM, O'Sullivan JA, Lalor EC (2015) Low-Frequency Cortical Entrainment to Speech
934 Reflects Phoneme-Level Processing. Current Biology 25:2457-2465.
935 Dionne JK, Meehan SK, Legon W, Staines WR (2010) Crossmodal influences in somatosensory
936 cortex: Interaction of vision and touch. Hum Brain Mapp 31:14-25.
937 Donoso M, Collins AG, Koechlin E (2014) Human cognition. Foundations of human reasoning in
938 the prefrontal cortex. Science 344:1481-1486.
939 Drugowitsch J, DeAngelis GC, Klier EM, Angelaki DE, Pouget A (2014) Optimal multisensory
940 decision-making in a reaction-time task. Elife 3.
941 Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically
942 optimal fashion. Nature 415:429-433.
943 Forstmann BU, Ratcliff R, Wagenmakers EJ (2016) Sequential Sampling Models in Cognitive
944 Neuroscience: Advantages, Applications, and Extensions. Annu Rev Psychol 67:641-666.
945 Frank MJ, Gagne C, Nyhus E, Masters S, Wiecki TV, Cavanagh JF, Badre D (2015) fMRI and EEG
946 predictors of dynamic decision parameters during human reinforcement learning. J
947 Neurosci 35:485-494.
948 Franzen L, Delis I, De Sousa G, Kayser C, Philiastides MG (2020) Auditory information enhances
949 post-sensory visual evidence during rapid multisensory decision-making. Nat Commun
950 11:5440.
951 Gaglianese A, Branco MP, Groen IIA, Benson NC, Vansteensel MJ, Murray MM, Petridou N,
952 Ramsey NF (2020) Electrocorticography Evidence of Tactile Responses in Visual Cortices.
953 Brain Topogr 33:559-570.

954 Genovese CR, Lazar NA, Nichols T (2002) Thresholding of statistical maps in functional
955 neuroimaging using the false discovery rate. Neuroimage 15:870-878.
956 Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? Trends Cogn Sci
957 10:278-285.
958 Gottlieb J, Oudeyer PY (2018) Towards a neuroscience of active sampling and curiosity. Nat Rev
959 Neurosci 19:758-770.
960 Griffith V, Koch C (2014) Quantifying Synergistic Mutual Information. In: Guided Self-
961 Organization: Inception. Emergence, Complexity and Computation (M. P, ed). Berlin,
962 Heidelberg: Springer.
963 Haufe S, Meinecke F, Gorgen K, Dahne S, Haynes JD, Blankertz B, Biessmann F (2014) On the
964 interpretation of weight vectors of linear models in multivariate neuroimaging.
965 Neuroimage 87:96-110.
966 Heekeren HR, Marrett S, Bandettini PA, Ungerleider LG (2004) A general mechanism for
967 perceptual decision-making in the human brain. Nature 431:859-862.
968 Heekeren HR, Marrett S, Ruff DA, Bandettini PA, Ungerleider LG (2006) Involvement of human
969 left dorsolateral prefrontal cortex in perceptual decision making is independent of
970 response modality. Proceedings of the National Academy of Sciences of the United
971 States of America 103:10023-10028.
972 Hein G, Doehrmann O, Muller NG, Kaiser J, Muckli L, Naumer MJ (2007) Object familiarity and
973 semantic congruency modulate responses in cortical audiovisual integration areas. J
974 Neurosci 27:7881-7887.
975 Ince RA, Giordano BL, Kayser C, Rousselet GA, Gross J, Schyns PG (2017) A statistical framework
976 for neuroimaging data analysis based on mutual information estimated via a gaussian
977 copula. Hum Brain Mapp 38:1541-1573.
978 Ince RAA (2017) Measuring multivariate redundant information with pointwise common change
979 in surprisal. Entropy 19:318.
980 Juavinett AL, Erlich JC, Churchland AK (2018) Decision-making behaviors: weighing ethology,
981 complexity, and sensorimotor compatibility. Curr Opin Neurobiol 49:42-50.
982 Kayser C, Logothetis NK (2007) Do early sensory cortices integrate cross-modal information?
983 Brain Struct Funct 212:121-132.
984 Koechlin E, Summerfield C (2007) An information theoretical approach to prefrontal executive
985 function. Trends Cogn Sci 11:229-235.
986 Kruschke JK (2010) What to believe: Bayesian methods for data analysis. Trends in Cognitive
987 Sciences 14:293-300.
988 Lacey S, Campbell C, Sathian K (2007) Vision and touch: multiple or multisensory representations
989 of objects? Perception 36:1513-1521.
990 Lakatos P, Chen CM, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and
991 multisensory interaction in primary auditory cortex. Neuron 53:279-292.
992 Lewis R, Noppeney U (2010) Audiovisual synchrony improves motion discrimination via
993 enhanced connectivity between early visual and auditory areas. J Neurosci 30:12329-
994 12339.
995 Liang M, Mouraux A, Hu L, Iannetti GD (2013) Primary sensory cortices contain distinguishable
996 spatial patterns of activity for each sense. Nat Commun 4:1979.
997 Lucan JN, Foxe JJ, Gomez-Ramirez M, Sathian K, Molholm S (2010) Tactile shape discrimination
998 recruits human lateral occipital complex during early perceptual processing. Hum Brain
999 Mapp 31:1813-1821.

McGill WJ (1954) Multivariate information transmission. Psychometrika:97-11.
Mercier MR, Cappe C (2020) The interplay between multisensory integration and perceptual decision making. Neuroimage 222:116970.
Mishra J, Martinez A, Sejnowski TJ, Hillyard SA (2007) Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. J Neurosci 27:4120-4131.
Murray MM, Thelen A, Thutd G, Romei V, Martuzzi R, Matusz PJ (2016) The multisensory function of the human primary visual cortex. Neuropsychologia 83:161-169.
Musall S, Urai AE, Sussillo D, Churchland AK (2019) Harnessing behavioral diversity to understand neural computations for cognition. Curr Opin Neurobiol 58:229-238.
Najafi F, Churchland AK (2018) Perceptual Decision-Making: A Field in the Midst of a Transformation. Neuron 100:453-462.
Noppeney U, Ostwald D, Werner S (2010) Perceptual decisions formed by accumulation of audiovisual evidence in prefrontal cortex. J Neurosci 30:7434-7446.
Nunez MD, Vandekerckhove J, Srinivasan R (2017) How attention influences perceptual decision making: Single-trial EEG correlates of drift-diffusion model parameters. Journal of Mathematical Psychology 76:117-130.
Palmer J, Huk AC, Shadlen MN (2005) The effect of stimulus strength on the speed and accuracy of a perceptual decision. J Vis 5:376-404.
Park H, Ince RAA, Schyns PG, Thut G, Gross J (2018) Representational interactions during audiovisual speech entrainment: Redundancy in left posterior superior temporal gyrus and synergy in left motor cortex. PLoS Biol 16:e2006558.
Parra L, Alvino C, Tang A, Pearlmutter B, Yeung N, Osman A, Sajda P (2002) Linear spatial integration for single-trial detection in encephalography. Neuroimage 17:223-230.
Parra LC, Spence CD, Gerson AD, Sajda P (2005) Recipes for the linear analysis of EEG. Neuroimage 28:326-341.
Philiastides MG, Auksztulewicz R, Heekeren HR, Blankenburg F (2011) Causal role of dorsolateral prefrontal cortex in human perceptual decision making. Curr Biol 21:980-983.
Plummer M (2003) JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling In: 3rd International Workshop on Distributed Statistical Computing.
Rahnev D, Nee DE, Riddle J, Larson AS, D'Esposito M (2016) Causal evidence for frontal cortex organization for perceptual decision making. Proc Natl Acad Sci U S A 113:6059-6064.
Raposo D, Sheppard JP, Schrater PR, Churchland AK (2012) Multisensory decision-making in rats and humans. J Neurosci 32:3726-3735.
Ratcliff R, McKoon G (2008) The diffusion decision model: Theory and data for two-choice decision tasks. Neural Computation 20:873-922.
Ratcliff R, Childers R (2015) Individual Differences and Fitting Methods for the Two-Choice Diffusion Model of Decision Making. Decision (Wash D C) 2015.
Rohe T, Ehlis AC, Noppeney U (2019) The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. Nat Commun 10:1907.
Sathian K (2016) Analysis of haptic information in the cerebral cortex. J Neurophysiol 116:1795-1806.
Schroeder CE, Wilson DA, Radman T, Scharfman H, Lakatos P (2010) Dynamics of Active Sensing and perceptual selection. Curr Opin Neurobiol 20:172-176.
Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379-423 \& 623-656.

1045 Spiegelhalter DJ, Best NG, Carlin BR, van der Linde A (2002) Bayesian measures of model
1046 complexity and fit. Journal of the Royal Statistical Society Series B-Statistical
1047 Methodology 64:583-616.
1048 Sterzer P (2016) Moving forward in perceptual decision making. Proc Natl Acad Sci U S A
1049 113:5771-5773.
1050 Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer JD (1992) Testing for Nonlinearity in Time-
1051 Series - the Method of Surrogate Data. Physica D 58:77-94.
1052 Timme N, Alford W, Flecker B, Beggs JM (2014) Synergy, redundancy, and multivariate
1053 information measures: an experimentalist's perspective. J Comput Neurosci 36:119-140.
1054 Tomov MS, Dorfman HM, Gershman SJ (2018) Neural Computations Underlying Causal Structure
1055 Learning. J Neurosci 38:7143-7157.
1056 Turner BM, van Maanen L, Forstmann BU (2015) Informing Cognitive Abstractions Through
1057 Neuroimaging: The Neural Drift Diffusion Model. Psychological Review 122:312-336.
1058 Wabersich D, Vandekerckhove J (2014) Extending JAGS: A tutorial on adding custom
1059 distributions to JAGS (with a diffusion model example). Behavior Research Methods
1060 46:15-28.
1061 Wiecki TV, Sofer I, Frank MJ (2013) HDDM: Hierarchical Bayesian estimation of the Drift-
1062 Diffusion Model in Python. Front Neuroinform 7:14.
1063 Williams PL, Beer RD (2010) Nonnegative Decomposition of Multivariate Information.
1064 arXiv:10042515v1.
1065 Winkler I, Haufe S, Tangermann M (2011) Automatic classification of artifactual ICA-components
1066 for artifact removal in EEG signals. Behav Brain Funct 7:30.
1067 Yang SC, Lengyel M, Wolpert DM (2016a) Active sensing in the categorization of visual patterns.
1068 Elife 5.
1069 Yang SCH, Wolpert DM, Lengyel M (2016b) Theoretical perspectives on active sensing. Current
1070 Opinion in Behavioral Sciences 11:100-108.
1071 Zhou YD, Fuster JM (2000) Visuo-tactile cross-modal associations in cortical somatosensory cells.
1072 Proc Natl Acad Sci U S A 97:9777-9782.

1073

1074

Figures and Tables

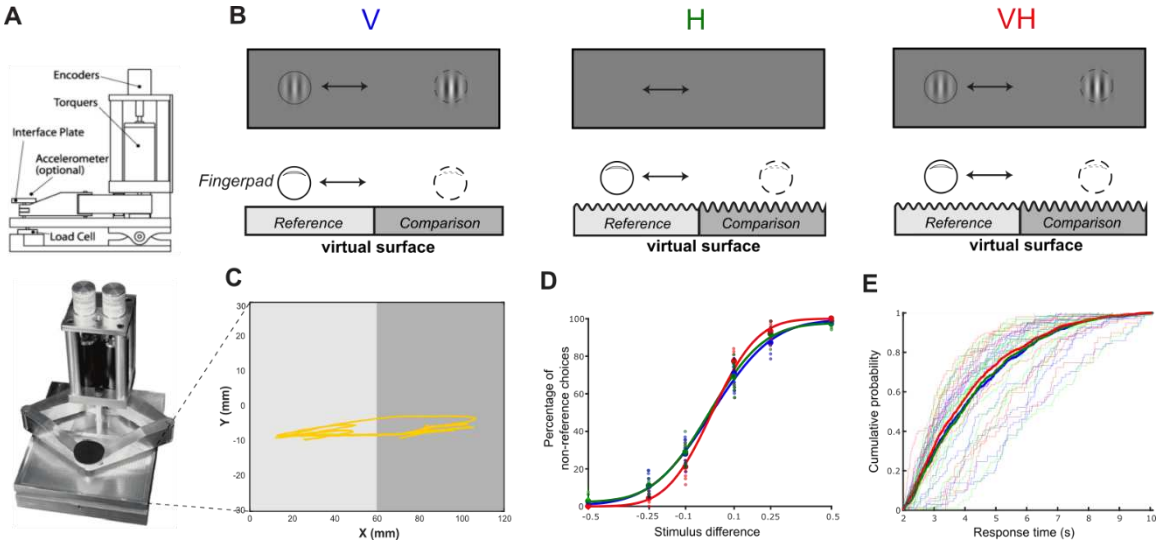

Figure 1. Experimental design and behavioral results. A. The Pantograph is a haptic device used to render virtual surfaces that can be actively sensed. Top: the parts of the Pantograph shown from a lateral view. Participants placed their index finger on the interface plate. Bottom: The Pantograph device used in this experiment. B. The stimulus in the three sensory conditions. We programmed the Pantograph to generate a virtual grating texture. The workspace was split into two subspaces (left - L and right - R) that differed in the amplitude of the virtual surface that the participants actively sensed. One of the two sides (randomly assigned) had the reference amplitude (equal to 1) and the other had the comparison amplitude that varied on each trial taking one of the values: 0.5, 0.75, 0.9, 1.1, 1.25, and 1.5. Participants performed the task using visual information only (V), haptic information only (H) or the two sensory streams together (VH). Amplitude of the stimulus in the haptic domain (H) was translated as contrast in the visual domain (V). Crucially, to match the H condition, only a moving dot following the participant's finger was revealed on the screen in V. C. Index finger trajectory indicating the scanning pattern of the virtual texture in one trial. On this trial, the participant actively sensed the left subspace first, then moved to the right subspace and explored it before coming back to the left subspace again and reporting their choice. D. Psychometric curves indicating the percentage of non-reference choices for all three sensory conditions (V in blue, H in green, VH in red) and for all stimulus differences. Large dots represent average percentage of choices across participants and smaller dots represent individual participant means. Data are fit using cumulative Gaussian functions. E. Cumulative distributions (CDF) of response times for all three sensory conditions V in blue, H in green, VH in red) across all trials of all participants. Thick lines indicate CDFs across all participant data and thin lines indicate individual participant CDFs for each sensory condition.

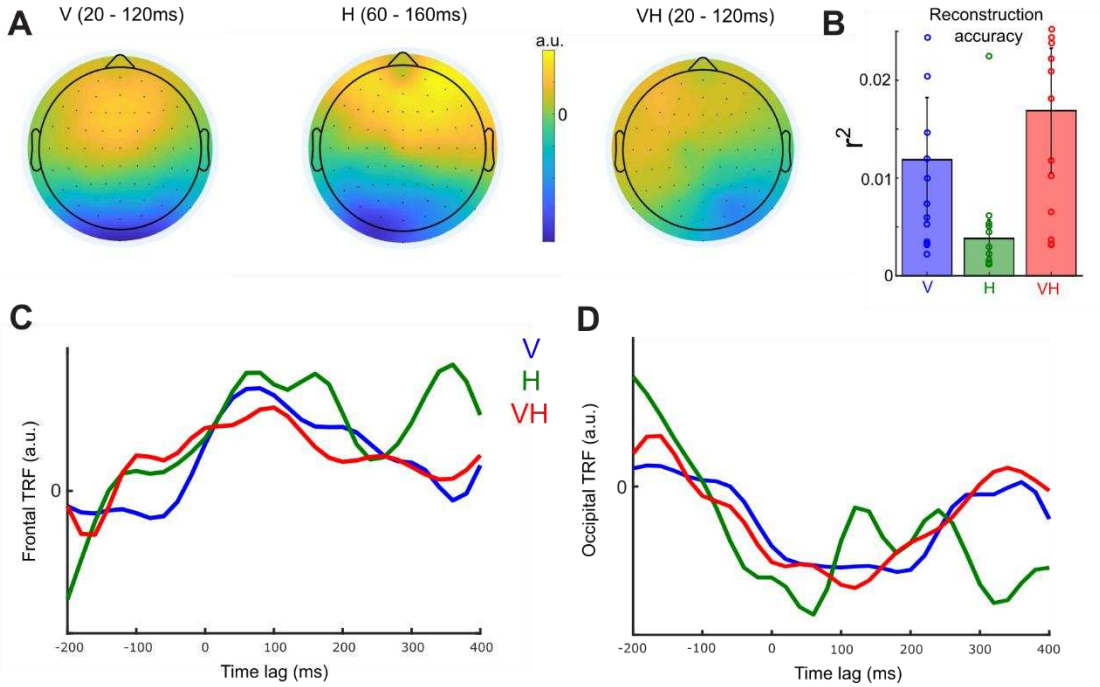

Figure 2. Results of velocity reconstruction analysis using EEG signals. A. Scalp topographies of the forward models representing neural encoding of instantaneous finger velocity for the three sensory conditions. The presented scalp maps show velocity-encoding EEG signals averaged over the following time windows: [20,120]ms lags between velocity and EEG for V and VH and [60,160]ms lags for H. B. Accuracy of the velocity reconstruction from the EEG signals measured using the squared correlation coefficient (r^2) between the original and the approximated velocity profile in the three sensory conditions (V in blue, H in green, VH in red). Bars represent means across participants and errorbars represent standard errors (sem). Dots represent individual participant data. C-D. Temporal response functions (TRFs) of the velocity-encoding EEG activity in the three sensory conditions (V in blue, H in green, VH in red) averaged over frontal electrodes (in C) and over occipital electrodes (in D).

Table1: Estimated values of the three core HDDM parameters for the best-fitting model

Parameter	Mean	Confidence Interval (5%)	Confidence Interval (95%)
Drift rate (δ)	0.897	0.628	1.162
Non-decision time (τ)	2.897	2.710	3.045
Decision boundary (α)	2.853	2.501	3.256

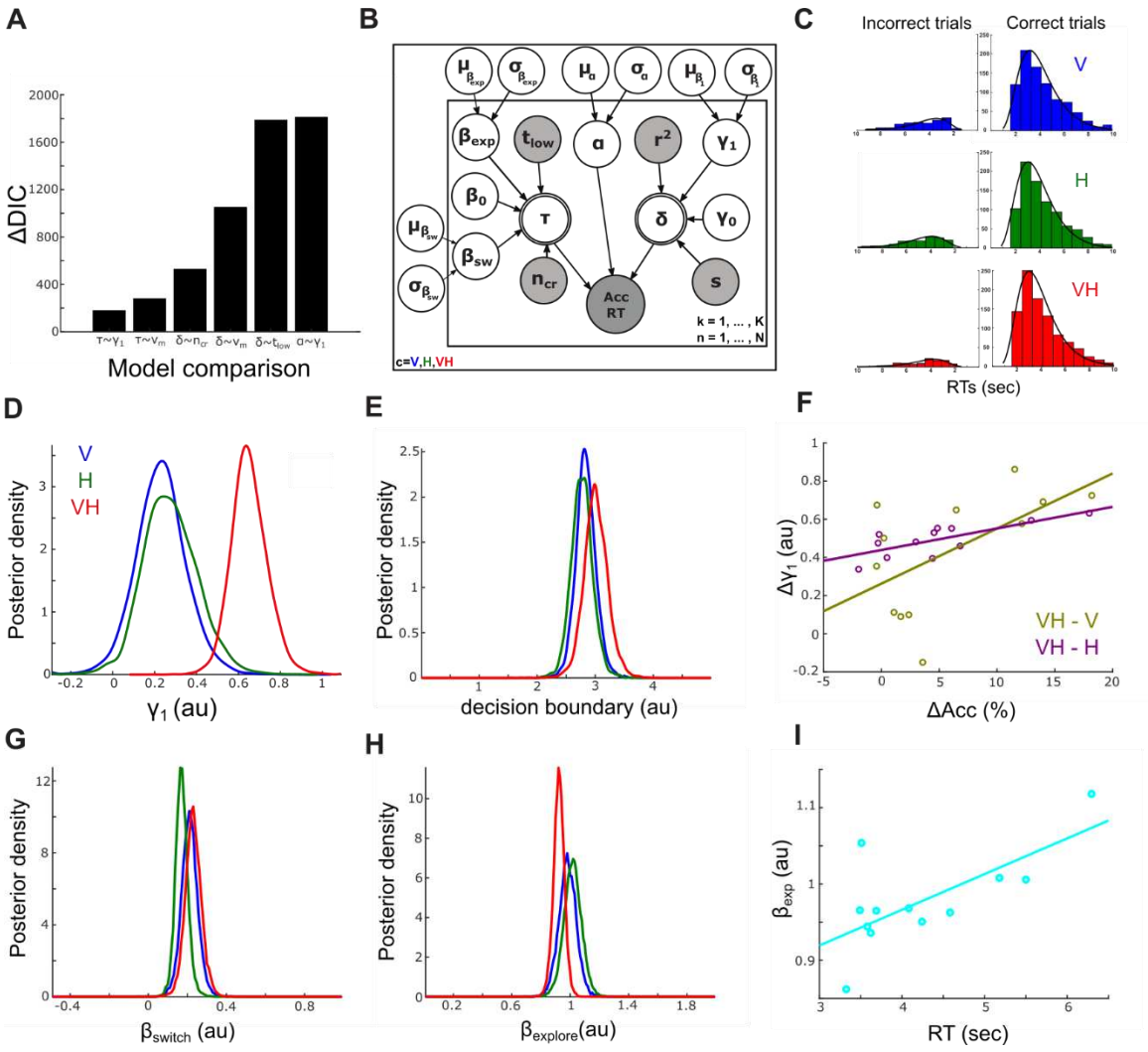

Figure 3. Informed modeling of decision-making behavior. A. Comparison of the best-fitting model (with r^2 as a regressor of drift rate δ only and n_{cr}, t_{low} as regressors of non-decision time τ only) with alternate models using the Deviance Information Criterion (DIC). Positive ΔDIC ($DIC_{model} - DIC_{optimal}$) values for all six models indicate that the model of choice achieved a better trade-off between goodness-of-fit and number of free parameters. B. Graphical representation showing hierarchical estimation of HDDM parameters. Round nodes represent continuous random variables and double-bordered nodes represent variables defined in terms of other variables. Shaded nodes represent recorded or computed signals, i.e. single-trial behavioral data (accuracy, RT and stimulus differences *s*), EEG-velocity couplings (r^2) and kinematic parameters (n_{cr}, t_{low}). Parameters are modelled as Gaussian random variables with inferred means μ and variances σ^2. Plates denote that multiple random variables share the same

parents and children. The outer plate is over sensory conditions (V,H, VH) and the inner plate is over all trials (K) and participants (N). C. Behavioral RT distributions are shown as histograms for each sensory condition (V in blue, H in green, VH in red) for correct (right) and incorrect (left) trials together with the HDDM fits (black lines). Higher histogram values on the right indicate higher proportion of correct choices. D. Posterior distributions of regression coefficients (γ_1) of the EEG-velocity couplings (r^2), as predictors of the drift rate (δ) of the HDDM shown in A. The three coloured curves indicate posterior distributions for the three sensory conditions (blue – V, green – H, red – VH). E. Posterior distributions of decision boundaries for the three sensory conditions (blue – V, green – H, red – VH). F. Cross-participant correlation of differences in choice accuracy (ΔAcc - x-axis) and differences in β_1 ($\Delta\beta_1$ – y-axis) between the multisensory (VH) and the two unisensory (V,H) conditions (VH-V in yellow, VH-H in purple). G. Posterior distributions of regression coefficients (β_{sw}) of the number of crossings between L and R (n_{cr}), as predictor of non-decision time (τ) of the HDDM shown in A. H. Posterior distributions of regression coefficients (β_{exp}) of the time spent on the low-amplitude stimulus (t_{low}), as predictor of non-decision time (τ) of the HDDM shown in A. I. Cross-participant correlation of average response times across trials and sensory conditions (x-axis) and β_{exp} (y-axis).

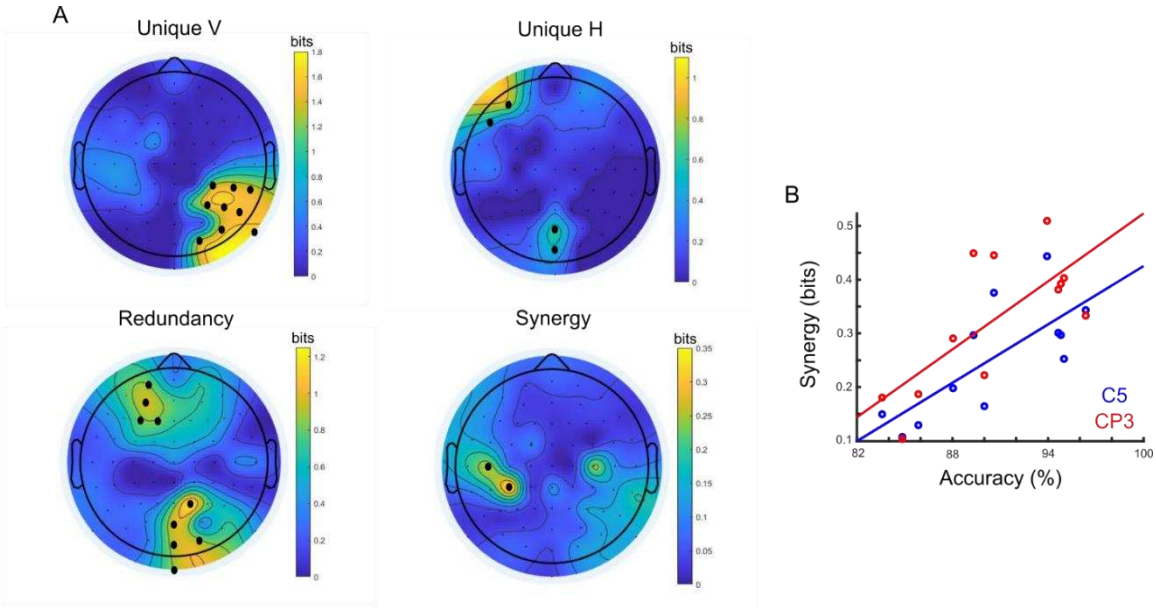

Figure 4. Neural representations and cross-modal interactions. A. Results of PID applied to predict the multisensory (VH) model of active sensing from the two unisensory (V and H) models. Dots on the scalp topographies indicate the EEG channels that provide significant (p<0.01, FDR corrected) visual unique (top left), haptic unique (top right), redundant (bottom left) and synergistic (bottom right) neural information respectively. B. Across-subject correlation between synergy in the two significant EEG channels (CP3 in red and C5 in blue) and choice accuracy in the VH condition.

Author Contributions: Conceptualization: ID, PS, QW
Methodology: ID, RAAI, PS, QW
Investigation: ID
Supervision: PS, QW
Writing—original draft: ID
Writing—review & editing: ID, RAAI, PS, QW