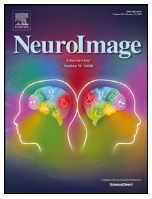
NeuroImage 247 (2022) 118841

Contents lists available at ScienceDirect

NeuroImage

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



Neurocomputational mechanisms underlying cross-modal associations and their influence on perceptual decisions

Joshua Bolam[a,*], Stephanie C. Boyle[b], Robin A.A. Ince[b], Ioannis Delis[a,*]

[a] School of Biomedical Sciences, University of Leeds UK
[b] Institute of Neuroscience and Psychology, University of Glasgow UK

ARTICLE INFO

Keywords:
Cross-modal associations
Congruency
Implicit Association Test
EEG
Perceptual decision-making
Hierarchical Drift Diffusion Model

ABSTRACT

When exposed to complementary features of information across sensory modalities, our brains formulate cross-modal associations between features of stimuli presented separately to multiple modalities. For example, auditory pitch-visual size associations map high-pitch tones with small-size visual objects, and low-pitch tones with large-size visual objects. Preferential, or congruent, cross-modal associations have been shown to affect behavioural performance, i.e. choice accuracy and reaction time (RT) across multisensory decision-making paradigms. However, the neural mechanisms underpinning such influences in perceptual decision formation remain unclear. Here, we sought to identify when perceptual improvements from associative congruency emerge in the brain during decision formation. In particular, we asked whether such improvements represent 'early' sensory processing benefits, or 'late' post-sensory changes in decision dynamics. Using a modified version of the Implicit Association Test (IAT), coupled with electroencephalography (EEG), we measured the neural activity underlying the effect of auditory stimulus-driven pitch-size associations on perceptual decision formation. Behavioural results showed that participants responded significantly faster during trials when auditory pitch was congruent, rather than incongruent, with its associative visual size counterpart. We used multivariate Linear Discriminant Analysis (LDA) to characterise the spatiotemporal dynamics of EEG activity underpinning IAT performance. We found an 'Early' component (~100–110 ms post-stimulus onset) coinciding with the time of maximal discrimination of the auditory stimuli, and a 'Late' component (~330–340 ms post-stimulus onset) underlying IAT performance. To characterise the functional role of these components in decision formation, we incorporated a neurally-informed Hierarchical Drift Diffusion Model (HDDM), revealing that the Late component decreases response caution, requiring less sensory evidence to be accumulated, whereas the Early component increased the duration of sensory-encoding processes for incongruent trials. Overall, our results provide a mechanistic insight into the contribution of 'early' sensory processing, as well as 'late' post-sensory neural representations of associative congruency to perceptual decision formation.

Introduction

In everyday life, we encounter situations where we are required to form rapid perceptual decisions based on ambiguous sensory information (Philiastides et al., 2017; Philiastides and Heekeren, 2009). This can involve processing information presented to multiple sensory modalities (Alais et al., 2010; Ghazanfar and Schroeder, 2006), a process commonly referred to as multisensory decision-making (Bizley et al., 2016; Drugowitsch et al., 2014; Franzen et al., 2020; Rapaso et al., 2012). Previous research has shown decision-making benefits deriving from complementary features of information across multiple sensory modalities (Ernst and Bülthoff, 2004). The brain's tendency to systematically map implicitly learnt associations between features of information across sensory modalities is referred to as cross-modal association (Parise and Spence, 2013; Spence and Deroy, 2013; Spence et al., 2011). When exposed to complementary features of sensory information, features that refer to the same object are redundantly associated, forming cross-modal associations, enabling the brain to exploit the correlation between such informational cues when forming perceptual decisions from ambiguous, and often noisy, unisensory information (Bien et al., 2012; Glicksohn and Cohen, 2013) .

Cross-modal associations have been shown to influence the consolidation of multisensory information when forming perceptual decisions (Bizley et al., 2016; Drugowitsch et al., 2014; Engel et al., 2012). This has been evidenced in studies that have used speeded classification paradigms, demonstrating behavioural effects such as increased response speed (i.e. decreased reaction times; RTs; Kayser and Kayser, 2018; Laurienti et al., 2004; Silva et al., 2017), increased choice

* Corresponding author.
E-mail addresses: bsjwb@leeds.ac.uk, i.delis@leeds.ac.uk (J. Bolam).

https://doi.org/10.1016/j.neuroimage.2021.118841.
Received 22 March 2021; Received in revised form 7 December 2021; Accepted 19 December 2021
Available online 21 December 2021.

accuracy (Franzen et al., 2020; Kayser and Kayser, 2018; Kayser et al., 2017; Kim et al., 2008), and improved stimulus detection (Adam and Noppeney, 2014; Aller et al., 2015). These associative influences towards multisensory decision-making are consistently attributed to the modulatory effects of *cross-modal (in)congruency* (Marks, 2004), i.e. modulations in behavioural performance when a multisensory stimulus has two or more features that are (un)favourably mapped. Preferential, or anticipated, cross-modal associations, are referred to as *congruent*, whereas non-preferential, or non-anticipated, cross-modal associations, are referred to as *incongruent*. A paradigmatic example demonstrated extensively across previous literature concerns auditory pitch-visual size associations (Bien et al., 2012; Evans and Tresiman, 2010; Gallace and Spence, 2006; Parise and Spence, 2009; 2008; 2012). Congruent auditory pitch-visual size associations map high-pitch tones with small-size objects, and low-pitch tones with large-size objects, whereas their incongruent counterparts map high-pitch tones with large-size objects, and low-pitch tones with small-size objects.

Previous research has demonstrated that the congruency of auditory pitch-visual size associations modulates behavioural performance, in particular, benefitting the formation of perceptual decisions. For example, Gallace and Spence (2006) found, using a visual discrimination paradigm, that participants responded more rapidly (i.e. decreased RTs) when auditory stimulus pitch (high/low-pitch tones) was congruent with the visual stimulus size (small/large-size disks) than when incongruent or no auditory stimulus was presented. Similarly, Parise and Spence (2008) found that when participants were asked to judge the temporal order of two different-sized visual stimuli (large/small-size grey circles), congruent auditory tones increased choice accuracy (i.e. higher sensitivity temporal order judgements). In contrast, in a follow-up study, participants judged the spatial discrepancy of an auditory stimulus less accurately (i.e. higher just noticeable difference discrimination thresholds) when presented congruently with the visual stimulus, suggesting a decisional bias of congruency and showing that the behavioural effects of congruency depend on the task at hand (Parise and Spence, 2009). Finally, decreased RTs for congruent, compared to incongruent, pairings were found when only one unisensory stimulus feature was presented per trial using an Implicit Association Test (IAT; Parise and Spence, 2012).

The neural basis of cross-modal associations within perceptual decision formation has recently become a focus of human electrophysiology and neuroimaging research (Spence et al., 2011; Bizley et al., 2016). However, the neural mechanisms facilitating these behavioural enhancements remain less well understood. In particular, it is not clear whether such improvements reflect the consequences of 'early' sensory processing benefits, or 'late' post-sensory changes in decision dynamics, or both. For example, auditory pitch-visual size congruency effects have been identified across two Event-Related Potentials (ERPs) at ~250 ms and ~300 ms at parietal and frontal electrodes respectively (Bien et al., 2012), whereas neural modulations of associative semantic congruency have been found in parahippocampal, dorsomedial, and orbitofrontal cortices at ~100 ms and ~400 ms post-stimulus (Diaconescu et al., 2011). Similarly, significant differences between congruent and incongruent learned label-object associations have been identified as early as ~140 ms across occipital regions, whereas mismatches to the learned associations evoked a modulation between ~340 ms and ~520 ms across parietal regions (Kovic et al., 2010). Overall, the above studies have started to shed light on the neural underpinnings of associative congruency across various association types, yet they have not provided a conclusive mechanistic account of how the brain uses cross-modal associations to improve the efficiency of perceptual decisions.

Difficulties in identifying the neural basis of cross-modal associations further stems from the utilisation of experimental paradigms that present two or more unisensory features. Previous research has associated multiple neural processes with the observed decision-making benefits, in particular i) multisensory integration; integrating information across sensory modalities into unified percepts (Angelaki et al., 2009; Calvert et al., 2004; Mercier and Cappe, 2020), or ii) a form of selective attention; dividing attentional resources towards attending to task-relevant information in one sensory modality, and ignoring task-irrelevant information in another modality (Bien et al., 2012; Choi et al., 2018; Gallace and Spence, 2006; Marks, 2004). Attending to two simultaneously presented stimulus features may facilitate enhancements to perceptual decision formation from benefits not directly attributed to genuine cross-modal associations. As such, any underlying neural activity will display mixed selectivity representing a variety of sensory, decision-related, and other task-relevant signals (Chandrasekaran, 2017; Dahl et al., 2009; Fusi et al., 2016; Kobak et al., 2016; Park et al., 2014; Raposo et al., 2014; Rigotti et al., 2013). Therefore, it remains difficult to characterise whether cross-modal associations represent early sensory processing benefits and/or late post-sensory changes to decision dynamics during the formation of perceptual decisions.

In this study, we sought to capitalise on the novelty of using a modified variant of the Implicit Association Test (IAT), demonstrated by Parise and Spence (2012), to induce auditory pitch-visual size cross-modal associations from the presentation of one unisensory stimulus feature (i.e. auditory pitch). The IAT presents one stimulus feature per trial and manipulates associative congruency by switching the stimulus feature-response key mappings across blocks of trials. Therefore, the proposed experimental manipulations overcome the methodological limitations present in previous research. First, the presentation of one sensory stimulus feature limits confounding effects from the processes of multisensory integration and selective attention. Second, the manipulation of associative congruency across blocks limits confounding effects from explicit stimulus feature mappings and subjective reporting of cross-modal associations. Thus, by coupling this paradigm with electroencephalography (EEG), we can record the neural activity underlying formulated auditory pitch-visual size associations, which is less likely to be affected by confounding activity attributed to processing multisensory stimuli.

Using this paradigm, we aim to mechanistically characterise the neural dynamics underlying cross-modal associations during perceptual decision formation. To achieve this, we first analysed single-trial EEG activity using multivariate Linear Discriminant Analysis (LDA; Parra et al., 2002; 2005; Philiastides and Sajda, 2006; Philiastides et al., 2006; Philiastides et al., 2014; Sajda et al., 2009). Then, to dissect the constituent processes underlying the effects of pitch-driven associations on perceptual decision formation, we adopted a *neurally-informed cognitive modelling approach* (Delis et al., 2018; Diaz et al., 2017; Franzen et al., 2020; Kayser and Shams, 2015; Turner et al., 2013, 2016, 2017). This approach links underlying latent behavioural variables to hypothesised cognitive processes, and further constrains model fits with recorded neuroimaging data, to interpret the modulation of neural activity under different experimental conditions. Previous neurally-informed cognitive modelling research has provided mechanistic characterisations of neural activity underlying perceptual decision formation, and recently, multisensory decision-making (Delis et al., 2018; Franzen et al., 2020; Mercier and Cappe, 2020). Here we used a neurally-informed Hierarchical Drift Diffusion Model (HDDM; Wiecki et al., 2013) to understand how the neural representations of auditory-driven pitch-size associations drive behavioural benefits to perceptual decision formation. Using this approach, we can extract sensory and decision-specific processes from brain activity and relate these to associative congruency benefits when forming perceptual decisions.

Methods and materials

Participants

20 participants (male = 7, female = 13; age range = 19–32 years) were recruited for this study. All participants reported normal/corrected-to-normal vision and normal hearing. Participants

were recruited using the University of Glasgow Subject Pool and received £6/hour (UK Sterling) for their participation. The study was approved by the ethics committee of the College of Science and Engineering at the University of Glasgow (CSE application number 300,130,001), and was conducted in accordance with the Declaration of Helsinki.

Stimuli

We used two auditory and two visual stimuli, which were created and presented using MATLAB (Mathworks) and the Psychophysics Toolbox Extensions (Brainard, 1997). Auditory stimuli consisted of two 300 ms pure tones ('high' and 'low' pitch, 2000 Hz and 100 Hz respectively). Visual stimuli consisted of two light grey circles ('small' and 'large', 2 cm and 5 cm, 1.1° and 2.8° of visual angle respectively). The sound intensity of each tone was matched to 72 dB(A) SPL for left and right ears using a sound level metre. Auditory stimuli were presented using Sennheiser headphones and visual stimuli were presented on a Hansol 2100A CRT monitor at a refresh rate of 85 Hz.

Implicit Association Test

The IAT is a two-alternative forced-choice (2AFC) task that measures implicit perceptual associations between two arbitrary stimulus features by manipulating stimulus feature-response key mappings (Greenwald et al., 1998). In one block of trials, two stimulus features are assigned, or *mapped*, to the same response key, whereas in a separate block of trials, they are assigned to different response keys. Reaction times (RTs) and choice accuracy are collected as dependant variable measurements of behavioural performance (and perceptual choice formation). The IAT assumes that the *congruency* of stimulus feature-response key mappings modulates behavioural performance, with perceptual choices faster (i.e. lower RTs) and more accurate (i.e. higher choice accuracy) when stimulus features are assigned to the same response key than when assigned to different response keys.

This study used a modified version of the IAT, adapted from Parise and Spence (2012), to formulate auditory pitch-visual size cross-modal associations (Fig. 1). In this version, on each block, one auditory (high-pitch/low-pitch tone) and one visual (small-size/large-size circle) stimulus feature are assigned to each response key. Participants are then instructed to categorise as quickly and as accurately as possible which stimulus feature was presented on a single-trial basis using the correctly assigned response key. Congruency was manipulated by switching the stimulus feature-response key mappings across blocks of trials. Congruent mappings assigned high-pitch tones and small-size circles to the left response key, and low-pitch tones and large-size circles to the right response key (Fig. 1, *top*). Incongruent mappings, however, switched the auditory stimulus feature-response key mappings only, so that high-pitch tones and low-pitch tones were assigned to the right and left response keys respectively (Fig. 1, *bottom*). These mappings justify previous findings, which suggested that high-pitch tones are often preferentially associated with small-size visual objects, and vice versa (Gallace and Spence, 2006; Evans and Treisman, 2010; Parise and Spence, 2012). The assigned visual stimulus features remained fixed across blocks for two reasons: (1) Pilot testing found that participants started to exhibit cross-modal associations between visual size and their assigned response keys, rather than their auditory pitch counterparts. Specifically, small-size and large-size visual objects were associated with left and right response keys respectively. (2) In total, experimental sessions ran for ~3 h (~2 h for EEG setup/cleanup, ~1.5 h for the task of 1280 trials per subject). Taken together, this made it difficult to design a cross-modal association experiment where the auditory and visual stimuli were counterbalanced, and participants were not asked to spend more than three hours in a single laboratory session. For these reasons, we chose to only manipulate auditory pitch-response key mappings, therefore manipulating auditory stimulus feature congruency, across blocks. These stimulus feature-response key mapping manipulations are consistent with the mapping manipulations used in the study by Parise and Spence (2012).

Procedure

Participants completed the experiment in a dark and electrically shielded room. Each block began with instructions on the auditory pitch-visual size mapping between stimuli and response keys (see *Implicit Association Test* section). Participants were given as much time as they needed to memorize the instructions for the upcoming block. Each trial started with a fixation cross presented centrally on-screen for a randomized period (uniform distribution from 500 to 1000 ms). Then, one of the four stimuli (see *Stimuli* section) were selected randomly and presented for 300 ms. Participants were instructed to categorize, as quickly and as accurately as possible, the presented stimulus using the left and right keyboard response keys, as defined by the instructions given for that specific block (see *Implicit Association Test* section). Feedback was given after each trial, with green fixation crosses given for correct response choices, and red fixation crosses given for incorrect response choices. Feedback was provided for a randomised duration (uniform distribution from 300 ms to 600 ms). In total, participants completed 8 blocks (4 blocks each for the congruency of stimulus feature-response key mappings presented in a randomized order) for a total of 1280 trials (160 trials per block; 40 trials for each stimulus feature).

Analysis of behavioural data

For each participant, median RTs and choice accuracy (calculated as the proportion of correct choices over all trials) were used as dependant variable measurements of behavioural performance. These were calculated separately for two independent variables: i) stimulus feature (auditory: high-pitch/low-pitch tones; visual: small-size/large-size circles), and ii) congruency of stimulus feature-response key mappings (congruent/incongruent). To further assess the effect of switching auditory stimulus feature-response key mappings, we calculated RTs for correct and incorrect choice responses. Trials with RTs less than 300 ms, or more than 1200 ms, were excluded from further analysis, as behavioural performance on such trials is often attributed to "fast guesses" (<300 ms) or attentional lapses (>1200 ms) during testing (Whelan, 2008). Overall, 905 trials (auditory: 677 trials; visual: 228 trials) were excluded from further analysis, leaving a total of 9850 trials for auditory stimuli and 10,323 trials for visual stimuli

As anticipated, RT data was not normally distributed (Whelan, 2008). Therefore, we statistically analysed median RTs and choice accuracy using Wilcoxon Matched-Pairs Signed-Rank tests, and further analysed RTs for correct and incorrect choices responses using Mann-Whitney U Paired tests. Effect sizes were calculated by dividing the Wilcoxon Signed-Rank test statistic (Z) by the square root of the test population ($N = 20$) for stimulus features (auditory: high-pitch/low-pitch tones; visual: small-size/large-size circles) and congruency (congruent/incongruent) respectively (Rosenthal et al., 1994). For Mann-Whitney U Paired testing analysing RTs of correct and incorrect choices within each congruency and stimulus feature condition for auditory stimuli, effect sizes were calculated by dividing the Mann-Whitney U test statistic (Z) by the square root of the total number of trials for correct and incorrect responses in each congruency condition (congruent/incongruent) and stimulus feature condition (high-pitch/low-pitch tones). For Mann-Whitney U Paired testing analysing RTs of correct and incorrect choices across congruency and auditory stimulus feature conditions, effect sizes were calculated by dividing the Mann-Whitney U test statistic (Z) by the square root of the total number of trials in each accuracy condition (correct/incorrect) across congruency and auditory stimulus feature conditions This enabled us to analyse the effects of both congruent/incongruent stimulus feature-response key mapping conditions, and high-pitch/low-pitch

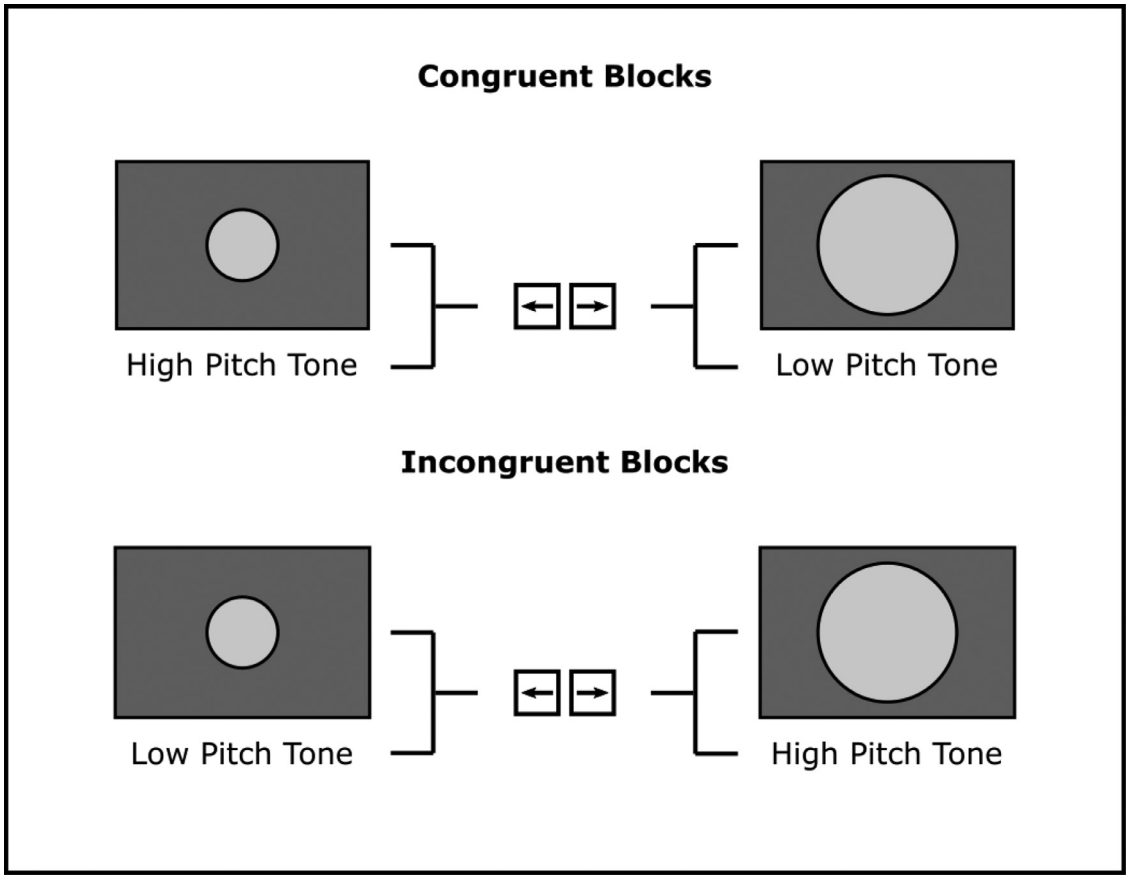


Fig. 1. Implicit association test. Participants were presented with one unisensory stimulus feature (auditory high/low-pitch tone; visual small/large-size circle) per trial, and were asked to categorize which stimulus feature, within that modality, was presented as quickly and accurately as possible, using the correct response key (left/right). Auditory congruency (congruent/incongruent) was manipulated by switching the stimulus feature-response key mappings across blocks (*top*, congruent block mappings; *bottom*, incongruent block mappings).

tone presentations for correct and incorrect RTs respectively. *Post hoc* power analyses were conducted using G*Power (Faul et al., 2007; 2009) to assess whether any identified significant results were of sufficient statistical power (Cohen 1992a; 1992b; see the *Power analyses* section in *Supplementary materials*). Statistical analysis of all behavioural data was completed using R.

EEG recording and preprocessing

Continuous EEG data was recorded in a sound-attenuated and electrostatically shielded room using a 128-channel BioSemi amplifier system and ActiView recording software (Biosemi, Amsterdam, Netherlands). Signals were sampled and digitized at 512 Hz, then band-pass filtered online between 0.16 and 100 Hz. Signals originating from ocular muscles were recorded from four additional electrooculography (EOG) electrodes placed below and at the outer canthi of each eye.

Individual blocks of data were preprocessed using the Fieldtrip Toolbox (Oostenveld et al., 2011), which was implemented in MATLAB using custom scripts. Epochs of 2 s, from −0.5 to 1.5 s relative to stimulus onset, were extracted and filtered between 0.5 and 90 Hz using a Butterworth filter, before being down-sampled to 200 Hz. Potential signal artefacts were removed using Independent Component Analysis (ICA), using the Fieldtrip toolbox (Oostenveld et al., 2011). Components related to typical eye movement activities, such as blinks, or noisy electrode channels were removed. Horizontal, vertical, and radial EOG signals were further processed using established procedures (Hipp and Siegel, 2013; Keren et al., 2010) and trials with high correlations between eye movements (e.g. saccades) and components in the EEG data removed. Remaining trials with amplitudes that exceeded ±120 μV were also removed. Successful cleaning was verified by visual inspection of single trials.

EEG signal analysis - Linear Discriminant Analysis

We applied single-trial multivariate Linear Discriminant Analysis (LDA; Parra et al., 2002;2005; Philiastides and Sajda, 2006; Philiastides et al., 2006; Philiastides et al., 2014; Sajda et al., 2009) to extract EEG components discriminating between congruent and incongruent trials for auditory stimulus-locked EEG data only. Specifically, for a pre-defined time window of interest, this method applies a linear multivariate classifier to EEG data in order to estimate a spatial weighting vector that quantifies the optimal combination of EEG sensor linear weights. When applied to multichannel EEG data, this yields a one-dimensional projection that maximally discriminates between two conditions of interest. This projection represents the 'discriminating component' that integrates all signal information across the multichannel EEG array, while reducing effects common to both conditions. Compared to univariate trial-averaging approaches, notably Event-Related/Evoked Response Potential (ERP) analyses, multivariate approaches are better able to spatially integrate information across the multidimensional EEG sensor space, yielding components which both preserve inter-trial signal variability and increase the signal-to-noise ratio (Sajda et al., 2011) for preserved task-relevant information. Note that the term 'component' is preferred instead of 'source' in order to make clear that this is a projection of all EEG activity correlated with the underlying source.

We used a sliding window approach (Parra et al., 2005; Sajda et al., 2009) to identify a projection of the multichannel EEG signal, $x_i(t)$, where $i = [1 \ldots N\ trials]$, and N is the total number of trials, within short time windows that maximally discriminated between congruent and incongruent trials for auditory stimulus features only. All time windows had a width of 50 ms, with the window centre t shifted from −100 ms to 800 ms, relative to auditory stimulus-onset, in 5 ms increments. Specifically, we used logistic regression (Parra et al., 2002; 2005) to learn a 128-channel spatial weighting vector $w(t)$ that achieved maximal discrimination within each time window. This yields a one dimensional projection, $y_i(t)$, for each trial i and given window t:

$$y(t) = w^T x(t) = \sum_{i=1}^{D} w_i x_i(t)$$

Here, D represents the number of channels in the multichannel EEG array and T refers to a matrix transpose operator. Our classifier was designed to map component amplitudes, $y_i(t)$, for congruent and incongruent trials, that separates activity maximizing differences and minimizing similarities of effects from neural processes common to both conditions. In discriminating the two congruency categories, the classifier maps negative and positive discriminant component amplitudes to congruent and incongruent trials respectively. Thus, larger negative values indicate a higher likelihood of categorizing auditory stimuli within congruent stimulus feature-response key mappings, and larger positive values indicate a higher likelihood of categorizing auditory stimuli within incongruent stimulus feature-response key mappings, with values near zero reflecting less discriminative component amplitudes.

We quantified classification performance of our classifier for each time window using the area under a receiver operating characteristic (ROC) curve (Green and Swets, 1966), referred to as an A_z value, using a leave-one-out cross-validation procedure (Gherman and Philiastides, 2015; Philiastides and Sajda, 2006). To determine group significance thresholds for discriminator performance, we implemented a permutation test, whereby congruent and incongruent trial labels were randomized and submitted to the leave-one-out procedure. This randomization procedure was repeated 1000 times, producing a probability distribution for A_z, which we used as reference to estimate the A_z value leading to a significance level of $p < 0.05$.

Finally, the linearity of our model allowed us to compute scalp projections of the discriminating components resulting from Eq. (1) by estimating a forward model as:

$$a(t) = \frac{x(t)y(t)}{y(t)^T y(t)}$$

where the EEG data (x) and discriminating components (y) are organized as matrix and vector notations, respectively, for convenience. Here, the EEG matrix, $x_i(t)$, denotes channel activity across rows and trials across columns for all 5 ms increments in time window t, whereas discriminating components, $y_i(t)$, are organized as single-trial vectors, $y(t)$, with each row is from trial i. Such forward model implementations can be displayed as scalp topographies and interpreted as the coupling between discriminating component amplitudes and observed multichannel EEG activity, whereby vector $a(t)$ reflects the coupling of the discriminating component $y(t)$ that explains most of the activity in $x(t)$, with maps illustrating this optimal component-activity coupling (Philiastides et al., 2014).

Hierarchical Drift Diffusion Model – description

We fit participants' behavioural performance i.e. RTs and choice accuracy, with a Hierarchical Drift Diffusion Model (HDDM; Wiecki et al., 2013). Similar to the traditional Drift Diffusion Model (DDM; Ratcliff et al., 2015; R. 2016; Forstmann et al., 2016; Ratcliff and McKoon, 2008; Ratcliff, 1978), the HDDM assumes sensory evidence is stochastically accumulated over time, towards one of two decision boundaries, corresponding to two choice alternatives (e.g. correct or incorrect choices; left or right response keys). For each decisional process, the HDDM returns parameter estimates of four internal components of perceptual decision-making, (1) the rate of evidence accumulation (drift rate), (2) possible *a priori* bias towards one of the two choice alternatives (starting point), (3) the distance between two decision boundaries controlling the amount of evidence required for one particular choice alternative (decision boundary), and (4) the duration of non-decisional processes, which can include time taken for stimulus encoding and motor-response production latency (non-decision time).

Hierarchical Drift Diffusion Model – fitting

To fit HDDM to participants' performance and estimate internal decisional processes, we used the HDDM toolbox (Wiecki et al., 2013), an open-source software package, written in Python, that permits custom fits of HDDM variants to participants' RTs and choice accuracy. The HDDM uses a Bayesian hierarchical framework to estimate the above four parameters, whereby sampled prior probability distributions of the model parameters are updated based on a likelihood function, formed from the data given to the model, to yield posterior probability distributions. The HDDM uses Markov-Chain Monte Carlo sampling within this framework, whereby prior distributions of estimated parameters are iteratively adjusted by a likelihood function that maximizes the log likelihood of predicted mean RTs and choice accuracy (Gamerman and Lopes, 2006). The use of Bayesian hierarchical frameworks, and specifically the HDDM, allows for several benefits relative to traditional (non-hierarchical) DDM analysis. First, such frameworks assume that participants' samples in a dataset are randomly drawn from a group (Vadekerchkove et al., 2011), thereby constraining participant- and group-level posterior distributions, which yield more stable parameter estimates for individual participants (Wiecki et al., 2013). Second, the HDDM has been found to be more robust in achieving stable parameter estimates in datasets with low numbers of trials, compared to non-hierarchical DDM approaches (Ratcliff and Childers, 2015). Third, rather than quantifying the most likely value for each parameter, uncertainty can be directly conveyed with posterior distributions for each estimated parameter (Wiecki et al., 2013; Navarro and Fuss, 2009; Gelman, 2003). Fourth, and most importantly for our analysis, the HDDM framework supports the use of external variables as regressors of estimated model parameters, to assess the relations between specific parameters with further behavioural or neuroimaging data (Delis et al., 2018; Frank et al., 2015; Franzen et al., 2020; Mercier and Cappe, 2020; Tremel and Wheeler, 2015).

To implement the HDDM, we used a process referred to as 'accuracy-coding' (Wiecki et al., 2013), which fits the HDDM to RT distributions that assume the upper and lower decision boundaries corresponding to correct and incorrect choices respectively. We sampled parameter estimates for drift rate (δ), decision boundary (θ), and non-decision time (τ). Starting point (z) was set as the midpoint between the two decision boundaries, since the IAT had no *a priori* bias towards either choice alternative (i.e. response key; Philiastides et al., 2011). We did not include any inter-trial variability parameters in our models as previous studies have shown that it is difficult to achieve stable posterior estimates, particularly with fewer trials (Boehm et al., 2018; Ratcliff and Childers, 2015). For each model, we ran 5 separate Markov chains with 11,000 samples each. For each chain, the first 1000 were discarded as "burn-in", and the rest subsampled ("thinned") by a factor of two, to reduce the autocorrelation within and between Markov chains. This is a conventional approach to MCMC sampling, whereby initial samples in the "burn-in" period are based on the selection of a random starting point, and neighbouring samples likely to be highly correlated. Both issues are likely to provide unreliable posterior distributions for estimated parameters. This left 25,000 remaining samples for our model, which constituted the probability distributions for each estimated parameter, allowing us to compute individual parameter estimates for participants

and condition categories. To ensure Markov chain convergence, we computed Gelman-Rubin $\hat{R}$ statistics between chains (Gelman and Rubin, 1992). This compares within-chain and between-chain variance of estimated parameters both for individual participants and group conditions. We verified that all $\hat{R}$ statistics fell between 0.98 and 1.02, which suggests reliable convergence between chains.

Hierarchical Drift Diffusion Model – EEG regressors

We sought to use our EEG discrimination analysis results to inform the fitting of the HDDM to our behavioural data (i.e. RTs and choices). Specifically, we used the HDDM toolbox (Wiecki et al., 2013) to construct regressors that assessed the trial-by-trial linear relationship between our single-trial EEG discriminator amplitudes (for congruent and incongruent trials) and posterior estimates for drift rate (δ), decision boundary (θ), and non-decision time (τ). In line with our behavioural results, in which we reported a significant effect of RTs decreasing for congruent trials, we hypothesized that component amplitudes would be predictive of increases in the rate of evidence accumulation (drift rate) and decreases in evidence required for categorising auditory stimuli (decision boundary). For the duration of non-decisional processes (non-decision time), we hypothesized that either a) component amplitudes for congruent trials would be predictive of decreases in the duration of non-decisional processes, or b) component amplitudes for incongruent trials would be predictive of increases in the duration of non-decisional processes. Therefore, as part of the model fitting within the HDDM framework, we used our single-trial EEG discriminator amplitudes for congruent and incongruent trials to construct regressors for drift rate (δ), decision boundary (θ), and non-decision time (τ) as follows:

$$\delta = \alpha_0 + \alpha_1 * \left|y^{max}_{early}\right| + \alpha_2 * \left|y^{max}_{late}\right|$$

$$\theta = \beta_0 + \beta_1 * \left|y^{max}_{early}\right| + \beta_2 * \left|y^{max}_{late}\right|$$

$$\tau = \gamma_0 + \gamma_1 * \left|y^{max}_{early}\right| + \gamma_2 * \left|y^{max}_{late}\right|$$

where $|y^{max}_{early}|$ and $|y^{max}_{late}|$ are the maximum, single-trial, discriminator amplitudes of subject-specific, stimulus-locked EEG components capturing the highest classification performance between congruent and incongruent trials (corresponding to group peak A_z values; Early ~ 110 ms; Late ~ 340 ms; see *Fig. 3*). Coefficients α_1, β_1, γ_1 and α_2, β_2, γ_2 weight the slope of each parameter by the absolute values of $|y^{max}_{early}|$ and $|y^{max}_{late}|$ respectively, with intercepts α_0, β_0, γ_0, on a trial-by-trial basis for each subject and congruency condition. Note that we used the absolute values of our single-trial EEG discriminator amplitudes to construct regressors, since congruent trials were predominantly categorised by negative $|y^{max}_{early}|$ and $|y^{max}_{late}|$ values, and incongruent trials were predominantly categorised by positive $|y^{max}_{early}|$ and $|y^{max}_{late}|$ values respectively (see *EEG signal analysis – Linear Discriminant Analysis* section). Hence, by using these regression coefficients, we were able to assess the trial-by-trial modulatory effects of each identified component on drift rate, decision boundary, and non-decision time in both congruency conditions. Consequently, we can characterise the behavioural benefits of cross-modal associative congruency on perceptual decision formation, dissecting which decisional processes best predict decreases in choice RT.

To assess the posterior predictive power of our regression coefficients, we first calculated the posterior probability densities of samples that differed from 0 using the built-in functions of the HDDM toolbox (Wiecki et al., 2013) corresponding to our pre-defined hypotheses predicting the effect of decreased RTs for congruent trials, and decreased RTs for incorrect responses for congruent trials (albeit not significantly affecting choice accuracy, see *Behavioural results* section). For drift rate and incongruent non-decision time regression coefficients, probability densities were calculated from the proportion of samples greater than 0 ($P(\delta > 0)$; $P(\tau > 0)$), whereas for decision boundary and congruent non-decision time regression coefficients, probability densities were calculated from the proportion of samples less than 0 ($P(\theta < 0)$; $P(\tau < 0)$). Then, we calculated each coefficient's posterior log-odds by applying the logit function to the proportion of posterior samples in favour of their corresponding hypothesis (Ince et al., 2020). This Bayesian Inference approach was utilised because Bayesian hierarchical modelling frameworks violate the assumption of independence in its posterior estimation sampling procedure, since group-level and participant-level parameter posteriors are simultaneously estimated (Wiecki et al., 2013). Therefore, null-hypothesis significance testing approaches commonly utilised in frequentist approaches to statistical analysis are not recommended. To determine the prevalence of true positive results, implicating strong predictive effects of our regression coefficients on posterior parameter estimations, we further calculated the log posterior odds proportion of a hypothetical sample corresponding to a false-positive rate of $\alpha = 0.05$ (i.e. a 95% true-positive threshold). Regression coefficient log-odds proportions greater than the hypothetical log-odds proportion of our false positive rate (which is equal to 2.944) suggests highly predictive effects of our regression coefficients on changes to estimated posterior parameters favoured by our hypotheses.

Results

Behavioural results

Participants responded faster in auditory trials with congruent compared to incongruent stimulus feature-response key mappings (Fig. 2b, Congruent: median = 608 ms post-stimulus offset; Incongruent: 643 ms post-stimulus offset). Wilcoxon Signed-Rank Testing determined this finding to be statistically significant ($Z = -2.135$, $p = 0.033$, effect size = −0.477, Wilcoxon Signed-Rank Testing). This result held for both correct (Fig. 2f, Congruent/Correct: median = 611 ms post-stimulus offset; Incongruent/Correct: median = 645 ms post-stimulus offset, $Z = -6.940$, $p < 0.001$, effect size = −0.073, Mann-Whitney U testing) and incorrect trials separately (Fig. 2f, Congruent/Incorrect: median = 578 ms post-stimulus offset; Incongruent/Incorrect: median = 621 ms post-stimulus offset, $Z = -2.628$, $p = 0.004$, effect size = −0.091, Mann-Whitney U testing). Furthermore, RTs were significantly longer for correct compared to incorrect responses for congruent stimulus feature-response key mappings (Fig. 2f, Correct: median = 611 ms post-stimulus offset; Incorrect: median = 578 ms post-stimulus offset, $Z = -2.142$, $p = 0.016$, effect size = −0.030, Mann-Whitney U Testing), but not for incongruent stimulus feature-response key mappings (Fig. 2f, Correct: median = 645 ms post-stimulus offset; Incorrect: median = 621 ms post-stimulus offset, $Z = -0.664$, $p = 0.253$, effect size = −0.010, Mann-Whitney U Testing). We found no significant effect of stimulus feature on median RTs (Fig. 2a, High-Pitch Tone: median = 625 ms; Low-Pitch Tone: median = 624 ms, $Z = -0.788$, $p = 0.430$, effect size = −0.176, Wilcoxon Signed-Rank Testing). There was also no significant effect when testing correct (Fig. 2e, High-Pitch Tone/Correct: median = 626 ms post-stimulus offset; Low-Pitch Tone/Correct = 626 post-stimulus offset, $Z = 0.421$, $p = 0.663$, effect size = 0.004, Mann-Whitney U Testing) or incorrect trials separately (Fig. 2e, High-Pitch Tone/Incorrect: median = 594 ms post-stimulus offset; Low-Pitch Tone/Incorrect: median = 597 ms post-stimulus offset, $Z = 0.420$, $p = 0.663$, effect size = 0.015, Mann-Whitney U Testing). Furthermore, we found no significant difference in RT between correct and incorrect responses for either high-pitch tones (Fig. 2e, Correct: median = 626 ms post-stimulus offset; Incorrect: median = 594 ms post-stimulus offset, $Z = -0.994$, $p = 0.172$, effect size = −0.013, Mann Whitney-U Testing), or low-pitch tones (Fig. 2e, Correct: median = 627 ms post-stimulus offset; Incorrect: median = 597 ms post-stimulus offset, $Z = -1.627$, $p = 0.052$, effect size = −0.023, Mann Whitney-U Testing).

Regarding choice accuracy, participants had a slightly but not significantly higher proportion of correct responses for auditory trials with

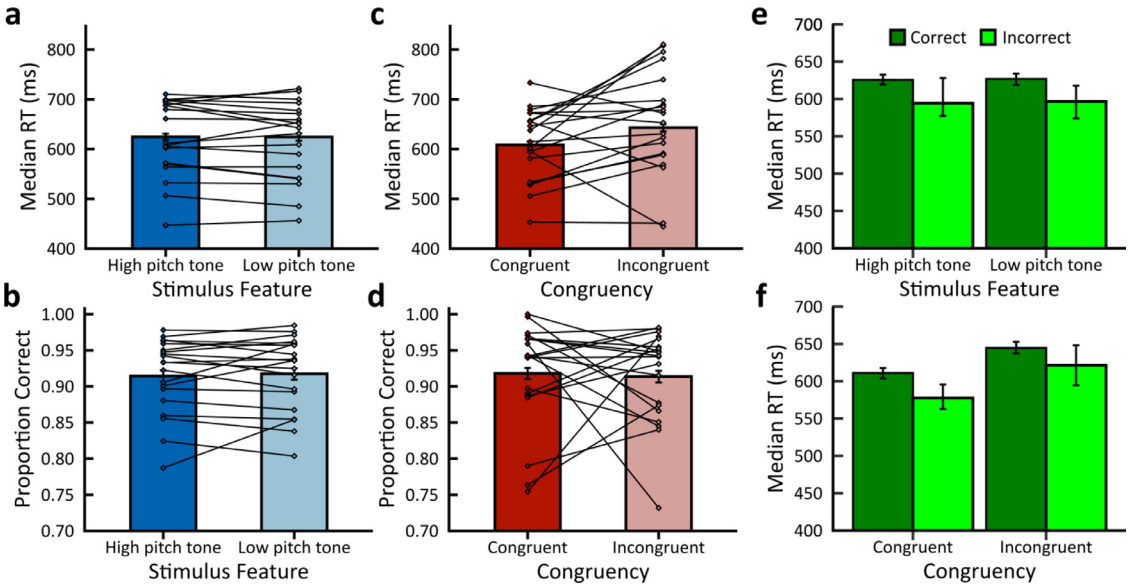


Fig. 2. Behavioural performance. *Left.* Median RTs and choice accuracy (proportion of correct responses) for condition (bars) and participants (scatter points) for **a, b** Stimulus Feature (high/low-pitch tones) and **c, d** Congruency (congruent/incongruent). *Right.* Median RTs for correct and incorrect RT for **e** Stimulus Feature (high/low-pitch tones) and **f** Congruency (congruent/incongruent). For all graphs, 95% Confidence Intervals (CIs) were computed using 1000 bootstrapping random sampling iterations to estimate the distribution of average performance measurements.

congruent compared to incongruent stimulus feature-response key mappings (Fig. 2d, Congruent: proportion correct = 0.918; Incongruent: proportion incorrect = 0.913, $Z = -0.128$, $p = 0.898$, effect size = −0.029, Wilcoxon Signed-Rank Testing). There was also no significant effect of stimulus feature on choice accuracy (Fig. 2b, High-Pitch Tone: proportion correct = 0.914; Low-Pitch Tone: proportion correct = 0.917, $Z = -0.237$, $p = 0.812$, effect size = −0.053, Wilcoxon Signed-Rank Testing).

We found no significant effect of associative congruency on median RTs for visual stimuli (Supplementary Figure 1b, Congruent: median = 581 ms post-stimulus offset; Incongruent: median = 604 ms post-stimulus offset; $Z = -1.161$, $p = 0.245$, effect size = −0.260, Wilcoxon Signed-Rank Testing). We further found no significant effect of visual stimulus feature on median RTs (Supplementary Figure 1a, Small-Size Circle: median = 597 ms post-stimulus offset; Large-Size Circle: median = 586 ms post-stimulus offset, $Z = -0.863$, $p = 0.388$, effect size = −0.193, Wilcoxon Signed-Rank Testing).

Regarding choice accuracy, participants had a slightly, but not significantly, higher proportion of correct responses for trials with congruent compared to incongruent visual stimulus feature-response key mappings (Supplementary Figure 1d, Congruent: proportion correct = 0.957; Incongruent: proportion correct = 0.955, $Z = -0.055$, p-value = 0.956, effect size = −0.012. Wilcoxon Signed-Rank Testing). There was also no significant effect of visual stimulus feature on choice accuracy (Supplementary Fig. 1c, Small-Size Circle: proportion correct = 0.954; Large-Size Circle: proportion correct = 0.958; $Z = -1.161$, p-value = 0.245, effect-size = −0.260, Wilcoxon Signed-Rank Testing).

To summarize, we found responses for congruent auditory trials were faster than responses for incongruent auditory trials and, in addition, within the set of congruent trials, correct responses were slower than incorrect responses. Furthermore, we found responses for congruent visual trials were not faster nor more accurate compared to incongruent visual trials. Therefore, no significant behavioural improvements as a result of associative congruency were demonstrated when categorising visual stimulus features (see *Supplementary materials* for Figs. 1 and 2 for the behavioural and modelling results for visual stimuli respectively).

EEG signal analysis results

Next, we analysed the EEG data to identify the neural components that discriminated between congruent and incongruent trials. Specifically, for each participant separately, we performed a single-trial multivariate discriminant analysis to identify linear spatial weightings (i.e. spatial filters) of the EEG sensors that discriminated congruent from incongruent trials. The identified weightings produced a projection in the 128-dimensional EEG space that maximally discriminated congruent-vs-incongruent trials within short pre-defined windows of 50 ms, locked to stimulus onset.

Application of the resulting linear spatial filters to single-trial EEG data produces a measurement quantifying the discriminating component amplitude (y, see *Methods and materials*). These component amplitudes can be used as an index of the quality of categorizing the congruency of stimulus feature-response key mappings in each trial. In other words, higher amplitudes, negative or positive, indicate higher neural evidence for congruent or incongruent stimulus feature-response key mappings, while values closer to zero indicate less evidence of categorizing associative congruency.

To quantify the discriminator's performance over time, we used the area under a receiver operating characteristic curve (i.e. A_z value), coupled with a leave-one-trial-out cross validation approach, to control for overfitting. Compared to traditional approaches, which assume an A_z value of 0.5 as chance performance, we performed a permutation analysis using a leave-one-trial-out procedure that produced an A_z randomization distribution, to compute a group-average A_z value, that lead to a conventional significance level of $p = 0.05$.

Our discriminator's performance as a function of stimulus-locked time revealed increased discriminant performance from 0 to 600 ms, above the significance level estimated from our permutation test. Specifically, discriminator performance within this range was characterized by two temporally specific components (Fig. 3a; C_{Early}: mean peak time = 100–110 ms, Az value = 0.846; C_{Late}: mean peak time = 330–340 ms, Az value = 0.797). These components were consistent across participants (see Fig. 4a for the A_z curves and Fig. 4b for the maximum A_z values of each participant). We then computed the corresponding

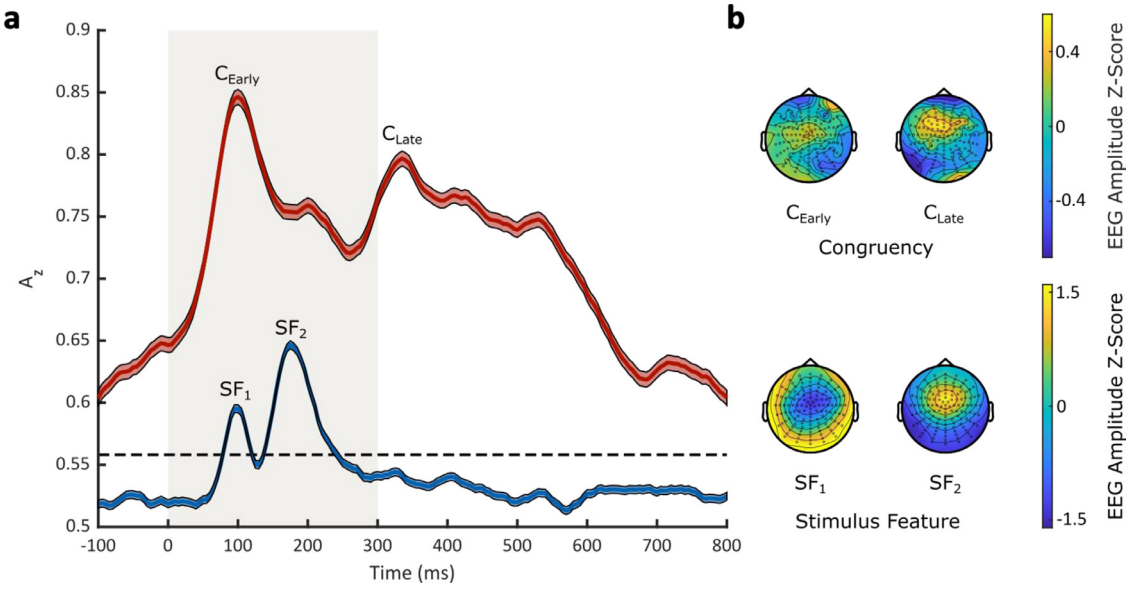


Fig. 3. Multivariate linear discriminant analysis results. a Mean multivariate discriminator performance (A_z), quantified by leave-one-out trial cross-validation procedure, during outcome discrimination of stimulus-locked EEG responses, as a function of congruency (congruent-vs-incongruent; red) and stimulus feature (high-pitch tones-vs-low-pitch tones; blue) conditions. Dashed black line represents the group average permutation threshold at $p < 0.05$ for congruent-vs-incongruent discriminator performance. Shaded error bars denote the standard error of the mean across participants. Shaded area denotes the presentation of auditory stimuli, from 0 ms (post-stimulus onset) to 300 ms (post-stimulus offset). **b** Scalp topographies at representative time windows corresponding to the two EEG components, defined for congruency (*Top*, C_{Early} and C_{Late}) and stimulus feature (*Bottom*, SF_1 and SF_2) conditions respectively.

scalp topographies, obtained using the forward model, correlating between peak discriminant output and EEG data (averaged over a 50 ms time window centred on the two classification performance peaks). For the 'Early' component, the strongest effects originated over central, left-lateralized centro-parietal, and left-lateralized occipital electrodes, whereas for the 'Late' component, the strongest effects predominantly originated over fronto-central electrodes. These results indicate that our multivariate LDA classifier identifies two EEG components that carry significant information about the congruency of stimulus feature-response key mappings.

Similarly, we applied the same single-trial multivariate discriminant analysis to the EEG data to identify the neural components that discriminated between trials that presented high-pitch and low-pitch auditory tones. Here, our discriminator's performance as a function of stimulus-locked time revealed increased discriminant performance post-stimulus onset, characterized by two temporally specific peaks (Fig. 3; SF_1: mean peak time = 90–100 ms, Az value = 0.595; SF_2: mean peak time = 170–180 ms, Az value = 0.647). The corresponding scalp topographies, again obtained using the forward model, revealed a bipolar EEG response that discriminated the two auditory stimuli. The first component (SF_1) had positive activations over outer occipital, parietal, and temporal electrodes and negative activations over a frontocentral cluster, whereas the second component (SF_2) showed activations at the same locations with inverse polarity. Notably, the stimulus-discriminating components occur approximately at the same temporal window as the Early congruency-discriminating EEG component. Thus, taken together, our EEG results attribute an early sensory-encoding role for the Early congruency-discriminating component and a post-sensory role for the Late congruency-discriminating component.

Neurally-informed cognitive modelling results

After characterizing the effect of congruency on the discriminating power of brain activity, we sought to gain a mechanistic insight into how the identified single-trial neural responses were linked to improvements in perceptual decision formation between congruent and incongruent trials. To achieve this, we used a neurally-informed variant of the Hierarchical Drift Diffusion Model (HDDM; Wiecki et al., 2013, see Fig. 5a for a graphical illustration and *Methods and materials* for details on the model). As previously mentioned, the HDDM is a Bayesian implementation of the well-known Drift Diffusion Model, used for characterizing perceptual decision formation in 2AFC paradigms (DDM; Ratcliff and McKoon, 2008).

We extracted the maximum single-trial discriminator amplitudes ($|y^{max}_{early}|$ and $|y^{max}_{late}|$) from subject-specific temporal windows corresponding to our stimulus-locked 'Early' and 'Late' peak EEG components. These values represent the neural evidence for discriminating the congruency of stimulus feature-response key mappings per trial (see Fig. 4c for histograms of y^{max}_{early} and y^{max}_{late} in congruent and incongruent trials). Depending on the stimulus feature-response key mapping, these values demonstrate where stimulus-induced neural responses systematically differ, explicitly linking perceptual decision formation benefits to time periods where early bottom-up and late top-down influences from associative congruency modulate the subsequent neural responses. Thus, we used them to construct regressors for drift rate, boundary separation, and non-decision time parameters in the model. We estimated regression coefficients to assess the relationship between trial-to-trial variations in EEG component amplitude and parameter posterior estimations (Coefficients α_1, β_1, γ_1 and α_2, β_2, γ_2 for $|y^{max}_{early}|$ and $|y^{max}_{late}|$ respectively). Note that we extracted the *absolute* single-trial discriminator amplitudes, as this would permit us to compare indexes of neural evidence, underlying our assumption that larger component amplitudes reflect higher discriminant activity within the brain for congruent compared to incongruent trials (see Fig. 4d for the average $|y^{max}_{early}|$ and $|y^{max}_{late}|$ of each participant).

We found a good fit of the behavioural data (i.e. choice accuracy and RTs) from our proposed neurally-informed HDDM (Fig. 5b). Crucially, we found that the single-trial amplitudes for the Early component were highly predictive of increases in non-decision time estimates for incongruent trials (Early: $P(\gamma_1^{Congruent} < 0) = 0.189$, log-odds $= -1.470$; $P(\gamma_1^{Incongruent} > 0) = 0.997$, log-odds $= 5.861$.

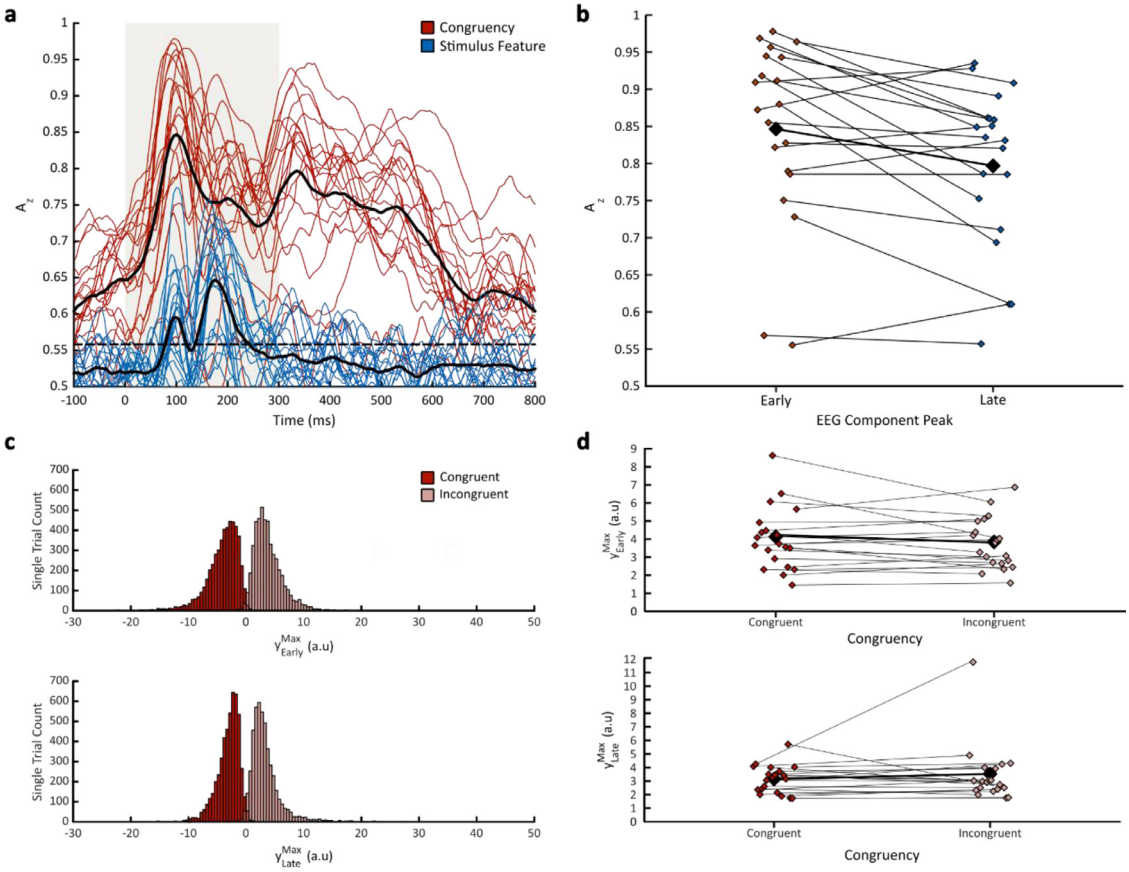


Fig. 4. Participant multivariate linear discriminant analysis results. a Participants' mean discriminator performance (A_z), obtained from a leave-one-out trial cross-validation procedure, during stimulus feature (blue) and congruency (red) discrimination of stimulus-locked EEG responses. Dashed black line represents the group average permutation threshold at $p < 0.05$ for congruent-vs-incongruent discriminator performance. Condition mean discriminator performance (black) is also illustrated for congruency and stimulus feature discrimination**.** Shaded area denotes the presentation of auditory stimuli, from 0 ms (post-stimulus onset) to 300 ms (post-stimulus offset). **b** Participants' mean discriminator performance (A_z) for the Early and Late congruency-discriminating EEG components. **c** Single-trial discriminator amplitudes (y) for the Congruent (red) and Incongruent (pink) component amplitudes are illustrated as histograms for the Early (top) and Late (bottom) EEG components respectively. Negative values indicate neural evidence for congruency whereas positive values indicate neural evidence for incongruency. **d** Absolute values of our single-trial discriminator amplitudes (y) for the Congruent (red) and Incongruent (pink) component amplitudes for the Early (top) and Late (bottom) components respectively.

Late: $P(\gamma_2^{Congruent} < 0) = 0.62$, log-odds = −2.712; $P(\gamma_2^{Incongruent} > 0) = 0.936$; log odds = 2.690; Fig. 6c). We should note that the non-decision time parameter captures the duration of non-decisional processes, such as the latency of early stimulus encoding and the motor preparatory response. This result is consistent with the longer RTs observed in incongruent trials, and combined with the early occurrence of this component (~100 ms post-stimulus onset), suggests a longer duration of early sensory processing during incongruent trials.

We further found evidence to indicate that single-trial amplitudes of the Late component were highly predictive of decreases in decision boundary parameter estimates for congruent trials only (Early: $P(\beta_1^{Congruent} < 0) = 0.370$, log-odds = −0.553; $P(\beta_1^{Incongruent} < 0) = 0.641$, log-odds = 0.580. Late: $P(\beta_2^{Congruent} < 0) = 0.973$, log-odds=3.574; $P(\beta_2^{Incongruent} < 0) = 0.234$, log-odds = −1186; Fig. 6b). Thus, this implies a modulation of the decision boundary in congruent trials by the Late component amplitudes. The lower decision boundary indicates that participants require less evidence to reach a decision in congruent trials, thus they a) respond faster and b) are more likely to make incorrect perceptual judgments when responding fast. These are consistent with our behavioural findings indicating a) shorter RTs in congruent trials and b) faster RTs for incorrect choices compared to correct choices in congruent trials (Fig. 2).

Discussion

In this work, we used single-trial multivariate linear discriminant analysis and neurally-informed cognitive modelling to investigate the neural mechanisms underlying auditory pitch-visual size cross-modal associations, formulated from the presentation of unisensory stimulus features (i.e. auditory pitch). Using a variant of the Implicit Association Test (Parise and Spence, 2012), we showed significant behavioural improvements as a result of associative congruency as participants responded faster to congruent than incongruent stimulus feature-response key mappings (Fig. 2). Our multivariate linear discriminant analysis on the EEG signals revealed neural information for congruent mappings in a 0–600 ms post-stimulus onset window. Moreover, we characterised two EEG components carrying congruency-relevant information in single trials: an 'Early' (~100–110 ms) component and a 'Late' (~330–340 ms) component. Using neurally-informed cognitive modelling, we linked these neural correlates of associative congruency with the corresponding behavioural benefits for forming perceptual decisions. We thus associated the observed shorter RTs in congruent trials with a) an increase in the duration of sensory processing time modulated by the Early component during incongruent trials, and b) a decrease in the quantity of post-sensory evidence needed to facilitate a perceptual choice modulated by the Late component in congruent trials.

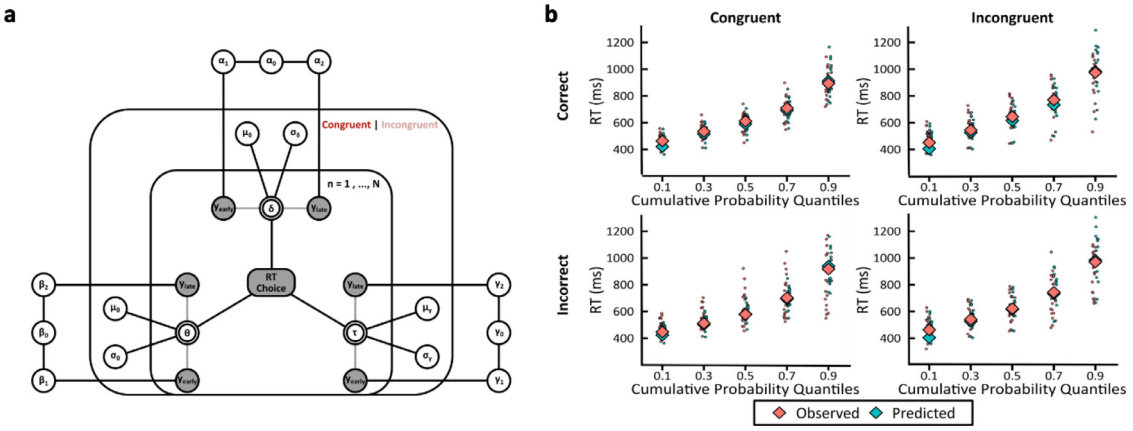


Fig. 5. Neurally-informed cognitive modelling. a Graphical representation illustrating the Bayesian hierarchical framework for estimating neurally-informed HDDM parameters. Round nodes represent continuous random variables, with shaded nodes representing recorded or computed signals, i.e., single-trial behavioural data (RTs and Choice) and EEG component discriminator amplitudes (y's). Double-bordered nodes represent deterministic variables, defined in terms of other variables. Plates denote a hierarchical framework for modelling multiple random variables. The inner plate is over participants ($n = 1, \ldots, N$) and the outer plate is over congruency conditions (Congruent | Incongruent). Parameters are modelled as random variables with inferred means μ and variances σ^2, constrained by inferred estimates over congruency conditions. External plates denote constructed single-trial regression coefficients as predictors of the drift rate (α), decision boundary (θ), and non-decision time (τ). **b** Posterior predictive checks of the Neurally-informed HDDM fitting to participant and group behavioural data. Modelling fit to behavioural data was assessed using a cumulative quantile-probability plot, showing quantiles of RT distributions split across congruency conditions (Congruent/Incongruent in columns) and choice accuracy (Correct/Incorrect in rows). Cumulative probability quantiles are plotted along the x-axis for observed RTs (in pink), i.e. single-trial behavioural data (RTs), and predicted RTs (in cyan), i.e. simulated RTs from HDDM posterior predictive estimates. Diamonds represent group averages and circles represent single-participant values.

Our behavioural results provide further evidence supporting the existence of auditory pitch-visual size cross-modal associations that have been previously reported (Bien et al., 2012; Evans and Treisman, 2010; Gallace and Spence, 2006; Parise and Spence, 2009; 2008; 2012). More importantly, our results demonstrate that auditory pitch-visual size cross-modal associations can be formulated even when only a single unisensory stimulus feature is presented on a single-trial basis. This replicates the findings of Parise and Spence (2012), who reported faster RTs for congruent compared to incongruent trials for five auditory-visual stimulus combinations, including frequency-pitch and object-size. It should be emphasized that the benefits of associative congruency observed in our study should be considered relative in nature. Specifically, it is the variation within blocks of trials, and subsequent trial-by-trial contrasts between 'high' and 'low' pitch tones, that influences behavioural performance, and not necessarily the absolute pitch frequency of the auditory tones presented (Spence, 2019).

We further provide neuroimaging evidence demonstrating a robust modulation to neural activity by the associative congruency of auditory-driven stimulus feature-response key mappings. Importantly, as the IAT only presents one unisensory stimulus feature per trial, it minimizes modulations from confounding neural activity attributed to further multisensory decision-making mechanisms, notably multisensory integration (Franzen et al., 2020; Mercier and Cappe, 2020) and a form of selective-attention/attention-dividing (Bien et al., 2012; Marks, 2004) between two simultaneously presented stimulus features.

To examine neural activity specifically related to the behavioural benefits of cross-modal associative congruency, we applied multivariate Linear Discriminant Analysis to decode congruent from incongruent stimulus-feature response key mapping trials. The application of multivariate Linear Discriminant Analysis to our EEG data revealed two temporally distinct neural components representing both early and late influences of associative congruency mappings. Furthermore, the two components share a broadly consistent scalp topography for localizing associative congruency benefits, clustering a positive discriminative topography that emerged over left-lateralized centro-parietal, and left-lateralized occipital electrodes, gradually emerging toward fronto-central regions of the brain.

The first component (Early: ~100–110 ms) arises near simultaneously with the defined components for encoding auditory stimuli (i.e. SF_1 and SF_2), with higher neural evidence for discriminating associative congruency prior to discriminating auditory pitch. The early latency onset of the discrimination of congruency coincides with our results revealing an increase in discrimination of the presented auditory stimulus feature, possibly implicating an overlapping mapping of perceptual priors of auditory-driven pitch-size associations that automatically influences early sensory encoding/processing. We suggest that the benefits of associative congruency, observed in the behavioural results, modulate neural activity due to a form of perceptual feedback, influencing the early processing of sensory information across the different modalities during the perceptual decision formation process. Previous research has demonstrated that repeated exposure to complementary stimulus features shapes their multisensory composition, thus forming implicit preferences to congruent mappings (Habets et al., 2017; Kayser and Kayser, 2018; Park and Kayser, 2019). Therefore, this reaffirms our assumption that associative congruency shapes multisensory decision formation, thus improving either or both the speed and accuracy of choice. Similarly, multisensory enhancements during perceptual decision formation have been found with interactions occurring in neural signals at very short latencies (Boyle et al., 2017; Cappe et al., 2010; Foxe et al., 2000; 2002; Foxe and Schroeder, 2005; Molholm et al., 2002; 2006; Sperdin et al., 2009). Importantly, in our study the early modulation we observed suggests such enhancements are not exclusively multisensory, since on each trial only a single sensory stimulus was presented. Consequently, we contend that the early onset of our results suggests that cross-modal associations are not exclusively decision-related, but may be perceptual in origin.

Alternatively, an existing underlying mapping of the perceptual priors of auditory pitch-visual size associations may automatically influence early sensory encoding. Cross-modal associations reflect a naturally occurring mapping between stimulus features (Parise et al., 2014; Parise and Spence, 2013). Auditory acoustic pitch-visual size associations demonstrate a strong statistical correspondence in our external environment, whereby larger objects resonate at lower pitch frequencies than smaller objects. Thus, an alternative interpretation suggests that the early onset of our results is related to the influence of such existing

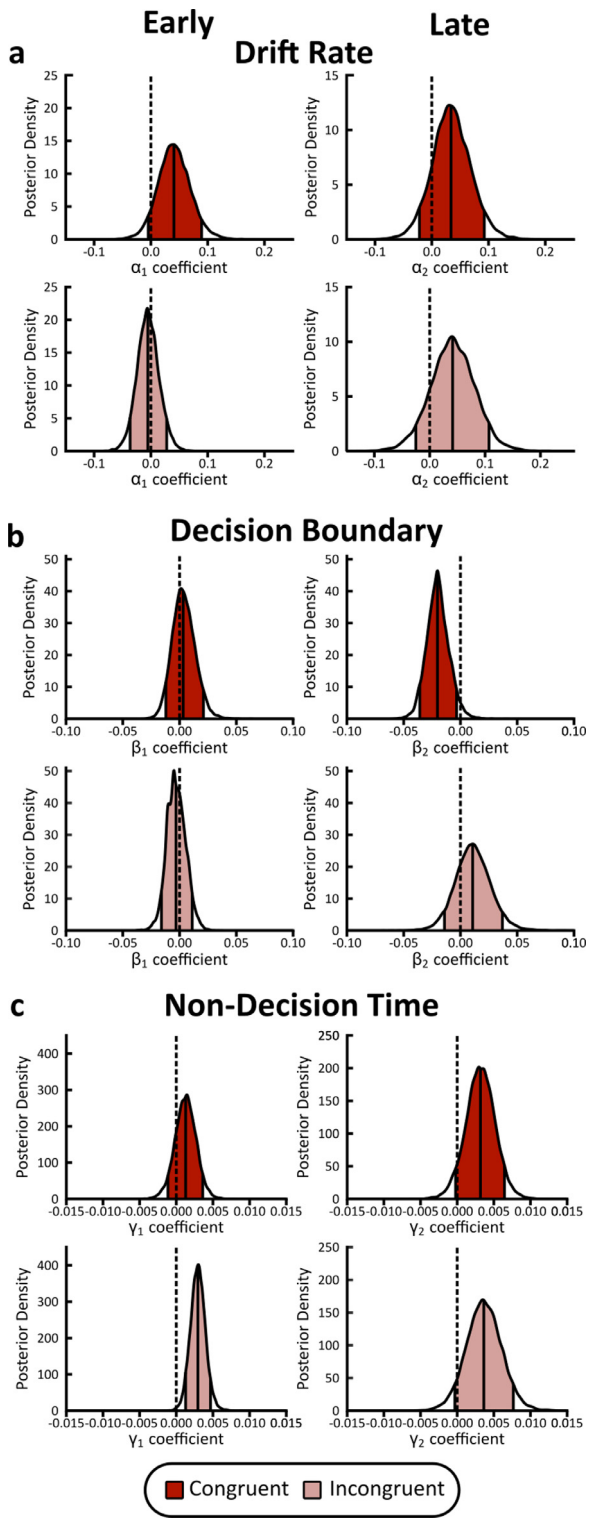


Fig. 6. Neurally-informed cognitive modelling results. Posterior density distributions of estimated regression coefficients for **a** drift rate (α's), **b** decision boundary (β's), and **c** non-decision time (γ's) for Early (Left) and Late (Right) EEG component discriminator amplitudes. All regression coefficients are derived from the neurally-informed HDDM, including $N = 20$ independent participants and 9850 trials. Thick lines denote the median point estimate and the shaded areas represent the 90% probability mass, enclosed between 5% and 95% probability confidence intervals. Dashed lines denote the zero point.

priors shaped in the statistics of our natural environment (Baier et al., 2006). For example, if top-down processes access this existing mapping, and signal to early sensory encoding regions, such feedback might embed the existing environmental prior mapping. The contention of an influence of existing priors between auditory pitch and visual size further contributes to the longstanding debate in the field concerning the degree of automaticity of cross-modal associations (Chen and Spence, 2017; Spence and Deroy, 2013). Our interpretation here supports findings suggesting that the automaticity of audiovisual associative congruency benefits involves both perceptual bottom-up and modulatory top-down processes (Getz and Kubovy, 2018). This interpretation is further supported by the observation that our discriminator's performance for congruency exceeded the significance level prior to auditory stimulus feature presentation (i.e. Az > 0.05, see Fig. 3). A possible explanation for this is that the discrimination of EEG component amplitudes, formulated by the congruency of stimulus feature-response key mappings prior to the formation of perceptual decisions, could indicate pre-mapping anticipation, or expectation, that actively modulates the effects of congruency benefitting the faster formation of perceptual decisions, without modulating the categorisation of auditory stimulus features, or their sensory signals themselves. Bang and Rahnev (2017) present psychophysical evidence to implicate the effects of pre-stimulus anticipation to support this interpretation.

Thus, we contend that cross-modal associations may benefit from being consolidated within a predictive coding framework as a mechanism benefitting choices for multisensory decision-making (Shi and Burr, 2016; Talsma, 2015). In this framework, repeated exposure to auditory pitch-visual size mappings could relate to some existing underlying mapping of the perceptual priors between high/low-pitch tones, and small/large-size objects respectively. In a predictive coding framework, we posit that the early sensory benefits we observed from associative congruency may be influenced by newly formed priors of auditory pitch-visual size associations, with top-down processing signalling to early sensory regions of the brain providing feedback that embeds this environmental prior. Evidence that applies a predictive coding framework stems from studies that implement Bayesian interpretations of the effect of existing priors (Huang and Rao, 2011; McGovern et al., 2016; Tong et al., 2020). Bayesian theories have implicated that cross-modal associative congruency strengthens the binding of stimulus features during multisensory integration (Parise and Spence, 2013), demonstrating the pronounced effect of associative priors for benefitting multisensory decision formation (Acerbi et al., 2018; Gau and Noppeney, 2016; Rohe et al., 2019; Rohe and Noppeney, 2015a; 2015b).

The late onset of the second component (Late: ~330–340 ms) further suggests that cross-modal associations may be decision-related, albeit not exclusively. Previous perceptual decision formation studies have consolidated a neural signature of decision formation, often termed Centro-Parietal Positivity (CPP; O'Connell et al., 2018; Polich, 2007; Tagliabue et al., 2019; Twomey et al., 2016), or the late decision-related component (Philiastides et al., 2006, 2011; 2014; Philiastides and Sajda, 2006; 2007), arising approximately 300–500 ms post-stimulus, reflecting neural activity for accumulating evidence to facilitate a choice. A previous study by Mercier and Cappe (2020) has further attributed that the CPP indexes the accumulation of sensory evidence for multisensory decision-making. In our study, the decoded Late component highly resembles the spatiotemporal characteristics of this indexed neural signature, with a positive discriminative topography emerging across centro-parietal regions. Given we further observed higher neural evidence for discriminating associative congruency as late as 600 ms, we contend that the congruency of cross-modal associations for accumulating sensory evidence at a further decisional stage is important, supporting studies demonstrating the CPP for both unisensory and multisensory decision-making, thus benefitting perceptual decision formation.

Previous multisensory decision-making studies have localized benefits to perceptual decision formation at a later stage (Franzen et al.,

2020; Kayser et al., 2017). However, these cannot be solely attributed to congruency effects as previously explained. Here, by using the IAT, we were able to demonstrate that associative congruency has a further role in accumulating sensory evidence at a later decisional stage, with neural activity aligned with CPP, or the late decision-related component, even when a single unisensory stimulus feature is presented. Thus, we can localize the benefits of cross-modal associations for forming perceptual decisions while simultaneously minimizing the benefits that may be attributed to bottom-up (i.e. multisensory integration), or top-down (i.e. selective attention) multisensory processes.

When forming decisions with multisensory information, multisensory interactions are pervasive within the human brain, constituting different processes along the cortical hierarchy (Cao et al., 2019; Rohe et al., 2019; Rohe and Noppeney, 2016; Keil and Senkowski, 2018; Sadaghiani et al., 2009). For identifying when multisensory information benefits perceptual decision-making, three prominent theories persist in the field (Bizley et al., 2016): a) the early integration hypothesis, b) the late integration hypothesis, and c) what we term as the dual integration hypothesis, which was formulated by Mercier and Cappe (2020). The early integration hypothesis posits that early sensory encoding stages facilitate the influences of multisensory benefits from complementary sensory information across modalities (Ghazanfar and Schroeder, 2006; Kayser and Logothetis, 2007; Schroeder and Foxe, 2005). The late integration hypothesis, however, postulates that unisensory information is processed separately at early sensory encoding stages, then combined into a unified source of evidence at a late post-sensory decisional stage (Bizley et al., 2016; Franzen et al., 2020). Finally, the dual integration hypothesis posits that unisensory information is integrated at both early sensory encoding and later decision formation stages, consolidating a role of causal inference in determining whether multisensory information is supramodal in defining incoming sensory information (Aller and Noppeney, 2019; Cao et al., 2019; Gau and Noppeney, 2016; Kayser and Shams, 2015; Mercier and Cappe, 2020; Rohe et al., 2019; Rohe and Noppeney, 2015; Su, 2014).

Evidence supporting the early integration hypothesis arises from identified neural pathways between sensory cortices (i.e. the visual and auditory cortices), and higher-order associative cortices of the brain (e.g. parietal, temporal and frontal associative cortices), with cross-modal influences on neural responses localized early within the sensory cortices (Eckert et al., 2008; Ghazanfar and Schroeder, 2006; Giart et al., 1999; Kayser et al., 2017 Petro et al., 2017; Rohe and Noppeney, 2016). Previous research has also argued in favour of the late integration hypothesis, implicating post-sensory enhancements of decision evidence from a late integration of multisensory information benefits perceptual decision formation (Franzen et al., 2020). The processes of object recognition and categorization, naturally multisensory processes given the information presented to multiple sensory modalities, has led researchers to contend that top-down processing is required to determine associative congruency, thereby expediting the speed of perceptual decision formation. However, as we've previously discussed, evidence supporting the late integration hypothesis remains stemmed from paradigms that present two or more unisensory features simultaneously. By utilizing a paradigm that presents only one unisensory stimulus feature per trial, recorded neural activity elicits a neural component for discriminating stimulus feature-response key mapping congruency early in a trial. To implicate cross-modal associations are only post-sensory, or decisional, in origin ignores this early associative benefit for forming perceptual decisions and contradicts previous research localizing ERPs for associative congruency early in the decision-making process (Kovic et al., 2010; Bien et al., 2012). To briefly summarize, our results do not provide exclusive support for one of these two theories.

Our data do however support the dual integration hypothesis. Mercier and Cappe (2020) demonstrated support for this hypothesis in their study, in which they identified two temporally-distinct neural processes underlying multisensory decision-making across both cue detection and cue categorization paradigms. Importantly, decoding of EEG activity underlying unisensory signal cues implicated these processes were responsible for early sensory encoding and late decisional formation. Multisensory benefits observed in the behavioural data (i.e. faster RTs, higher accuracy, increased sensitivity towards multisensory cues) were concurrent with an acceleration of both processing stages, suggesting that associative congruency benefitted both a faster integration of sensory information and consolidation of decisional evidence. Here, we identified a similar temporal trajectory of EEG activity, characterized by two mechanisms complementing prior research demonstrating early sensory encoding and decision formation processes that benefit from cross-modal associative congruency (Bizley et al., 2016). Without confounds due to the processes of multisensory integration, and higher-order cognitive processes such as selective attention using the IAT, we also localized the effects of associative congruency as both early sensory-perceptual and late-decisional, thus further consolidating the benefits traditionally observed by early multisensory integrative processes and late decision accuracy. Ultimately, this is in line with a dual integration hypothesis, reconciling the early and late integration hypotheses respectively.

Importantly, our findings suggest that key mechanistic insights can be elicited by coupling models of perceptual decision formation with neuroimaging data. The inclusion of the two characterised EEG components enabled the disambiguation of the internal processes that yielded two IAT behavioural performance results. First, decreased RTs for congruent compared to incongruent stimulus feature-response key mappings, and second, decreased RTs for incorrect compared to correct congruent trials. Our Late component was linked with a decrease in the amount of evidence required to reach a decision as a result of congruent associations, thus congruent trials had shorter RTs and larger proportions of incorrect responses for short RTs. This result is complemented by the observation that incongruent stimulus-response mappings yielded increased non-decision time estimates modulated by the Early component, suggesting longer stimulus encoding times and consequently slower responses in incongruent trials.

Previous studies have used DDMs to study multisensory decision-making (Delis et al., 2018; Franzen et al., 2020; Kayser et al., 2017; Mercier and Cappe, 2020). To our knowledge, such studies have not focused purely on cross-modal associations and modelled behavioural and neuroimaging data from experimental paradigms that present two sensory stimuli simultaneously, or within close spatial or temporal proximity. The application of the IAT means we can model multisensory decision-making, yielding parameter estimates informed by neural measurements linked to the processing of one sensory stimulus feature, thus producing neurally compatible outcomes underlying benefits purely driven by cross-modal associations.

In conclusion, using a neurally-informed cognitive modelling approach, we first characterized the spatiotemporal dynamics of neural activity underlying associative congruency, and then probed its functional role in perceptual decision formation. By presenting only one unisensory stimulus feature per trial, we were able to overcome previous difficulties interpreting the mixed selectivity of neural responses to simultaneously presented stimulus features. Consequently, we could identify the effects of cross-modal associations on neural processing and draw a direct link between these neural processes and the behavioural benefits of associative congruency in perceptual decision-making. We recommend that future research consolidates our observations by utilizing similar unisensory approaches for investigating cross-modal associations with alternative statistical correspondences (e.g. auditory pitch-visual lightness; Brunel et al., 2015; Zeljko et al., 2019; auditory pitch-visual brightness; Marks, 1987; Klapetek et al., 2012; auditory pitch-visual elevation; Jamal et al., 2017; McCormick et al., 2018; Zeljko et al., 2019; auditory pitch-visual shape; Köhler, 1929; Marks, 1987; Parise and Spence, 2012, and higher-order semantic coherence; Marks, 2004; Parise and Spence, 2013; Revill et al., 2014; Sadaghiani et al., 2009; Spence and Deroy, 2013; Spence, 2011).

CRediT author statement

Joshua Bolam: Conceptualization, Formal Analysis, Visualization, Writing – Initial Draft Preparation, Writing – Reviewing & Editing. **Stephanie C. Boyle:** Data Curation, Funding Acquisition, Investigation, Methodology. **Robin A.A Ince:** Data Curation, Investigation, Methodology, Validation **Ioannis Delis:** Conceptualization, Funding Acquisition, Resources, Supervision, Validation, Writing – Initial Draft Preparation, Writing – Review & Editing.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgements

This work was supported by the Wellcome Trust ([214120/Z/18/Z] to R.A.A.I.); the European Commission (H2020-MSCA-IF-2018/845884, "NeuCoDe" to I.D.) and the Physiological Society (2018 Research Grant Scheme to I.D.). S.C.B was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) through a BBSRC DTP Studentship (Grant Number BB/L027534/1).

Data and Code Availability

The supplementary datasets (behavioural and neural) and custom-script MATLAB/Python codes used for analysis and modelling during the current study are available from the study's online data repository (www.github.com/DelisLab).

References

Acerbi, L., Dokka, K., Angelaki, D.E., Ma, W.J., 2018. Bayesian comparison of explicit and implicit causal inference strategies in multisensory heading perception. PLoS Comput. Biol. 14 (7), e1006110.
Adam, R., Noppeney, U., 2014. A phonologically congruent sound boosts a visual target into perceptual awareness. Front Integr Neurosci 8 (70), 1–13.
Aller, M., Giani, A., Conrad, V., Watanabe, M., Noppeney, U., 2015. A spatially collocated sound thrusts a flash into awareness. Front. Integr. Neurosci. 9 (16), 1–8.
Aller, M., Noppeney, U., 2019. To integrate or not to integrate: temporal dynamics of hierarchical Bayesian causal inference. PLoS Biol. 17 (4), e3000210.
Alais, D., Newell, F., Mamassian, P., 2010. Multisensory processing in review: from physiology to behaviour. Seeing Perceiving 23 (1), 3–38.
Angelaki, D.E., Gu, Y., DeAngelis, G.C., 2009. Multisensory integration: psychophysics, neurophysiology, and computation. Curr. Opin. Neurobiol. 19 (4), 452–458.
Baier, B., Kleinschmidt, A., Müller, N.G., 2006. Cross-modal processing in early visual and auditory cortices depends on expected statistical relationship of multisensory information. J. Neurosci. 26 (47), 12260–12265.
Bang, J.W., Rahnev, D., 2017. Stimulus expectation alters decision criterion but not sensory signal in perceptual decision making. Sci. Rep. 7 (1), 1–12.
Bien, N., Ten Oever, S., Goebel, R., Sack, A.T., 2012. The sound of size: crossmodal binding in pitch-size synesthesia: a combined TMS, EEG and psychophysics study. Neuroimage 59 (1), 663–672.
Bizley, J.K., Jones, G.P., Town, S.M., 2016. Where are multisensory signals combined for perceptual decision-making? Curr. Opin. Neurobiol. 40, 31–37.
Boehm, U., Annis, J., Frank, M.J., Hawkins, G.E., Heathcote, A., Kellen, D., Krypotos, A.M., Lerche, V., Logan, G.D., Palmeri, T.J., Ravenzwaaj, D., Servant, M., Jeffrey, H.S., Starns, J.J., Voss, A., Wiecki, T.V., Matzke, D., Wagenmakers, E.J., 2018. Estimating across-trial variability parameters of the diffusion decision model: expert advice and recommendations. J. Math. Psychol. 87, 46–75.
Boyle, S.C., Kayser, S.J., Kayser, C., 2017. Neural correlates of multisensory reliability and perceptual weights emerge at early latencies during audio-visual integration. European J. Neurosci. 46 (10), 2565–2577.
Brainard, D.H., 1997. The psychophysics toolbox. Spat Vis. 10 (4), 433–436.
Brunel, L., Carvalho, P.F., Goldstone, R.L., 2015. It does belong together: cross-modal correspondences influence cross-modal integration during perceptual learning. Front. Psychol. 6, 358–368.
Calvert, G., Spence, C., Stein, B.E. (Eds.), 2004. The Handbook of Multisensory Processes. MIT press.
Cao, Y., Summerfield, C., Park, H., Giordano, B.L., Kayser, C., 2019. Causal inference in the multisensory brain. Neuron 102 (5), 1076–1087.
Cappe, C., Thut, G., Romei, V., Murray, M.M., 2010. Auditory–visual multisensory interactions in humans: timing, topography, directionality, and sources. J. Neurosci. 30 (38), 12572–12580.
Chandrasekaran, C., 2017. Computational principles and models of multisensory integration. Curr. Opin. Neurobiol. 43, 25–34.
Chen, Y.C., Spence, C., 2017. Assessing the Role of the 'Unity Assumption' on Multisensory Integration: A Review. Frontiers in psychology 8, 445.
Choi, I., Lee, J.Y., Lee, S.H., 2018. Bottom-up and top-down modulation of multisensory integration. Curr. Opin. Neurobiol. 52 (1), 115–122.
Cohen, J., 1992a. Statistical power analysis. Curr Dir Psychol Sci 1 (3), 98–101.
Cohen, J., 1992b. A power primer. Psychol Bull 112 (1), 155–159.
Dahl, C.D., Logothetis, N.K., Kayser, C., 2009. Spatial organization of multisensory responses in temporal association cortex. J. Neurosci. 29 (38), 11924–11932.
Diaconescu, A.O., Alain, C., McIntosh, A.R., 2011. The co-occurrence of multisensory facilitation and cross-modal conflict in the human brain. J. Neurophysiol. 106, 2896–2909.
Diaz, J.A., Queirazza, F., Philiastides, M.G., 2017. Perceptual learning alters post-sensory processing in human decision-making. Nat. Hum. Behav. 1 (2), 1–9.
Delis, I., Dmochowski, J.P., Sajda, P., Wang, Q., 2018. Correlation of neural activity with behavioral kinematics reveals distinct sensory encoding and evidence accumulation processes during active tactile sensing. Neuroimage 175, 12–21.
Drugowitsch, J., DeAngelis, G.C., Klier, E.M., Angelaki, D.E., Pouget, A., 2014. Optimal multisensory decision-making in a reaction-time task. Elife 3, e03005 1-19.
Eckert, M.A., Kamdar, N.V., Chang, C.E., Beckmann, C.F., Greicius, M.D., Menon, V., 2008. A cross-modal system linking primary auditory and visual cortices: evidence from intrinsic fMRI connectivity analysis. Hum. Brain Mapp. 29 (7), 848–857.
Engel, A.K., Senkowski, D., Schneider, T.R., 2012. Multisensory integration through neural coherence. The Neural Bases of Multisensory Processes. CRC Press/Taylor and Francis.
Ernst, M.O., Bülthoff, H.H., 2004. Merging the senses into a robust percept. Trends Cogn. Sci. (Regul. Ed.) 8 (4), 162–169.
Evans, K.K., Treisman, A., 2010. Natural cross-modal mappings between visual and auditory features. J. Vis. 10 (1) 6-6.
Faul, F., Erdfelder, E., Buchner, A., Lang, A.G., 2009. Statistical power analyses using G* Power 3.1: tests for correlation and regression analyses. Behav. Res. Methods 41 (4), 1149–1160.
Faul, F., Erdfelder, E., Lang, A.G., Buchner, A., 2007. G* Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39 (2), 175–191.
Forstmann, B.U., Ratcliff, R., Wagenmakers, E.J., 2016. Sequential sampling models in cognitive neuroscience: advantages, applications, and extensions. Ann. Rev. Psychol. 67 (1), 641–666.
Foxe, J.J., Morocz, I.A., Murray, M.M., Higgins, B.A., Javitt, D.C., Schroeder, C.E., 2000. Multisensory auditory–somatosensory interactions in early cortical processing revealed by high-density electrical mapping. Cogn. Brain Res. 10 (1–2), 77–83.
Foxe, J.J., Schroeder, C.E., 2005. The case for feedforward multisensory convergence during early cortical processing. Neuroreport 16 (5), 419–423.
Foxe, J.J., Wylie, G.R., Martinez, A., Schroeder, C.E., Javitt, D.C., Guilfoyle, D., … Murray, M.M., 2002. Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. J. Neurophysiol. 88 (1), 540–543.
Frank, M.J., Gagne, C., Nyhus, E., Masters, S., Wiecki, T.V., Cavanagh, J.F., Badre, D., 2015. fMRI and EEG predictors of dynamic decision parameters during human reinforcement learning. J. Neurosci. 35 (1), 484–494.
Franzen, L., Delis, I., De Sousa, G., Kayser, C., Philiastides, M.G., 2020. Auditory information enhances post-sensory visual evidence during rapid multisensory decision-making. Nat. Commun. 11 (1), 1–14.
Fusi, S., Miller, E.K., Rigotti, M., 2016. Why neurons mix: high dimensionality for higher cognition. Curr. Opin. Neurobiol. 37 (1), 66–74.
Gallace, A., Spence, C., 2006. Multisensory synesthetic interactions in the speeded classification of visual size. Percept. Psychophys 68 (7), 1191–1203.
Gamerman, D., Lopes, H.F., 2006. Markov Chain Monte Carlo: Stochastic Simulation For Bayesian Inference. CRC Press.
Gau, R., Noppeney, U., 2016. How prior expectations shape multisensory perception. Neuroimage 124, 876–886.
Gelman, A., 2003. A Bayesian formulation of exploratory data analysis and goodness-of-fit testing. Int. Stat. Rev. 71 (2), 369–382.
Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 7 (4), 457–472.
Getz, L.M., Kubovy, M., 2018. Questioning the automaticity of audiovisual correspondences. Cognition 175 (1), 101–108.
Ghazanfar, A.A., Schroeder, C.E., 2006. Is neocortex essentially multisensory? Trends Cogn. Sci. (Regul. Ed.) 10 (6), 278–285.
Gherman, S., Philiastides, M.G., 2015. Neural representations of confidence emerge from the process of decision formation during perceptual choices. Neuroimage 106, 134–143.
Giard, M.H., Peronnet, F., 1999. Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J. Cogn. Neurosci. 11 (5), 473–490.
Glicksohn, A., Cohen, A., 2013. The role of cross-modal associations in statistical learning. Psychon Bull Rev 20 (6), 1161–1169.
Green, D.M., Swets, J.A., 1966. Signal Detection Theory and Psychophysics, Vol. 1. Wiley, New York.
Greenweld, A.G., McGhee, D.E., Schwartz, J.L.K., 1998. Measuring individual differences in implicit cognition: the implicit association test. J. Pers. Soc. Psychol. 47 (6), 1464–1480.
Habets, B., Bruns, P., Röder, B., 2017. Experience with crossmodal statistics reduces the sensitivity for audio-visual temporal asynchrony. Sci. Rep. 7 (1), 1–7.
Hipp, J.F., Siegel, M., 2013. Dissociating neuronal gamma-band activity from cranial and ocular muscle activity in EEG. Front. Hum. Neurosci. 7, 338.
Huang, Y., Rao, R.P., 2011. Predictive coding. Wiley Interdisc. Rev. 2 (5), 580–593.
Ince, R.A., Kay, J.W., and Schyns, P.G. (2020). Bayesian inference of population prevalence. *bioRxiv*.

Jamal, Y., Lacey, S., Nygaard, L., Sathian, K., 2017. Interactions between auditory elevation, auditory pitch and visual elevation during multisensory perception. Multisens Res. 30 (3–5), 287–306.
Kayser, S.J., Kayser, C., 2018. Trial by trial dependencies in multisensory perception and their correlates in dynamic brain activity. Sci. Rep. 8 (1), 3742.
Kayser, C., Logothetis, N.K., 2007. Do early sensory cortices integrate cross-modal information? Brain Struct. Funct. 212 (2), 121–132.
Kayser, S.J., Philiastides, M.G., Kayser, C., 2017. Sounds facilitate visual motion discrimination via the enhancement of late occipital visual representations. Neuroimage 148, 31–41.
Kayser, C., Shams, L., 2015. Multisensory causal inference in the brain. PLoS Biol. 13 (2), e1002075.
Keil, J., Senkowski, D., 2018. Neural oscillations orchestrate multisensory processing. The Neuroscientist 24 (6), 609–626.
Keren, A.S., Yuval-Greenberg, S., Deouell, L.Y., 2010. Saccadic spike potentials in gamma-band EEG: characterization, detection and suppression. Neuroimage 49 (3), 2248–2263.
Kim, R.S., Seitz, A.R., Shams, L., 2008. Benefits of stimulus congruency for multisensory facilitation of visual learning. PLoS ONE 3 (1), e1532.
Klapetek, A., Ngo, M.K., Spence, C., 2012. Does crossmodal correspondence modulate the facilitatory effect of auditory cues on visual search? Attent. Percept. Psychophys. 74 (6), 1154–1167.
Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C.E., Kepecs, A., Mainen, Z.F., … Machens, C.K., 2016. Demixed principal component analysis of neural population data. Elife 5, e10989.
Köhler, W., 1929. Gestalt Psychology. Liveright, New York.
Kovic, V., Plunkett, K., Westermann, G., 2010. The shape of words in the brain. Cognition 114 (1), 19–28.
Laurienti, P.J., Kraft, R.A., Maldjian, J.A., Burdette, J.H., Wallace, M.T., 2004. Semantic congruence is a critical factor in multisensory behavioral performance. Exp. Brain Res. 158 (4), 405–414.
Marks, L.E., 1987. On cross-modal similarity: auditory–visual interactions in speeded discrimination. J. Exper. Psychol. 13 (3), 384–394.
Marks, L.E., 2004. Cross-modal interactions in speeded classification. In: Calvert, G., Spence, C., Stein, B.E. (Eds.), The Handbook of Multisensory Processes. The MIT Press, Cambridge, MA, pp. 85–105.
McCormick, K., Lacey, S., Stilla, R., Nygaard, L.C., Sathian, K., 2018. Neural basis of the crossmodal correspondence between auditory pitch and visuospatial elevation. Neuropsychologia 112, 19–30.
McGovern, D.P., Roudaia, E., Newell, F.N., Roach, N.W., 2016. Perceptual learning shapes multisensory causal inference via two distinct mechanisms. Sci. Rep. 6 (24673), 1–11.
Mercier, M.R., Cappe, C., 2020. The interplay between multisensory integration and perceptual decision making. Neuroimage 116970, 1–16.
Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E., Foxe, J.J., 2002. Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. Cognit. Brain Res. 14 (1), 115–128.
Molholm, S., Sehatpour, P., Mehta, A.D., Shpaner, M., Gomez-Ramirez, M., Ortigue, S., … Foxe, J.J., 2006. Audio-visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. J. Neurophysiol. 96 (2), 721–729.
Navarro, D.J., Fuss, I.G., 2009. Fast and accurate calculations for first-passage times in Wiener diffusion models. J. Math Psychol. 53 (4), 222–230.
O'Connell, R.G., Shadlen, M.N., Wong-Lin, K., Kelly, S.P., 2018. Bridging neural and computational viewpoints on perceptual decision-making. Trends Neurosci. 41 (11), 838–852.
Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011 (156869), 1–9.
Parise, C.V., Knorre, K., Ernst, M.O., 2014. Natural auditory scene statistics shapes human spatial hearing. Proc. Natl. Acad. Sci. 111 (16), 6104–6108.
Parise, C., Spence, C., 2008. Synesthetic congruency modulates the temporal ventriloquism effect. Neurosci. Lett. 442 (3), 257–261.
Parise, C.V., Spence, C., 2009. 'When birds of a feather flock together': synesthetic correspondences modulate audiovisual integration in non-synesthetes. PLoS ONE 4 (5), e5664.
Parise, C.V., Spence, C., 2012. Audiovisual crossmodal correspondences and sound symbolism: a study using the implicit association test. Exp. Brain Res. 220 (3–4), 319–333.
Parise, C.V., Spence, C., 2013. Audiovisual cross-modal correspondences in the general population. In: Simner, Julia, Hubbard, Edward M. (Eds.), Oxford Handbook of Synesthesia. Oxford University Press, Oxford. Oxfordshire, pp. 790–815.
Parra, L.C., Alvino, C., Tang, A., Pearlmutter, B., Yeung, N., Osman, A., Sajda, P., 2002. Linear spatial integration for single-trial detection in encephalography. Neuroimage 17 (1), 223–230.
Parra, L.C., Spence, C.D., Gerson, A.D., Sajda, P., 2005. Recipes for the linear analysis of EEG. Neuroimage 28 (2), 326–341.
Park, H., Kayser, C., 2019. Shared neural underpinnings of multisensory integration and trial-by-trial perceptual recalibration in humans. Elife 8, e47001.
Park, I.M., Meister, M.L., Huk, A.C., Pillow, J.W., 2014. Encoding and decoding in parietal cortex during sensorimotor decision-making. Nat. Neurosci. 17 (10), 1395–1403.
Petro, L.S., Paton, A.T., Muckli, L., 2017. Contextual modulation of primary visual cortex by auditory signals. Philos. Trans. Royal Soc. B 372 (1714), 20160104.
Philiastides, M.G., Auksztulewicz, R., Heekeren, H.R., Blankenburg, F., 2011. Causal role of dorsolateral prefrontal cortex in human perceptual decision making. Curr. Biol. 21 (11), 980–983.
Philiastides, M.G., Diaz, J.A., Gherman, S., 2017. Spatiotemporal characteristics and modulators of perceptual decision-making in the human brain. In: Decision Neuroscience. Academic Press, pp. 137–147.
Philiastides, M.G., Heekeren, H.R., 2009. Spatiotemporal characteristics of perceptual decision making in the human brain. In: Handbook of Reward and Decision Making. Academic Press, pp. 185–212.
Philiastides, M.G., Heekeren, H.R., Sajda, P., 2014. Human scalp potentials reflect a mixture of decision-related signals during perceptual choices. J. Neurosci. 34 (50), 16877–16889.
Philiastides, M.G., Ratcliff, R., Sajda, P., 2006. Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. J. Neurosci. 26 (35), 8965–8975.
Philiastides, M.G., Sajda, P., 2006. Temporal characterization of the neural correlates of perceptual decision making in the human brain. Cerebral Cortex 16 (4), 509–518.
Philiastides, M.G., Sajda, P., 2007. EEG-informed fMRI reveals spatiotemporal characteristics of perceptual decision making. J. Neurosci. 27 (48), 13082–13091.
Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. 118 (10), 2128–2148.
Raposo, D., Kaufman, M.T., Churchland, A.K., 2014. A category-free neural population supports evolving demands during decision-making. Nat. Neurosci. 17 (12), 1784.
Raposo, D., Sheppard, J.P., Schrater, P.R., Churchland, A.K., 2012. Multisensory decision–making in rats and humans. J. Neurosci. 32 (11), 3726–3735.
Ratcliff, R., 1978. A theory of memory retrieval. Psychol. Rev. 85 (2), 1–59.
Ratcliff, R., Childers, R., 2015. Individual differences and fitting methods for the two–choice diffusion model of decision making. Decision 2 (4), 237–279.
Ratcliff, R., McKoon, G., 2008. The diffusion decision model: theory and data for two–choice decision tasks. Neural Comput 20 (4), 873–922.
Ratcliff, R., Smith, P.L., McKoon, G., 2015. Modeling regularities in response time and accuracy data with the diffusion model. Curr. Dir. Psychol. Sci. 24 (6), 458–470.
Ratcliff, R., Smith, P.L., Brown, S.D., McKoon, G., 2016. Diffusion decision model: current issues and history. Trends Cogn. Sci. (Regul. Ed.) 20 (4), 260–281.
Revill, K.P., Namy, L.L., DeFife, L.C., Nygaard, L.C., 2014. Cross-linguistic sound symbolism and crossmodal correspondence: evidence from fMRI and DTI. Brain Lang 128 (1), 18–24.
Rigotti, M., Barak, O., Warden, M.R., Wang, X.J., Daw, N.D., Miller, E.K., Fusi, S., 2013. The importance of mixed selectivity in complex cognitive tasks. Nature 497 (7451), 585–590.
Rohe, T., Ehlis, A.C., Noppeney, U., 2019. The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. Nat. Commun. 10 (1), 1–17.
Rohe, T., Noppeney, U., 2015a. Sensory reliability shapes perceptual inference via two mechanisms. J. Vis. 15 (5) 22-22.
Rohe, T., Noppeney, U., 2015b. Cortical hierarchies perform Bayesian causal inference in multisensory perception. PLoS Biol. 13 (2), e1002073.
Rohe, T., Noppeney, U., 2016. Distinct computational principles govern multisensory integration in primary sensory and association cortices. Curr. Biol. 26 (4), 509–514.
Rosenthal, R., Cooper, H., Hedges, L., 1994. Parametric measures of effect size. The Handbook Res. Synthesis 621 (2), 231–244.
Sadaghiani, S., Maier, J.X., Noppeney, U., 2009. Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing. J. Neurosci. 29 (20), 6490–6499.
Sajda, P., Philiastides, M.G., Heekeren, H.H., Ratcliff, R., 2011. Linking neuronal variability to perceptual decision making via neuroimaging. In: Ding, M., Glanzman, D.Oxford (Eds.), Neuronal Variability and Its Functional Significance. Oxford University Press, pp. 214–232.
Sajda, P., Philiastides, M.G., Parra, L.C., 2009. Single-trial analysis of neuroimaging data: inferring neural networks underlying perceptual decision-making in the human brain. IEEE Rev. Biomed. Eng. 2, 97–109.
Schroeder, C.E., Foxe, J., 2005. Multisensory contributions to low-level, 'unisensory' processing. Curr. Opin. Neurobiol. 15 (4), 454–458.
Shi, Z., Burr, D., 2016. Predictive coding of multisensory timing. Curr. Opin. Behav. Sci. 8, 200–206.
Silva, A.E., Barakat, B.K., Jimenez, L.O., Shams, L., 2017. Multisensory congruency enhances explicit awareness in a sequence learning task. Multisens Res. 30 (7–8), 681–689.
Spence, C., Deroy, O., 2013. How automatic are crossmodal correspondences? Conscious Cogn. 22 (1), 245–260.
Spence, C., 2011. Crossmodal correspondences: a tutorial review. Attent. Percept. Psychophys. 73 (4), 971–995.
Spence, C., 2019. On the relative nature of (pitch-based) crossmodal correspondences. Multisens Res 32 (3), 235–265.
Sperdin, H.F., Cappe, C., Foxe, J.J., Murray, M.M., 2009. Early, low-level auditory-somatosensory multisensory interactions impact reaction time speed. Front. Integr. Neurosci 3 (2), 1–10.
Su, Y.H., 2014. Content congruency and its interplay with temporal synchrony modulate integration between rhythmic audiovisual streams. Front. Integr. Neurosci. 8 (92), 1–13.
Tagliabue, C.F., Veniero, D., Benwell, C.S., Cecere, R., Savazzi, S., Thut, G., 2019. The EEG signature of sensory evidence accumulation during decision formation closely tracks subjective perceptual experience. Sci. Rep. 9 (1), 1–12.
Talsma, D., 2015. Predictive coding and multisensory integration: an attentional account of the multisensory mind. Front. Integr. Neurosci. 9 (19), 1–13.
Tong, J., Li, L., Bruns, P., Röder, B., 2020. Crossmodal associations modulate multisensory spatial integration. Attent. Percept. Psychophys. 82 (7), 3490–3506.
Tremel, J.J., Wheeler, M.E., 2015. Content-specific evidence accumulation in inferior temporal cortex during perceptual decision-making. Neuroimage 109 (1), 35–49.
Turner, B.M., Forstmann, B.U., Love, B.C., Palmeri, T.J., Van Maanen, L., 2017. Approaches to analysis in model-based cognitive neuroscience. J. Math Psychol. 76, 65–79.

Turner, B.M., Forstmann, B.U., Wagenmakers, E.J., Brown, S.D., Sederberg, P.B., Steyvers, M., 2013. A Bayesian framework for simultaneously modeling neural and behavioral data. Neuroimage 72, 193–206.
Turner, B.M., Rodriguez, C.A., Norcia, T.M., McClure, S.M., Steyvers, M., 2016. Why more is better: simultaneous modeling of EEG, fMRI, and behavioral data. Neuroimage 128, 96–115.
Twomey, D.M., Kelly, S.P., O'Connell, R.G, 2016. Abstract and effector-selective decision signals exhibit qualitatively distinct dynamics before delayed perceptual reports. J. Neurosci. 36 (28), 7346–7352.
Vandekerckhove, J., Tuerlinckx, F., Lee, M.D., 2011. Hierarchical diffusion models for two-choice response times. Psychol. Methods 16 (1), 44–62.
Whelan, R., 2008. Effective analysis of reaction time data. Psychol. Rec. 58 (3), 475–482.
Wiecki, T.V., Sofer, I., Frank, M.J., 2013. HDDM: hierarchical Bayesian estimation of the drift-diffusion model in Python. Front. Neuroinform. 7, 1–14.
Zeljko, M., Kritikos, A., Grove, P.M., 2019. Lightness/pitch and elevation/pitch cross-modal correspondences are low-level sensory effects. Attenti. Percept. Psychophys. 81 (5), 1609–1623.