Investigating the Cognitive and Neural Mechanisms underlying Multisensory Perceptual Decision-Making in Humans

Mr Joshua William Bolam

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
Faculty of Biological Sciences
School of Biomedical Sciences

December 2022

Intellectual Property and Publication Statements

The candidate confirms that the work submitted is his/her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapter 2 contains content which has also been published in a jointly authored publication. The reference for this publication is provided below:

Bolam, J., Boyle, S. C., Ince, R. A., & Delis, I. (2022). Neurocomputational mechanisms underlying cross-modal associations and their influence on perceptual decisions. *NeuroImage*, *247*, 118841. https://doi.org/10.1016/j.neuroimage.2021.118841

The accompanying Contributor Roles Taxonomy (CRediT) author statement recognises the following contributions towards this publication:

Joshua Bolam – Conceptualisation, Formal Analysis, Visualisation, Writing – Initial Draft Preparation, Writing – Review and Editing.

Dr Stephanie C. Boyle – Data Curation, Funding Acquisition, Investigation, Methodology.

Dr Robin A.A. Ince – Data Curation, Investigation, Methodology, Funding Acquisition.

Dr Ioannis Delis – Conceptualisation, Funding Acquisition, Resources, Supervision, Validation, Writing – Initial Draft Preparation, Writing – Review and Editing.

Acknowledgements

First and foremost, I would like to thank my supervisors Dr Ioannis Delis and Dr Sarah Astill, for everything throughout these four years. I will forever be grateful to be a PhD candidate of the School of Sports & Exercise Sciences at the University of Leeds, as well as for all the learning opportunities you provided during this time. Your guidance, knowledge, advice, and encouragement to progress (especially throughout the SARS-CoV-2 pandemic which tried and failed to disrupt so much of what we aimed to achieve) has provided me so many opportunities to develop my skills and experience as an independent cognitive neuroscientist. Your confidence in me has done more than enough to keep me going when I thought I couldn't. Thank you. It really has been an honour.

Second, a thank you to everyone who volunteered to be a participant in my experimental studies, both online and in the EEG laboratory. Arriving at the laboratory for EEG testing on three consecutive days required a lot of patience and concentration, and I'm hugely appreciative none of you missed a day or arrived late. Thanks especially to Mandvi Bajpai and Jennifer Deakin for your assistance I needed on the long days of EEG data collection. You're both fantastic people, and I hope you've taken your developed skills forward to progress as researchers on your own in whatever field your ambitions lie. I would also like to show my gratitude to Dr Robin A.A. Ince and Dr Stephanie C. Boyle, both at the University of Glasgow's Institute of Neuroscience & Psychology, for provision of the EEG and behavioural datasets, and discussions concerning my research, that contributed towards my first lead author publication (i.e., Bolam et al., 2022) and Chapter 2 of my thesis. I hope the published article and thesis chapter have done your valuable contributions justice.

A thank you as well to all the PhDs and postdoctoral researchers I met in the School of Exercise Sciences, School of Psychology, and wider Faculty of Biological Sciences at the University of Leeds. I apologise if I have missed your name (please don't think you're not considered!), but a particular thank you to Dr Jessica Ann Diaz, Rafael Aguiar, Barbara Passadouro, David O'Reilly, and Amir Nadimi for making the Delis Laboratory a vibrant and encouraging group to progress with my research. In particular, I'm further grateful to Dr Jessica Ann Diaz for the thorough groundwork in developing the

audiovisual face-versus-car categorisation paradigm which contributed towards Chapters 3 and 4 of my thesis.

Finally, a special thank you to my family, especially my Mum and Dad. Your unconditional love and unwavering support have not gone amiss. Your constant faith has meant so much when I’ve needed it the most. I love you both.

Abstract

On a frequent day-to-day basis, we encounter situations that require the formation of decisions based on ambiguous and often incomplete sensory information. *Perceptual decision-making* defines the process by which sensory information is consolidated and accumulated towards one of multiple possible choice alternatives, which inform our behavioural responses. Perceptual decision-making can be understood both theoretically and neurologically as a process of stochastic sensory evidence accumulation towards some choice threshold. Once this threshold is exceeded, a response is facilitated, informing the overt actions undertaken. Prevalent progress has been made towards understanding the cognitive and neural mechanisms underlying perceptual decision-making. Analyses of Reaction Time (RTs; typically constrained to milliseconds) and choice accuracy; reflecting decision-making behaviour, can be coupled with neuroimaging methodologies; notably electroencephalography (EEG) and functional Magnetic Resonance Imaging (fMRI), to identify spatiotemporal components representative of the neural signatures corresponding to such accumulation-to-bound decision formation on a single-trial basis. Taken together, these provide us with an experimental framework conceptualising the key computations underlying perceptual decision-making.

Despite this, relatively little remains known about the enhancements or alternations to the process of perceptual decision-making from the integration of information across multiple sensory modalities. Consolidating the available sensory evidence requires processing information presented in more than one sensory modality, often near-simultaneously, to exploit the salient percepts for what we term as *multisensory (perceptual) decision-making*. Specifically, *multisensory integration* must be considered within the perceptual decision-making framework in order to understand how information becomes stochastically accumulated to inform overt sensory-motor choice behaviours. Recently, substantial progress in research has been made through the application of *behaviourally-informed*, and/or *neurally-informed*, modelling approaches to benefit our understanding of multisensory decision-making. In particular, these approaches fit a number of model parameters to behavioural and/or neuroimaging datasets, in order to (a) dissect the constituent internal cognitive and neural processes underlying perceptual decision-making with both multisensory and unisensory information, and (b) mechanistically infer how multisensory enhancements arise from the integration of information across multiple sensory modalities to benefit perceptual decision formation. Despite this, the spatiotemporal locus of the neural and cognitive underpinnings of enhancements from multisensory integration remains subject to debate. In particular, our understanding of *which* brain regions are predictive of such enhancements, *where* they arise, and *how* they influence decision-making behaviours requires further exploration.

The current thesis outlines empirical findings from three studies aimed at providing a more complete characterisation of multisensory perceptual decision-making, utilising EEG and accumulation-to-bound modelling methodologies to incorporate both behaviourally-informed and neurally-informed modelling approaches, investigating where, when, and how perceptual improvements arise during multisensory perceptual decision-making. Pointedly, these modelling approaches sought to probe the exerted modulatory influences of three factors: *unisensory formulated cross-modal associations* (Chapter 2), *natural ageing* (Chapter 3), and *perceptual learning* (Chapter 4), on the integral cognitive and neural mechanisms underlying observable benefits towards multisensory decision formation.

Chapter 2 outlines secondary analyses, utilising a neurally-informed modelling approach, characterising the spatiotemporal dynamics of neural activity underlying auditory pitch-visual size cross-modal associations. In particular, how unisensory auditory pitch-driven associations benefit perceptual decision formation was functionally probed. EEG measurements were recorded from participants during performance of an Implicit Association Test (IAT), a two-alternative forced-choice (2AFC) paradigm which presents one unisensory stimulus feature per trial for participants to categorise, but manipulates the stimulus feature-response key mappings of auditory pitch-visual size cross-modal associations from unisensory stimuli alone, thus overcoming the issue of mixed selectivity in recorded neural activity prevalent in previous cross-modal associative research, which near-simultaneously presented multisensory stimuli. Categorisations were faster (i.e., lower RTs) when stimulus feature-response key mappings were associatively congruent, compared to associatively incongruent, between the two associative counterparts, thus demonstrating a behavioural benefit to perceptual decision formation. Multivariate Linear Discriminant Analysis (LDA) was used to characterise the spatiotemporal dynamics of EEG activity underpinning IAT performance, in which two EEG components were identified that discriminated

neural activity underlying the benefits of associative congruency of stimulus feature-response key mappings. Application of a neurally-informed Hierarchical Drift Diffusion Model (HDDM) demonstrated early sensory processing benefits, with increases in the duration of non-decisional processes with incongruent stimulus feature-response key mappings, and late post-sensory alterations to decision dynamics, with congruent stimulus feature-response key mappings decreasing the quantity of evidence required to facilitate a decision. Hence, we found that the trial-by-trial variability in perceptual decision formation from unisensory facilitated cross-modal associations could be predicted by neural activity within our neurally-informed modelling approach.

Next, Chapter 3 outlines cognitive research investigating age-related impacts on the behavioural indices of multisensory perceptual decision-making (i.e., RTs and choice accuracy). Natural ageing has been demonstrated to diversely affect multisensory perceptual decision-making dynamics. However, the constituent cognitive processes affected remain unclear. Specifically, a mechanistic insight reconciling why older adults may exhibit preserved multisensory integrative benefits, yet display generalised perceptual deficits, relative to younger adults, remains inconclusive. To address this limitation, 212 participants performed an online variant of a well-established audiovisual object categorisation paradigm, whereby age-related differences in RTs and choice accuracy (binary responses) between audiovisual (AV), visual (V), and auditory (A) trial types could be assessed between Younger Adults (YAs; Mean ± Standard Deviation = 27.95 ± 5.82 years) and Older Adults (OAs; Mean ± Standard Deviation = 60.96 ± 10.35 years). Hierarchical Drift Diffusion Modelling (HDDM) was fitted to participants’ RTs and binary responses in order to probe age-related impacts on the latent underlying processes of multisensory decision formation. Behavioural results found that whereas OAs were typically slower (i.e., ↑ RTs) and less accurate (i.e., ↓ choice accuracy), relative to YAs across all sensory trial types, they exhibited greater differences in RTs between AV and V trials (i.e., ↑ AV-V RT difference), with no significant effects of choice accuracy, implicating preserved benefits of multisensory integration towards perceptual decision formation. HDDM demonstrated parsimonious fittings for characterising these behavioural discrepancies between YAs and OAs. Notably we found slower rates of sensory evidence accumulation (i.e., ↓ drift rates) for OAs across all sensory trial types, coupled with (1) higher rates of sensory evidence accumulation (i.e., ↑ drift rates) for

OAs between AV versus V trial types irrespective of stimulus difficulty, coupled with (2) increased response caution (i.e., ↑ decision boundaries) between AV versus V trial types, and (3) decreased non-decisional processing duration (i.e., ↓ non-decision times) between AV versus V trial types for stimuli of increased difficulty respectively. Our findings suggest that older adults trade-off multisensory decision-making speed for accuracy to preserve enhancements towards perceptual decision formation relative to younger adults. Hence, they display an increased reliance on integrating multimodal information; through the principle of inverse effectiveness, as a compensatory mechanism for a generalised cognitive slowing when processing unisensory information. Overall, our findings demonstrate how computational modelling can reconcile contrasting hypotheses of age-related changes in processes underlying multisensory perceptual decision-making behaviour.

Finally, Chapter 4 outlines research probing the exerted influence of perceptual learning on multisensory perceptual decision-making. Views of unisensory perceptual learning imply that improvements in perceptual sensitivity may be due to enhancements in early sensory representations and/or modulations to post-sensory decision dynamics. We sought to assess whether these views could account for improvements in perceptual sensitivity for multisensory stimuli, or even exacerbations of multisensory enhancements towards decision formation, by consolidating the spatiotemporal locus of where and when in the brain they may be observed. We recorded EEG activity from participants who completed the same audiovisual object categorisation paradigm (as outlined in Chapter 3), over three consecutive days. We used single-trial multivariate LDA to characterise the spatiotemporal trajectory of the decision dynamics underlying any observed multisensory benefits both (a) within and (b) between visual, auditory, and audiovisual trial types. While found significant decreases were found in RTs and increases in choice accuracy over testing days, we did not find any significant effects of perceptual learning on multisensory nor unisensory perceptual decision formation. Similarly, EEG analysis did not find any neural components indicative of early or late modulatory effects from perceptual learning in brain activity, which we attribute to (1) a long duration of stimulus presentations (300ms), and (2) a lack of sufficient statistical power for our LDA classifier to discriminate face-versus-car trial types. We end this chapter with considerations for discerning multisensory benefits towards perceptual decision formation, and recommendations for altering our experimental design to observe the effects of perceptual learning as a decision neuromodulator.

These findings contribute to literature justifying the increasing relevance of utilising behaviourally-informed and/or neurally-informed modelling approaches for investigating multisensory perceptual decision-making. In particular, a discussion of the underlying cognitive and/or neural mechanisms that can be attributed to the benefits of multisensory integration towards perceptual decision formation, as well as the modulatory impact of the decision modulators in question, can contribute to a theoretical reconciliation that multisensory integrative benefits are not ubiquitous to specific spatiotemporal neural dynamics nor cognitive processes.

Table of Contents

List of Abbreviations

2AFC	Two-Alternative Forced-Choice Paradigms
A	Auditory (trials)
ACC	Anterior Cingulate Cortex
AV	Audiovisual (trials)
AVE	Ventriloquist After-Effect
CPP	Centro-Parietal Positivity
DDM	Drift Diffusion Model / Drift Diffusion Modelling
DIC	Deviance Information Criterion
DLPFC	Dorsolateral Pre-Frontal Cortex
EEG	Electroencephalography
ERP	Event-Related Potential
FEF	Frontal Eye Fields
fMRI	Functional Magnetic Resonance Imaging
HC	High Coherence (trials)
HDI	Highest Density Interval
HDDM	Hierarchical Drift Diffusion Model / Hierarchical Drift Diffusion Modelling
IFG	Inferior Frontal Gyri
IFS	Inferior Frontal Sulcus
IPS	Intraparietal Sulcus
LC	Low Coherence (trials)
LBA	Linear Ballistic Accumulator
LDA	Linear Discriminant Analysis
LHB	Left-Hemisphere Beta ERP
LIP	Lateral Intraparietal Area
LOC	Lateral Occipital Complex
MCMC	Markov Chain Monte Carlo
MEG	Magnetoencephalography

MFG	Middle Frontal Gyrus
MMN	Mismatch Negativity ERP
MT	Medial Temporal
OA	Older Adults (trials)
P(…)	Probability (Bayesian Hypothesis Testing)
PPI	Psychophysical Interaction
PPC	Posterior Predictive Check(s)
RDK	Random Dot Kinematogram
ROI	Region of Interest
RROB	Reward Rate Optimal Boundary
rTMS	Transcranial Magnetic Stimulation
SATO	Speed-Accuracy Trade-Off
SC	Superior Colliculus
SFS	Superior Frontal Sulcus
SMA	Supplementary Motor Area
SNR	Signal-to-Noise Ratio
SSVEP	Steady State Visual-Evoked Potential
STS	Superior Temporal Sulcus
TBW	Temporal Binding Window
TWIN	Time-Window-of-Integration Model
V	Visual (trials)
VE	Ventriloquist Effect
YA	Younger Adults (trials)

List of Tables

List of Figures

Chapter 1
General Introduction

In everyday life, we frequently encounter situations whereby a multitude of decisions must be formed on the basis of ambiguous and inherently noisy sensory information. Each decision requires a consolidation of incoming sensory information, or sensory-motor information, as determinant evidence towards particular choice alternatives. Once sufficient evidence has been accumulated towards a specific choice alternative, an overt behavioural response is facilitated, which prepares motor responses for efficient interaction with our immediate surrounding environment. The importance of the role of sensory information in directing behavioural responses has led this process to be defined as *perceptual decision-making* (see Gold & Shadlen, 2007; Heekeren, Marrett, & Ungerleider, 2008; Philiastides & Heekeren, 2009; Philiastides, Diaz, & Gherman, 2017, for detailed reviews). In contrast with broader definitions of decision-making, perceptual decision-making considers choices to be facilitated at a non-deliberate lower level, which operates independently from a higher level of conscious awareness. Specifically, the criteria defining the quality of decisions is dependent on immediate and readily available sensory information, and less dependent on the subjective preference and goals of the observer (Dutilh & Rieskamp, 2016; Hauser & Salinas, 2014). In addition, prior-informed expectations, derived from repeated exposure, can bias perceptual decisions formed under time pressure, whereby we prioritise choice alternatives that are probabilistically more likely to be correct (Kelly, Corbett, & O’Connell, 2021). With a higher level of conscious awareness, as well as a lack of time pressure, this bias can be reduced.

Consequently, there has been a longstanding interest in cognitive neuroscientific research to investigate the underlying cognitive and neural mechanisms of perceptual decision-making within the human brain. In order to investigate how we formulate perceptual decisions from readily available sensory information, researchers typically use speeded classification paradigms, whereby a participant must rapidly categorise a particular level of sensory stimulus input (which may vary in coherence, or “noise”), and make a decision constrained to a limited number of choice alternatives, for example, two choices in 2AFC paradigms (Smith & Ratcliff, 2004). Decisional behaviours can then be assessed with single-trial measurements such as RTs and choice accuracy (and/or choice probability), which reflect the speed and accuracy of perceptual decision-making respectively. As such, differences in behavioural performance across the formation of perceptual decisions can be quantified in order to assess modulations in decisional behaviour from modulators, for example, changes in sensory input strength (Palmer, Hulk, & Shadlen, 2005) or the task demands of the paradigm itself (Heekeren et al., 2004; Otto, Dassy, & Mamassian, 2013). A neuroimaging methodology such as EEG, fMRI, or magnetoencephalography (MEG) can be simultaneously applied to measure brain activity that underlies the formation of perceptual decisions on a single-trial basis. Such dual methodologies permit for an investigation of the spatiotemporal trajectory of the overall process of perceptual decision-making, whereby its neural correlates can be pinpointed within an interconnected hierarchy of cortical regions within the brain.

The collected behavioural and neuroimaging datasets can be implemented, either independently or simultaneously, into computational models, to complete what are termed as *behaviourally-informed modelling approaches* (modelling behavioural datasets) or *neurally-informed modelling approaches* (modelling neuroimaging and behavioural datasets) respectively (see Hawkins et al., 2017; Turner et al., 2013; 2017, for reviews) for investigating the internal constituent mechanisms; both cognitive and neural, of perceptual decision formation. In particular, the Drift Diffusion Model (DDM; see Forstmann, Ratcliff, & Wagenmakers, 2016; Ratcliff, 1978; Ratcliff et al., 2016; Ratcliff & McKoon, 2008; Ratcliff & Smith, 2004; Ratcliff, Smith, & McKoon, 2015; for detailed reviews), has proven invaluable in its modelling of RTs and choice accuracy. Furthermore, simultaneously collected neuroimaging datasets can be used to inform parameter fittings to predict latent states of perceptual decision-making patterns. Through the application of such approaches, mechanistic insights into the functions that predict decisional behaviours can be uncovered. In particular, a framework has been conceptualised implicating the overall process of perceptual decision-making involves a non-serial cascade of separate cognitive stages, or events, each of which is represented in an interconnected hierarchy of cortical regions within the brain.

However, our external environment is inherently multisensory. The available sensory information must be seamlessly integrated and consolidated across multiple sensory modalities, often near-simultaneously, in order to form unique and coherent perceptual representations, or percepts, delineating salient information about the environmental objects that inform choice behaviours (see Alais et al., 2010; Ghazanfar & Schroeder, 2006, for detailed reviews). The bound multisensory percepts subsequently guide adaptive behaviours. When compared to unisensory percepts (formed from information presented to single sensory modalities), perceptual decision formation from multisensory information shows benefits to decisional responses, notably improved stimulus detection (Lewis & Noppeney, 2010), increased response speed (i.e. reduced RTs in speeded classification paradigms; Lippert, Logothetis, & Kayser, 2007), and increased response choice accuracy (Franzen et al., 2020; Kayser, Philiastides, & Kayser, 2017).

As such, the framework of perceptual decision-making has been extended to consider the integration and consolidation of sensory information across multiple modalities, and recently has been referred to as *multisensory (perceptual) decision-making*[1] (Bizley, Jones, & Town, 2016a; Bizley, Maddox, & Lee, 2016b; Drugowitsch et al., 2014; Franzen et al., 2020; Raposo et al., 2012). The employment of both behaviourally-informed and neurally-informed modelling approaches in investigations of the consolidation of multisensory information for perceptual decision formation, through comparisons of bimodal (presented in close spatial and/or temporal proximity, e.g., auditory and visual stimuli) and unimodal stimulus presentations have demonstrated neural and cognitive underpinnings discerning the benefits we see with multisensory information, compared to unisensory information, towards perceptual decision formation. Thus, this implicates that the integration and consolidation of multisensory information increases the likelihood of optimised decisional responses (Chandrasekaran, 2017; Drugowitsch et al., 2014; 2015; Hou et al., 2018). Thus, in order to benefit a wider understanding of perceptual decision-making, the proposed framework needs to consider the distinct integrative mechanisms; both cognitive and neural, that underlies an increased likelihood of optimised decisional responses through consolidating multisensory representations to benefit

---

[1] Where necessary, this thesis will refer to multisensory decision-making as “multisensory perceptual decision-making” in order to distinguish the process from perceptual decision-making alone, which focuses on findings from experimental paradigms that presented only unisensory stimuli.

choice formation. These include cognitive mechanisms for forming multisensory representations; for example, binding unisensory signals weighted on sensory reliability and/or task relevance (Rohe & Noppeney, 2016), and consolidating the causal relationship between bound signals to expedite the accumulation of sensory evidence (Zangaladze et al., 1999), as well as the neural mechanisms for differentiating multisensory representations in sensory and higher associative cortices. In particular, sensory cortices have been prominently linked to encoding sensory information into decisional evidence (e.g., auditory; Bizley et al., 2013; somatosensory; Romo, Lemus, & Lafuente, 2012), whereas regions such as the Inferior Frontal Sulcus (IFS; Werner & Noppeney, 2010), Posterior Parietal Cortex (PPC), and prefrontal cortex (Erlich et al., 2015) being found to reflect the accumulation of decisional evidence benefitting choice formation.

Despite an increased understanding of these cognitive and neural mechanisms underlying perceptual decision-making with unisensory stimuli, our understanding of the comparative similarities and differences with the cognitive and neural mechanisms underlying multisensory perceptual decision-making requires further exploration. Cognitive neuroscientific research is only beginning to yield insights into how enhancements from multisensory stimuli in perceptual decision formation emerge, and subsequently modulate, the cascade of mechanisms resulting from an integration of sensory information across multiple modalities. This profoundly impacts a wider understanding of perceptual decision-making. In addition, this further impacts our comprehension of the exerted influence of *decision (neuro)modulators* on the separate events of decisional processing within the established framework. Consequently, insights into the prominent cognitive and neural mechanisms for multisensory perceptual decision-making in ecological surroundings, where sensory information is inherently probabilistic and highly variable (i.e. “noisy”; Philiastides, Diaz, & Gherman, 2017), is limited. As such, the previously defined framework needs to continue considering enhancements arising within multisensory perceptual decision-making.

This introductory chapter attempts to summarise recent literature exploring multisensory perceptual decision-making. First, it will outline the current understanding of perceptual decision-making with unisensory stimuli. Second, it will outline the current understanding of the mechanisms of multisensory integration, and the ongoing debate of where multisensory signals are combined to benefit the formation of perceptual decisions. Finally, it will outline the outstanding questions that motivate this thesis, and how the decision (neuro)modulators in question affect the determinant cognitive and neural mechanisms of multisensory perceptual decision-making.

1.1 Unisensory Perceptual Decision-Making

The following section will describe converging non-human neurophysiological, human neuroimaging, and theoretical modelling research that demonstrates empirical support for the various stages of processing for unisensory perceptual decision-making. It will also outline the predictions made from behaviourally-informed and neurally-informed modelling approaches that inform the current framework used to conceptualise unisensory perceptual decision-making.

1.1.1 Non-Human Neurophysiological Research

Earlier non-human electrophysiological findings first revealed patterns of single-neuron activity that suggested an integrative mechanism for translating sensory information into overt motor behaviours (Gold & Shadlen, 2007). Previous findings recorded single-unit patterns that pointed towards modulations in neural activity that were not attributed to the commitment of choice behaviours, and instead corresponded to a cascade of neural events underlying the overall process of perceptual decision formation. For example, Newsome, Britten, and Movshon (1989) recorded single-neuron activity from MT area V5 in the extrastriate visual cortex of Rhesus monkeys, which spanned three different psychophysical thresholds of a Random Dot Kinematogram (RDK). This paradigm required the monkeys to categorise the dominant direction of motion of a random dot display by sustaining a visual fixation on their preferred direction. They observed that the response profiles of neurons recorded displayed increased sensitivity and reliability in the measurement of visual motion coherence when compared to the overt visual fixations the monkeys made themselves. Moreover, further studies found that the perceptual choice formation of monkeys correlated with the single-trial variability of direction-selective MT/V5 neurons, whereby the neurons displayed increased firing that correlated positively with the likelihood of the neuron’s preferred direction of motion (Britten et al., 1992; 1993; Shadlen et al., 1996). These results implicated that perceptual decisions are formed from the increased firing of cortical neurons prior to the behavioural response itself, and therefore that sensory evidence representation forms the early basis for decision formation, but is accumulated with increased activity within the brain.

Further research in this field also demonstrated that the sensory evidence provided by sensory cortical regions of the brain can be used to drive the later stages of perceptual decision formation in non-sensory, higher-order, regions. For example, it has been found that single-neuron activity in the lateral intraparietal (LIP) area, within the parietal cortex, of Rhesus and Macaque monkeys gradually increased over time at a rate proportional to the difficulty of a visual direction-discrimination RDK. This was thought to implicate an accumulation of consolidated sensory evidence needed to facilitate a decision based on task difficulty. In addition, firing rates were reported to have gradually converged when the monkeys committed to their visual fixation choices (Huk & Shadlen, 2005; Roitman & Shadlen, 2002; Shadlen & Newsome, 2001). The increased neuronal firing, corresponding to the availability of incoming visual information, and hence the subsequent build-up of decisional evidence, was also reported in the frontal eye fields (FEFs) and dorsolateral pre-frontal (DLPFC) cortex (Kim & Shadlen, 1999). These higher-order cortical regions are independent of lower-order sensory cortical regions, which receive incoming sensory information first. Therefore, these findings can be regarded as evidence that analogue representations of visual motion were not received in higher-order cortical regions. Instead, they were consolidated, accumulated, and translated into a decision variable which informed choice behaviours (i.e., visual fixations). Such an interpretation can be reaffirmed by further observations that the predictive ability of firing rates, for the time and direction of visual saccades, were not observed once primate participants were committed to their behavioural movement.

In summary, earlier non-human neurophysiological findings provided seminal insights into the neural mechanisms underlying perceptual decision-making, albeit at a single-neuron level. Undoubtedly, they outline distinct processes of perceptual decision formation that interacted in a hierarchical manner of activation. First, there is the process of sensory evidence representation in sensory cortices, representative of the consolidation of incoming signal information. Second, there is the corresponding downstream process from sensory cortices to associative cortices, whereby decisional evidence is accumulated to form choice responses through the comparison of sensory neural firing inputs. Importantly, the identification of a downstream process for accumulating sensory information matches computational frameworks of perceptual decision, which implicate an integrative mechanism for accumulating sensory evidence over time to a pre-set internal decision boundary, thus determining choice responses (Philiastides, Diaz, & Gherman, 2017; Siegel, Engel, & Donner, 2011). These will be outlined in the following subsection.

1.1.2 Human Neuroimaging Research

The *de facto* preference of speeded classification paradigms, often utilised when investigating perceptual decision formation in humans, is typically the two-alternative forced-choice (2AFC) paradigm. In this paradigm, a participant observes a sensory stimulus input, and is constrained to make a binary perceptual judgement. The sensitivity level, or stimulus coherence, of the input can be manipulated to observe performance differences in decision-making behaviour on a single-trial basis. The two commonly collected behavioural measurements of performance consist of RTs; usually recorded on the scale of milliseconds, and choice accuracy; usually recorded as a binary variable of whether or not the participant was correct in their facilitated choice. Using these two dependent variable measurements, how the formation of perceptual decisions changes between experimental conditions can be quantified. A methodology of neuroimaging such as EEG, fMRI, or MEG, can then be simultaneously applied to measure brain activity underlying the formation of perceptual decisions. Uncovering brain activity that reflects key neural events underpinning perceptual decision formation, quantified through analyses of RTs and choice accuracy, opens new avenues in regard to empirically validating the neural mechanisms underlying the distinct cognitive events, whereby predictions can be made for gaining a functional insight into the translation of sensory information into choice behaviours.

In particular, key computations can be predicted utilising *behaviourally-informed* or *neurally-informed modelling approaches* (see Hawkins et al., 2017; O’Connell et al., 2018; Turner et al., 2013; 2017, for detailed reviews). In such approaches, individually or simultaneously collected behavioural and neuroimaging datasets can be joined in a single framework for empirically testing and refining model predictions about the constituent cognitive and neural stages of perceptual decision formation respectively. One prominent variant of such model-based cognitive neuroscience approaches concerns *sequential sampling models*, in particular, the Drift Diffusion Model (DDM; Forstmann, Ratcliff, & Wagenmakers, 2016; Ratcliff et al., 2016; Ratcliff & McKoon, 2008; Ratcliff & Smith, 2004; Ratcliff, Smith, & McKoon, 2015; Wiecki et al., 2013). These models are behaviourally-informed, assuming a process of stochastic sensory evidence accumulation during decision formation, captured by certain parameters estimating different components of the decision-making process. Moreover, this process involves an integrative mechanism, concurrent with the mechanism processed for accumulation of representative sensory information from earlier animal neurophysiological findings, in which once enough sensory evidence is accumulated over time to exceed an internal boundary for one choice alternative, that choice is initiated, facilitating a behavioural response. They model single-trial speed (i.e., RTs) and choice accuracy, inextricably coupling speed and choice accuracy, and not treating them as independent response variables. Conceptualising perceptual decision-making as a gradual evidence accumulation process is advantageous in understanding its own internal variability in choice responses, and not quantifying the relation between RTs and accuracy as fixed. Instead, predictions can be accounted for of the effects of internal and external decisional modulators with the two behavioural performance measurements. For example, arbitrarily high choice accuracy can be accounted for by increases in response speed, and arbitrarily low (i.e., fast) response speed can be accounted for by decreases in choice accuracy, thus functionally probing the speed-accuracy trade-off (SATOs) when forming perceptual decisions (Wenzlaff et al., 2011; Zhang & Rowe, 2014). In DDMs, estimated parameters of internal cognitive components modulated by the experimental context are decomposed as: (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*). In addition, a further advantage of utilising DDMs as a sequential sampling modelling approach are that trial-by-trial correlations between a measurement of brain function (e.g. EEG amplitude, fMRI BOLD activity), and the aforementioned DDM parameters, can be estimated to demonstrate an association between brain activity and internal cognitive components under investigation, thus constraining the model fits further to inform the role of brain function for certain latent processes.

Consequently, the application of behaviourally-informed and neurally-informed modelling approaches for linking single-trial behavioural measurements (i.e., RTs and/or choice accuracy) with corresponding recorded neural activity in the human brain has uncovered an interconnected hierarchy of cortical areas that underlie a cascade of separate stages, or events, of perceptual decision formation (Heekeren et al., 2004; 2006). Specifically, findings from the application of such approaches have coupled distinct regions in the human brain that underlie separate constituent components for the internal process of stochastic sensory evidence accumulation that predicts decision-making performance. For instance, Domenech and Dreher (2010) applied model-based fMRI to investigate perceptual decision formation in a go/no-go task, whereby participants had to press a response key for specific shapes in a sequence of visual stimuli. Their results highlighted voxels in the anterior cingulate cortex (ACC) that proportionally accounted for the slope of sensory evidence needed to be accumulated to initiate the motor response (i.e., response key presses) when the correct shape arose in the sequence. In addition, they highlighted a role for the dorsolateral prefrontal cortex (DLPFC) for the coding of accumulated sensory evidence. They proposed that the flow of sensory information from the ACC to the DLPFC was optimally adjusted to account for the modulations in decision thresholds, particularly for computations that required increased regulation to inhibit incorrect motor responses.

Comparably, Forstmann et al. (2008) utilised fMRI and an experimental paradigm that instructed participants to judge whether the overall motion of a cloud of moving dots moved to the left or the right. Instructions in one block were manipulated whereby on one condition participants were asked to prioritise decisional speed, whereas separate instructions in an alternate block instructed participants to prioritise decisional accuracy. Statistical analysis of behavioural data using a Linear Ballistic Accumulator (LBA) model; a sequential sampling model similar to the DDM, assessing accumulation to response thresholds, discerned an account of changes in the response threshold for facilitating decisions based on the prioritisation of speed or accuracy (i.e., a lowering of speed for prioritising accuracy and vice versa). Their fMRI analysis showed that activity in the striatum and pre-supplementary motor area (pre-SMA), brain regions associated with voluntary motor planning (Shima & Tanji, 1998), increased when block instructions prioritised a cueing for decisional speed. Their fMRI findings suggested that individual variations in striatum and pre-SMA activity correlated to the accumulation of visual sensory evidence evaluated by their modelling approach. They conducted a follow-up study applying the same methodology, instead incorporating within-block variations in instructions to emphasise decisional speed or accuracy. Their fMRI analysis highlighted that connectivity strength between the pre-SMA and striatum was predictive of the efficiency of which participants could modulate their response threshold based on SATO prioritisation (Forstmann et al., 2010). Yau et al. (2020) further characterised the neural underpinnings of evidence accumulation rate and response threshold modulations, utilising fMRI and a hierarchical extension of the DDM (Hierarchical Drift Diffusion Model(ling); HDDM; Wiecki et al., 2013), to examine the dynamic perception of emotional facial features. They decoded multivariate fMRI signals underlying subsequent perceptual decision formation to discern how they predict changes to HDDM parameters. They identified trial-by-trial variations in patterns of neural activity, localised in the fusiform gyrus, which corresponded with a consolidation of facial emotional information. Furthermore, the variability in these patterns was accounted for by modulations in drift rate, hence characterising the process of evidence accumulation in discerning the emotion of facial features within the fusiform gyrus. Moreover, a separate subcortical structure, the caudate nucleus, exhibited greater paternal activity variability corresponding to an inverse-urgency signal. In the HDDM, this was modelled with decision boundary increases when choice formation was slower, suggesting that in contexts where incoming sensory information remains ambiguous to consolidate, an increase in response caution compensates to ensure appropriate choice facilitation.

While fMRI has excellent spatial resolution localising where the constituent processes of perceptual decision-making may occur in the human brain, it has poor temporal resolution. This makes it difficult to measure the time-course and temporal characteristics of perceptual decision-making processes. To this end, the application of neuroimaging methodologies such as MEG and EEG have been used in neurally-informed modelling approaches to investigate the temporal trajectory of the internal processes of perceptual decision formation. Their advantage over spatial methodologies; for example fMRI, is their excellent temporal resolution, typically on the order of milliseconds, to dissociate when the decisional processes and their constituent neural modulations arise. Studies applying EEG-based modelling approaches, in particular, have consolidated distinct neural markers, and corresponding time points, for sensory information consolidation and an indexing of accumulation-to-bound sensory evidence accumulation. These studies further empirically validate a hierarchical framework for the processes within perceptual decision formation.

In particular, seminal research from Philiastides et al. (see Philiastides & Sajda, 2006a; 2006b; Philiastides, Ratcliff, & Sajda, 2006) used a novel 2AFC paradigm; the face-versus-car categorisation paradigm, to research non-invasive neural modulations that correlate to perceptual decision-making processes in the visual modality. In this paradigm, single-trial images of faces or cars are rapidly presented. Participants are required to initiate a choice, as quickly and as accurately as possible, categorising the embedded stimulus. Their neurally-informed modelling approach aimed to find a basis vector whereby EEG signals are projected maximally discriminating neural patterns in trials of one stimulus type compared to the other, using multivariate Linear Discriminant Analysis (LDA; Parra et al., 2002; 2005; 2007; Sajda et al., 2009). They highlighted two spatiotemporal EEG components, at two distinct time points, that best discriminated neural activity between the two stimulus categories (Philastides et al., 2006a; 2006b). First, they highlighted an earlier major discriminating EEG component at a post-stimulus onset of 130-170ms (which they termed the *Early* component). This was consistent with the characteristics of the face-selective N170 Event-Related Potential (ERP; Elmer, 2011), demonstrating evoked activity for processing stimulus features. The second *Late* component, however, arose variably from 300-450ms post-stimulus onset. Interestingly, their choice probability analysis indicated a strong correlation between the profile of the Late component and participant's judgements, whereby when evidence for face stimuli increased and evidence for car stimuli decreased, the onset of the Late component shifted earlier in time, correlating with reported lower RTs and increased choice accuracy. Consequently, these findings demonstrated an evolution in the temporal trajectory of the perceptual decision formation process within early visual perception. Specifically, evidence accumulation benefitted from increased strength of incoming sensory information (i.e. increased image coherence), modulating the underlying neural components for accumulation-to-bound evidence gathering to benefit choice formation. Projected scalp topographies reaffirmed this interpretation of the temporal trajectory, in which the localized discriminatory EEG activity occurred over occipital-temporal brain regions, but shifted to frontocentral brain regions prior to choice facilitation, validating neuroimaging findings suggesting structures such as the inferior occipital and fusiform gyri have functional roles in encoding incoming visual information (McCarthy et al., 1997; Starrfelt & Gerlach, 2007), as well as the DLPFC having a functional role with initiating actions (Heekeren et al., 2004; Philiastides et al., 2011).

They further investigated the temporal shifts in the Late component from variations in task difficulty (Philiastides, Ratcliff, & Sajda, 2006). They used a cued variant of the face-versus-car categorisation paradigm, instructing participants in certain trials to categorise image colour (as either red or green), changing the difficulty of categorisations (i.e., from categorizing stimulus type). Similar to their previous findings, they reported post-stimulus onsets of Early and Late EEG components at approximately 170ms and 300ms respectively. However, when instructions changed to red-green colour categorisations, the magnitude of the Late component, but not Early component, reduced to almost zero. This implicated a change in decision evidence characterization, with the authors suggesting a recruitment of top-down attentional resources in accordance with a change in task instructions. Furthermore, DDM application highlighted that the Late component was a significantly better predictor of single-trial modulations in drift rate, quantifying the rate of sensory evidence accumulation, motioning that an integrative mechanism determined choice response facilitation in the human brain according to paradigm demands (Philiastides, Ratcliff, & Sajda, 2006; 2009).

Fundamentally, these preliminary findings implicated different neural mechanisms, captured with temporally distinct components, which were representative of distinct cognitive events during perceptual decision formation. Pointedly, earlier neural mechanisms appeared to reflect the events of incoming sensory information encoding and translation into decisional evidence, whereas later neural mechanisms appeared to reflect the events of decisional evidence accumulation. The events appeared to be serially independent and operated downstream into regions of the brain underlying the formation of decisions (Philiastides & Heekeren, 2009). The serial cascade of these events predicted the onset at which information is first represented in sensory cortices, depending on the activated sensory domains (e.g. auditory; Kaiser, Lennert, & Lutzenberger, 2007; somatosensory; Spitzer & Blankenburg, 2011). Then, this information is translated in inferior frontal, temporal, and parietal associative cortices, which operate to form decisions from accumulation-to-bound evidence accumulation (Ploran et al., 2007; Tremel & Wheeler, 2015; Wheeler et al., 2015). Further findings empirically validate the notion of separate processing markers for perceptual decision formation. For example, Philiastides et al. (2011) used repetitive Transcranial Magnetic Stimulation (rTMS) on regions of the DLPFC of participants engaged in the face-versus car categorisation task. They found that low frequency rTMS over the left DLPFC reduced decision speed and accuracy, with the application of a DDM corresponding to decreases in drift rate, implicating a reduced rate of sensory evidence accumulation. These results further suggested a causal role of the DLPFC in accumulating incoming visual information to facilitate choice responses, with similar modelling findings with a stimulus-locked EEG-DDM analysis on population responses across the frontal scalp (Philiastides, Heekeren, & Sajda, 2014). Comparatively, O’Connell, Dockree, & Kelly (2012) investigated Steady State Visual-Evoked Potential (SSVEP), Left-Hemisphere Beta (LHB), and Central Parietal Positivity (CPP) ERPs using a gradual target detection paradigm. They found steady increases in the CPP ERP underlying decisions made about a stimulus fading, while SSVEP and LHB activity gradually faded. This captured the characteristics of accumulation-to-bound decision formation, independent of sensory-specific representations of flicking visual evidence; captured by gradual changes in the SSVEP potential, and motor-response production latency; captured by gradual changes in the LHB potential. In addition, Filimon et al. (2013) combined event-related fMRI, effective connectivity analysis, and Psychophysical Interaction (PPI) analysis to investigate whether the human sensorimotor system is embodied in perceptual decision implementation. Using cues for response execution varying between eye and hand movements in a face-versus-house categorisation paradigm, they found greater effectivity connectivity from sensory regions to the left IFS, corresponding to amounts of sensory evidence, but greater task-positive BOLD responses in the lateral intraparietal (LIP) area for motor response execution when cues instructed for categorisations with hand movements. This implicated that eye and hand motor response preparation was disentangled from the formation of perceptual decisions. PPI analysis modelled increased connectivity from the Lateral Occipital Complex (LOC) within the occipitotemporal brain area to the left IFS, anterior Inferior Frontal Gyri (IFG), posterior Superior Frontal Sulcus

(SFS), and middle frontal gyrus (MFG) in the frontal lobe regions, which predictably varied with the coherence of visual stimuli. This reaffirms interpretations of different systems for perceptual accumulation of sensory evidence and production of motor responses and underlying neural mechanisms for distinct cognitive and motor events in the facilitation of behavioural responses.

In conclusion, the application of behaviourally-informed and neurally-informed modelling approaches in unisensory perceptual decision-making research has consolidated the underlying neural mechanisms, and their functional role in decision-making cognition, within the formation of perceptual decisions. Thus, they established a framework for perceptual decision formation which conceptualised a spatiotemporal locus of sensory information representation and integration into decision evidence, with task manipulations varying choice formation (i.e. RTs and choice accuracy), hence characterising inherent cognitive processes that operate to facilitate appropriate behavioural responses.

1.2 Multisensory Perceptual Decision-Making

The perceptual decision-making framework highlighted in the previous section is limited to an extent. Undoubtedly, it is based on research that investigated unisensory stimuli only. Our external environment is inherently multisensory. Everyday perceptual decisions frequently involve situations whereby incoming information is presented to multiple sensory modalities (see Alais et al., 2010; Ghazanfar & Schroeder, 2006, for detailed reviews). This influx of sensory signals occurs often near-simultaneously. Thus, choice responses must be facilitated that consider the consolidation of sensory information presented to more than one sensory modality. Integration of sensory modality-specific signals is an inherent process in combining information into unified percepts, which not only form unique coherent representations of our external environment, but optimise choice response facilitation when compared to the consolidation of unisensory information alone (Drugowitsch et al., 2014; 2015). Consequently, this has led researchers to define *multisensory perceptual decision-making* (Bizley, Jones, & Town, 2016a; Bizley, Maddox, & Lee, 2016b; Drugowitsch et al., 2014; Franzen et al., 2020; Raposo et al., 2012).

The following section will introduce research highlighting how multisensory integration is an important mechanism to consider within perceptual decision formation. It will outline human neuroimaging and theoretical modelling multisensory research that empirically validates the framework constructed from unisensory perceptual decision-making research, but builds upon it in consideration of the cognitive and neural mechanisms underlying the integration of multisensory information. It will further consider how a multisensory perceptual decision-making framework accounts for perceptual enhancements towards behavioural decision formation. Finally, it will introduce the outstanding question in the field, namely *where* we may observe enhancements towards perceptual decision formation from multisensory integration of unisensory stimuli.

1.2.1 Multisensory Integration

Incoming sensory input from single modalities interacts, converges upon, and influences further multisensory processes (Stein & Meredith, 1993). This convergence of multisensory information forms single integrated experiences from which relevant environmental percepts are constructed to inform decisional behaviours, and is commonly referred to as *multisensory integration* (Calvert et al., 2004; Stein & Stanford, 2008; Spence & Squire, 2003). A classical viewpoint posited that each sensory input was processed independently prior to multisensory integration, and the multisensory convergence of afferent inputs occurred in higher-order associative cortices and specialised subcortical structures (Treisman & Gelade, 1980). However, this “unisensory before multisensory” perspective (Alais et al., 2010) has become less influential overtime (Schroeder & Foxe, 2005).

The lessening influence of the “unisensory before multisensory” viewpoint occurred primarily due to two significant findings across earlier research. First, low-level brain regions, assumed to be unisensory, were found to respond to different combinations of multisensory stimuli (See Murray et al., 2016; Kayser et al., 2005, for reviews on neuroimaging evidence that commonly reports early audiovisual and auditory-somatosensory integrative effects within the primary visual and auditory cortices respectively). Foxe and colleagues, for example, reported that simultaneous auditory-somatosensory stimulations manifested interaction effects approximately 50ms post-stimulus onset over right postcentral gyrus (localised to the side of auditory-somatosensory stimulation over the respective cortices) using ERPs (Foxe et al., 2000). Similar findings were obtained using fMRI over the same brain regions that corresponded to detection of simultaneous auditory-somatosensory stimulations (Foxe et al., 2002). Similarly, Cappe (2010) identified ERPs 60-95ms post-stimulus onset, which revealed a topographic organisation of auditory-visual stimuli integrative effects for primary auditory and visual cortical areas, as well as preliminary posterior superior temporal regions mediating these effects. These findings provided evidence for the dissociation of multisensory integration as a process from separate bottom-up perceptual processes.

Second, pioneering neurophysiological research by Stein and colleagues led to identifying the Superior Colliculus (SC) as a subcortical structure inherent in the integration of multisensory information (Meredith, 2002; Calvert et al., 2001; Stein & Meredith, 1993). This midbrain structure both topographically maps retinotopic coordinates obtained from visual modality input (Sprague & Meikle, 1965), and is linked to the proprioception of head and eye orientations (Sparks, 1986). In investigating its role in multisensory integration, Meredith and Stein (1983) evoked dynamic modulations in neural activity from deep laminae SC neurons in cats using different simultaneous combinations of visual, auditory, and somatosensory inputs. Further studies implicated that the spatial overlap of SC neuron Receptive Field (RF) alignments was predictive of the multisensory stimulus-dependent combinations and was essential for converging afferent multisensory inputs (Kadunce et al., 2001; Meredith & Stein, 1996; Ursino et al., 2014; Wallace et al., 1998). This could be through either multisensory enhancement of cross-modal inputs in close spatial RF proximity (Stanford et al., 2005; Perrault et al., 2005), multisensory depression of inputs not spatially proximate, or inhibited by greater cross-modal neural responses (Jiang & Stein, 2003; Kadunce et al., 1997). Without behavioural measurements, these studies support multisensory integration as functionally distinct from further multisensory processes. Instead, modulations of neural responses from multisensory integration operate independently of lower-order perceptual or higher-order cognitive awareness (Alsiuis & Munhall, 2013; Jones, 2015), are innately automatic as a mechanism (Onat et al., 2007; Bresciani et al., 2006), and occur at early post-stimulus latency onsets; often less than 100ms and as early as ~50ms (Molholm et al., 2002; Foxe et al., 2000; 2002; Giard & Peronnet, 1990). Despite the lessening influence of the "unisensory before multisensory" perspective, it still highlights a hierarchical organisation in the brain for consolidation of multisensory information towards perceptual choice formation (Ursino et al., 2014).

Accordingly, four key principles have been governed outlining the innate benefits of multisensory integration towards perceptual decision formation (see Angelaki, Gu, & DeAngelis, 2009; Freiherr et al., 2013; Stevenson et al., 2014; for detailed reviews). The first two principles are referred to as the *spatial* and *temporal* rules respectively (Holmes & Spence, 2005; King & Palmer, 1985; Meredith & Stein, 1986; Spence, 2013). These principles argue that unisensory information presented in close spatial proximity and/or temporal concordance results in an increased likelihood of multisensory integration, and therefore an increased likelihood of multisensory enhancements towards decision formation. The third principle constitutes contextual *semantic congruency* between unisensory informational features strengthens the likelihood of multisensory integration. Here, unisensory informational streams that correspond to the same environmental object in meaning underlies the formation of percepts benefitting perceptual choice formation (Spence, 2007; 2011; Stevenson, Wallace, & Altieri, 2014). The fourth principle is termed the *Principle of Inverse Effectiveness* (Holmes, 2009; Meredith & Stein, 1986a; 1986b; Stein & Stafford, 2008). The Principle of Inverse Effectiveness refers to increased likelihood of multisensory integration when responsiveness to individual unisensory stimuli decreases. Hence, when less intense (in their stimulations) unisensory stimuli are processed separately, their consistent summation is exceeded by evoked responses from simultaneous stimuli processing. This principle is supported by observations that even weak multisensory combinations of unimodal stimuli have been found to enhance behavioural responses compared to strong, salient, unimodal stimulus presentations alone (Cappe et al., 2010).

The aforementioned principles, reinforced by single-unit findings integration in the SC, imply that multisensory integration is supramodal within the brain. Furthermore, it is pivotal for the processing of incoming sensory information, presented to multiple sensory modalities, to optimise appropriate behavioural responses when interacting with our environment (Romo and de Lafuente, 2013; Schroeder et al., 2003). Importantly, it needs to be considered within the unisensory research-formulated perceptual decision-making framework outlined in previous sections (see section 1.1 *Unisensory Perceptual*

Decision-Making). Specifically, the cognitive processes of sensory information encoding, and integration into evidence for decision formation, need to consider the role of multisensory integration, and how it can benefit perceptual decision formation. The following subsection, therefore, will discern research that applied modelling approaches to gain a functional insight into how perceptual decisions benefit from the integration of multisensory information.

1.2.2 The Benefits of Multisensory Integration for Perceptual Decision-Making

The benefits of processing complementary information across multiple sensory modalities towards perceptual decision formation; notably decreases in decisional speed and increases in choice accuracy, have been extensively demonstrated in previous research (Stein & Meredith, 1993; Welch & Warren, 1980; Mercier & Cappe, 2020). Similar to previous unisensory perceptual decision-making research, such benefits can also be investigated in speeded classification paradigms, including 2AFC paradigms, by comparing the effects of multimodal and unimodal stimulus presentations on behavioural performance. Benefits to choice formation implicate increased likelihood of multisensory integration, hence increased likelihood of optimal perceptual decision formation (Drugowitsch et al., 2014). Considering the previously outlined principles of multisensory integration (see subsection 1.2.1 *Multisensory Integration*), findings have found multisensory enhancements to choice formation corresponding to each of the principles.

Concerning the spatial and temporal rules, for example, Jones (2015) reported an enhancement of temporal expectancy towards audiovisual spatial attention in a rhythmic cuing task. In this study, participants facilitated faster and more accurate behavioural responses when rhythmically cueing audiovisual stimuli compared to unisensory (i.e., auditory or visual) stimuli, suggesting multisensory percepts enhancing decision formation are more likely when based on spatially concordant rhythmic cuing of the two unisensory stimuli. Nidiffer et al. (2018) similarly found that temporally correlated audiovisual inputs facilitated greater speech perception. This suggested correlational strength of temporal coherence was an important factor for multisensory perceptual flexibility, with greater strength more likely to optimise choice behaviours. For spatial coherence, Aller et al. (2015) used dynamic Continuous Flash Suppression (CFS) to spatially collocate sounds to flashes, and found spatial audiovisual integration stimulated more robust percepts that contributed to perceptual awareness. Similar findings have further been elucidated when both spatial and temporal correspondence factors are synchronized (Stevenson et al., 2012; Van der Burg, Alais, & Crass, 2018; Zhao et al., 2022).

Concerning semantic congruency, Heikkilä et al. (2015) found improved audiovisual declarative memory (i.e. higher choice accuracy) in participants who memorised verbal and non-verbal auditory (spoken words and/or sounds) and visual (written words and/or pictures) stimuli, when presented simultaneously with semantically congruent stimuli in the alternate modality (verbal: spoken words-written words; non-verbal: sounds-pictures), compared to semantically incongruent stimuli in the alternative modality or a neutral stimulus (“noise”). Additionally, Delong & Noppeney (2021) investigated the extent that spatial and semantic congruency jointly influence the binding of audiovisual information depending on observers’ explicit conscious awareness. They employed a spatial ventriloquist paradigm across two experiments. Their findings from the first experiment demonstrated that not only did spatially correspondent audiovisual stimuli improve sound localization accuracy, but this improvement was amplified when audiovisual stimuli were semantically correspondent, thus implicating that semantic correspondences decreased susceptibility to spatial biases in auditory stimulus presentations. Their findings from the second experiment demonstrated that when forward-backward masking was applied within the spatial ventriloquist paradigm, susceptibility to spatial biases in auditory stimulus presentations increased when visual stimuli were masked (i.e., outside of conscious perceptual awareness), with improvements from semantically correspondent audiovisual stimuli diminished. Together, their findings infer that different correspondences influence multisensory integration depending on an observer’s perceptual awareness, thus modulating the benefits towards perceptual decision formation. In particular, they implicate that semantic congruency modulates auditory spatial localisation bias when visual stimuli are presented within conscious awareness.

Finally, for the principle of inverse effectiveness, Hairston et al. (2003) examined visual localisation precision of visual, auditory, and audiovisual stimuli under conditions of unaffected and degraded visual stimulus coherence. Under conditions of visual stimulus degradation, participants wore

+6-Dimensional glasses to artificially induce myopia (near-sightedness), increasing the difficulty of detecting light-emitting diode (LED; visual stimuli), broadband noise bursts (auditory stimuli), or the two respective stimuli simultaneously presented. They calculated the difference, in degrees, between perceived and actual target location, termed the localisation error. Their results revealed equal localisation performance for visual and audiovisual stimuli and impaired localisation performance for auditory stimuli in unaffected (i.e. “normal”) blocks. However, under conditions of artificially-induced myopia, participants’ localisation judgements for audiovisual stimuli significantly improved relative to localisation judgements for visual and auditory stimuli. Comparably, Rach, Diederich, & Colonius (2011) quantified multisensory interaction effects, using measurements derived from a variant of sequential sampling modelling, that integrated RTs and Detection Rates (DRs) for unimodal and bimodal stimuli across two audiovisual cue detection paradigms. Measurements of multisensory speed-up and detectability; quantified as Multisensory Response Enhancement (MRE) and Multisensory Detection Enhancement (MDE) metrics respectively, predicted observable patterns of reduced RTs and increased DRs for audiovisual stimuli when unisensory (i.e., auditory and visual) stimulus coherence decreased, and increased RTs and decreased DRs when unisensory stimulus coherence increased. Finally, Tatz, Undorf, & Peynircioğlu (2021) observed multisensory benefits towards the effect of perceptual fluency of Judgement of Learning (JOL), whereby decreased RTs and choice accuracy were found across three experiments investigating written and spoken word processing fluency in audiovisual conditions when visual and/or auditory stimulus coherence became harder to perceive. This suggested that multisensory information processing heightened JOL, relative to unisensory information processing, in accordance with the principle of inverse effectiveness. Overall, these findings demonstrate multisensory-mediated enhancements towards perceptual decision formation when information presented to unisensory modalities becomes increasingly ambiguous to process.

These findings demonstrate not only the principles of multisensory integration, but the benefits they facilitate towards perceptual decision formation and overt behavioural responses. Thus, multisensory enhancements are likely to have a key role for perceptual decision formation in experimental paradigms with multimodal information. Consequently, the neural substrates of multisensory integration, and its constituent loci within the non-serial hierarchy of computations underpinning the benefits towards choice formation, has been the focus of a myriad of perceptual decision-making studies (see Keil & Senkowski, 2018; Talsma et al., 2010; ten Oever et al., 2016, for reviews). In particular, the application of behaviourally-informed and/or neurally-informed modelling approaches has furthered our understanding of how complimentary multisensory information is integrated and consolidated within the perceptual decision-making framework; derived from studies that used unisensory stimuli (see Chandrasekaran, 2017; Liu & Otto, 2020; Seilheimer, Rosenberg, & Angelaki, 2014; Ursino, Cuppini, & Magosso, 2014, for reviews). As such, the neurocomputational processes within perceptual decision formation can also be inferred by linking behavioural measurements (i.e., RTs and/or choice accuracy) with simultaneously recorded neural activity on a single-trial basis. Previous literature utilising such modelling principles and architectures has uncovered a similar interconnected hierarchy of sensory-cortical and higher-associative brain regions underlying the cascade of separate, but not-linear, constituent stages of perceptual decision-making with multisensory information.

For example, Delis et al. (2018) probed the neural correlates of visual-tactile active sensing; a process whereby sensory receptors are directed to extract task-relevant sensory information in order to interact with the external environment. They extracted brain-behaviour couplings using both EEG and finger kinematics, and characterised the prominent neural components that were predictive of participants’ perceptual judgements during a 2AFC texture discrimination paradigm. They found three distinct EEG components that showed significant brain-behaviour couplings, correlating visual-tactile integration with RTs and choice probability of finger movements during texture discrimination, hence predicting patterns of active sensing. These components were localised to the right-lateralized occipital cortex (LOC); specifically, Area V1, middle frontal gyrus (MFG), and SMA. In order to generate mechanistic insights into the functional role of these components underlying active sensing, a HDDM framework was incorporated, thus providing a principled approach towards linking behavioural output (i.e., visual-tactile sensory processing) with representations derived from brain-behaviour couplings. First, they functionally implied that the LOC modulated the encoding of multisensory information, since component activity predicted decreases in non-decision time parameter estimations when visual and tactile information was complementary in its correspondence towards texture coarseness. Second, the MFG was seen to represent the rate of multisensory evidence accumulation towards a particular choice alternative, since it predicted greater increases in drift rate parameter estimations for visual-tactile correspondences relative to parameter estimations for visual or tactile information alone. The third component, localised to the SMA did not produce significant EEG-kinematic couplings, nor correlations with estimated HDDM parameters, but was most prominent after peak MFG component discrimination, and prior to the facilitation of choice responses, empirically validating previous research implicating it as a region responsible for motor-response production (Pleger et al., 2006; Kim et al., 2015). Furthermore, MFG activation disappeared from components uncovered from control participants when performing active sensing but not facilitating decisional responses, further elucidating functionally distinct neural correlates for sensory stimulus encoding and evidence accumulation processes. Overall, the study discerned integral processes predictive of multisensory perceptual decision-making behaviour, and neural mechanisms predicting the benefits of integrating visual and somatosensory information. In a follow-up study, using a similar variant of their previously applied 2AFC texture discrimination paradigm, coupled with an information-theoretic approach for characterising neural activity underlying multisensory enhancements towards perceptual decision formation, they found EEG activity in parieto-occipital and prefrontal brain regions that were predictive of participants' movement patterns across visual, haptic, and visual-haptic trials. Application of a HDDM framework further revealed trial-to-trial fluctuations in their neural representations of active sensing that predicted increases in drift rate (i.e. the rate of sensory evidence accumulation), across all three trial types. However, the multisensory (i.e. visual-haptic) representations of active sensing were strongly predictive of greater drift rate estimations, relative to the unisensory representations of active sensing (i.e. visual or haptic), therefore discerning a functional role of supramodal integration within active sensing, which benefits perceptual decision formation when presented with complimentary multisensory information (Delis et al., 2022).

Similarly, Regenbogen et al. (2018) incorporated fMRI BOLD signals into a HDDM framework, probing the differences in cortical processing that drive multisensory integration for clear, coherent, stimuli compared to unclear, degraded, stimuli. Their findings implicated connectivity between the IPS and fronto-parietal brain regions as a prevalent interconnected neural network for enhancing decision-making performance (i.e., decreased RTs and increased choice accuracy) when multisensory stimuli were increasingly unclear, and thus ambiguous, to process when facilitating choice responses. HDDM application further elicited predictive increases in drift rate, capturing increased sensory evidence accumulation for degraded audiovisual stimuli relative to their auditory and visual counterparts. This indicated that an interconnected network within frontal brain regions drove selective recruitment of relay networks for integrating percepts of noisy unisensory information, not only validating the principle of inverse effectiveness governing multisensory integration, but reaffirming functional insights into the benefits of integrating multimodal compared to unimodal sensory information.

Moreover, Franzen et al. (2020) coupled EEG and LDA with an audiovisual extension of the visual face-versus-car-categorisation paradigm (Philiastides & Sajda 2006a; 2006b; Phliastides, Sajda, & Ratcliff, 2007). Their study aimed to identify when and where multisensory enhancements towards perceptual decision formation arose when complementary auditory stimuli is presented simultaneously with visual stimuli. Using multivariate LDA, they capitalised on distinct neural components, which correlated to multisensory encoding of decision-relevant audiovisual evidence, revealing *Early* (approximately 180ms) and *Late* (approximately 380ms) EEG components over frontocentral and bilateral occipitotemporal brain clusters, and prominent centroparietal brain clusters respectively. Amplitude comparisons between their EEG components demonstrated that the Late component for audiovisual stimulus presentations captured increases in drift rate parameter estimations, when incorporated into a neurally-informed HDDM, further predicting higher proportions of perceptual choice accuracy. This implicated multisensory behavioural improvements could arise with enhanced quality of incoming decisional evidence, which in this instance, was complementary auditory information, thus benefitting the process of (multi)sensory evidence accumulation when facilitating choice responses. This implication is consistent with findings demonstrating the enhanced emergence of neural activity in higher-order, associative, cortices brain regions, and not sensory associative brain regions, when presented with multisensory evidence (Newell & Shams, 2007).

Finally, Mercier & Cappe (2020) used EEG to identify two temporally-distinct neural events underlying multisensory perceptual decision-making across both audiovisual cue detection and cue categorisation paradigms. These events arose from approximately 0-400ms post-stimulus onset, which overlapped with the presentation of multisensory information, and at approximately 600ms within a 200ms period until average response facilitation. The spatiotemporal dynamics of the identified EEG events constituted an evolution of neural activity emerging in parietal brain regions, then progressing towards central brain regions, in which emergent increases in amplitude peaked before decision responses. These findings reaffirm previous research identifying distinct regions within a network for processing incoming sensory information in sensory cortices, and consolidating such information as decision evidence, when facilitating perceptual decisions. Importantly, the decoding of EEG activity underlying unisensory signal cues; utilizing a supervised machine learning approach, showed that these processes were predictive of early sensory encoding and late decisional formation processes. Multisensory benefits observed in the behavioural data (i.e., faster RTs, higher accuracy, and increased sensitivity towards multisensory cues) were concurrent with an acceleration of both processing stages when a HDDM was applied. Non-decision time parameter estimations and its variability significantly decreased in multisensory compared to unisensory conditions for the earlier EEG component, but parameter estimations for drift rate variability significantly increased for the later EEG component, across both paradigms, suggesting multisensory enhancements benefitted both a faster integration of sensory information and consolidation of decisional evidence.

To conclude, application of behaviourally-informed and/or neurally-informed modelling approaches has revealed that multisensory benefits are pervasive within perceptual decision formation. Importantly, they reveal a dynamic, continuous, but somewhat non-linear, interplay between multisensory integrative and perceptual decision-making processes. This defines a critical role of multimodal information integration within the framework that previous unisensory perceptual decision-making findings have theorised. Specifically, multisensory integrative benefits have been observed within earlier sensory encoding stages and later decision formation stages. In utilising such approaches, a mechanistic insight into how this integrative mechanism consolidates information arriving at multiple sensory modalities has been gained. However, such theoretical considerations have further demonstrated that such multisensory integrative and perceptual decision-making processes are not sequential, despite earlier observations that information is integrated innately within the human brain. Instead, the processes can be considered nested (Mercier & Cappe, 2020). Due to this, the fundamental question currently asked in the field concerns *where* and *when* multisensory signals are integrated to benefit perceptual decision formation (see Bizley, Jones, & Town, 2016a; Bizley, Maddox, & Lee, 2016b for detailed reviews). Comparably, it is important to consider the underlying neural correlates for such benefits within multisensory perceptual decision-making. Undoubtedly, this involves discerning the spatiotemporal characteristics of neural mechanisms that correlate with the benefits observed from multisensory integration towards overt choice behaviours. The prominent hypotheses of whether multisensory benefits arise before and/or during perceptual decision formation will therefore be discussed in the following subsection.

1.2.3 Where Do the Benefits of Multisensory Integration Occur Within Perceptual Decision-Making?

So far, we have discerned that when forming perceptual decisions with multimodal information, multisensory interactions are pervasive within the human brain, constituting different serial, but non-linear, processes along the cortical hierarchy (Cao et al., 2019; Rohe et al., 2018; Rohe & Noppeney, 2016; Keil & Senkowski, 2018; Sadaghiani et al., 2009). Despite recent progress towards an understanding of the neural signatures associated with multisensory perceptual decision-making, and commonalities associated with neural signatures of unisensory perceptual decision-making, it remains unclear *how* the brain translates the relevant integrated multisensory information into benefits towards perceptual decision formation. Specifically, it is unclear whether multisensory representations localised in primary and non-primary sensory cortices benefit perceptual decision formation “early”, and/or whether a “late” integration of multisensory signals as decisional evidence, combined after the accumulation of unisensory evidence, occurs in higher associative brain regions such as the frontal and parietal cortices (see Bizley, Jones, & Town, 2016a; Bizley, Maddox, & Lee, 2016b; Calvert & Thesen, 2004, for reviews). Thus, for *where* and *when* multisensory information benefits perceptual decision-making, three prominent theories persist in the field: (1) the *early integration hypothesis*, (2) the *late integration hypothesis*, and (3) what we term as the *dual integration hypothesis*.

The early integration hypothesis posits that stages for early sensory encoding, and for combining complementary features of unisensory information, facilitate the influence of multisensory benefits towards perceptual decision formation (Ghazanfar & Schroeder, 2006; Kayser & Logothetis, 2007; Schroeder & Foxe, 2005). This hypothesis theorises that multisensory integration begins in the sensory cortices and contributes to an accumulation of multisensory evidence to benefit perceptual decision formation. Therefore, perceptual decision-making in high-level associative cortices is facilitated on readily available multisensory representations of information (Bizley, Jones, & Town, 2016a). Substantial evidence that supports the early integration hypothesis arises from identified interactions in neural pathways between primary sensory cortices (e.g. auditory, visual, and somatosensory cortices) and non-primary, higher-order, associative cortices (e.g. parietal, temporal, and frontal associative cortices) in the brain, but with prominent neural responses localized early within sensory cortices, thus highlighting multisensory interactions in putative “unisensory” cortices (Eckert et al., 2008; Ghazanfar & Schroeder, 2006; Giart et al., 1999; Kayser et al., 2017; Petro et al., 2017; Rohe & Noppeney, 2016). Comparably, neural pathways have been demonstrated to project directly to primary sensory cortices within the processing hierarchies of distinct modalities in animal model studies (Cappe, Rouiller, & Barone, 2009), as demonstrated by, for example, Falchier et al. (2002) and Majka et al. (2019); who found cortico-cortical pathways between core and parabelt areas of the auditory cortex, superior temporal polysensory (STP; area TPO-1) cortex, and area 17/V1 within the visual cortex in Cynomolgus Macaque (*Macaca fascicularis*) and Marmoset (*Callithrix jacchus*) monkeys respectively, and Henschke et al. (2015); who found matched and non-matched thalamocortical nuclei projections between areas V1 (visual cortex), A1 (auditory cortex), and S1 (somatosensory cortex) in Mongolian gerbils (*Meriones unguiculatus*); using retrograde tracing.

Furthermore, numerous studies have correlated multisensory interactions in sensory cortices, through identified neural indices of multisensory integration, with outcomes representative of enhancements in decision behaviours (i.e., decreased RTs and/or choice accuracy). For instance, Mercier et al. (2013) focused on auditory and multisensory activity in posterior visual cortices using continuous intracranial EEG (iEEG) and a simple detection paradigm; responding with keyboard button presses whenever they observed auditory-alone, visual-alone, or audiovisual stimuli, in order to investigate whether phase resetting of ongoing intrinsic neural oscillations would impact the integration of audiovisual information in visual cortical regions. They reported early-latency auditory-induced phase resets over visual cortex, as well as auditory-evoked potentials (AEPs), that differed from AEPs traditionally observed over auditory cortex in amplitude, latency, and oscillatory frequency (Molholm et al., 2006). Interestingly, multisensory integrative effects were observed when audiovisual stimuli were presented through increased phase concentration and amplitude of identified ERPs over Brodmann Areas 17, 18, 19, 37 and 39, which predicted faster RTs relative to unisensory stimulus presentations. These findings indicate neural markers for early multisensory integration in visual cortices, most notably, early cross-sensory phase setting to optimally integrate incoming multimodal sensory information.

Moreover, Rajj et al. (2010) utilized fMRI and MEG to identify pathways and timings of early audiovisual activations and interactions in primary auditory cortex (A1) and primary visual cortex (V1) in human brains. Using auditory and visual “noise” bursts and chequerboard stimuli, presented for 300ms, they reported early latency onsets of neural activity for both modality-specific and modality-alternate stimuli across both primary cortices (Heschl’s gyri and Calcarine fissure for A1 and V1 respectively); with the onset of auditory-driven activity observed at 23ms and 53ms, and the onset of visual-driven activity observed at 43ms and 53ms in A1 and V1 regions respectively, thus implicating early neural origins for cross-sensory activations between the auditory and visual sensory modalities. In addition, they implicated that cross-modal activations originate from the sensory cortex of the alternate stimulus modality, as evidenced by the asymmetrical activations between modalities, thought to be due to sensory-specific conduction delays, and given that audiovisual interactions occurred within a consistent timeframe range of 3-21ms (after 300ms onsets). Finally, Lauzon et al. (2022) investigated the neural indices underlying the interrelation between multisensory associative learning and multisensory integration. They applied EEG to participants who engaged in a three-auditory stimulus oddball detection paradigm; a target detection task whereby participants’ responses, behavioural and/or neural, to a deviant stimulus presented within sequences of repetitive stimuli, are recorded as an index of pre-attentive processing (Courchesne, Hillyard, & Galambos, 1975). They found that ERPs of Mismatch Negativity (MMN; between ~216-252ms) and P300 (i.e. a positive deflection in electrical potential voltage; between ~332-440ms), over bilateral fronto-central and occipital brain regions, correlated strongly with ERP fluctuations, and were recorded as early as 48ms (and approximately <100ms) over left-localised centro-parietal brain regions as participants engaged in a speeded-response task for detecting deviant audiovisual shape-tone pairings relative to auditory-tone or visual-shape stimuli. Importantly, they found that their neural markers of audiovisual associative learning and integration predicted lower RTs, thereby faster perceptual decision-making, inferring that learned associations strengthen the early integration of novel multisensory information.

While research implicating early cross-modal integration offers the possibility that multisensory evidence representations expedite later decision-making processes, discerning the functions of earlier cross-modal integration in shaping perceptual decision formation, as well as subsequent behaviour, remains subject to interpretation. Previous research has instead argued in favour of the late integration hypothesis, which argues that post-sensory enhancements of decision evidence arise from a late integration of multisensory information, therefore benefitting perceptual decision formation (Bizley, Jones, & Town, 2016a; Franzen et al., 2020). It 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. To an extent, this hypothesis thereby posits that multisensory integration occurs immediately prior of perceptual decision formation, with unisensory evidence accumulated in independent modalities (thereby within the respective associative cortices) and consolidated as unisensory decision variables within higher-order cortices, hence integrated later in time as multisensory decision evidence to facilitate appropriate behavioural responses. Empirical support for this hypothesis comes from research investigating the processes of object categorisation and recognition (Lee, Maddox, & Bizley, 2019). Given these are naturally multisensory processes, since information is presented to multiple sensory modalities near-simultaneously, researchers have contended that top-down control processes are required to integrate salient accumulated unisensory information in order to expedite the speed and accuracy of perceptual decision formation (Calvert & Thesen, 2004; Chen & Spence, 2017; De Meo et al., 2015). Findings from Raposo et al. (2012) and Shepperd et al. (2013), for example, demonstrated improved multisensory perceptual decision-making behaviour, through increased choice accuracy, using an audiovisual rate categorisation paradigm. This paradigm instructs participants (both humans and rats) to categorise brief auditory (tones) and/or visual (flashes) events as “fast” or “slow” when events varied both in stimulus strength (through manipulations of trial-averaged event rates) and reliability (through manipulations of Signal-to-Nosie ratios; SNR ratios). Both studies reported consistent findings, in which clear multisensory enhancements were demonstrated as a function of stimulus reliability and not stimulus strength. Crucially, psychophysical analyses implicated that when unisensory stimuli were presented in asynchronous conditions with alternating trial-averaged event rates, they were less likely to be integrated into unified percepts to benefit choice formation, and instead were estimated separately for each modality. This led the authors of both studies to conclude that decisions are formed separately for dynamic, time-varying, stochastic unisensory stimuli, and prior to their integration in order reach a multisensory decision. These studies empirically validate the late integration hypothesis through the consideration that observers weigh individual sensory cues in proportion to the variability of their reliability. This mechanism is referred to as *reliability-weighting* (Ernst & Banks, 2002; Ernst & Bülthoff, 2004; Fetsch et al., 2012; Rohe & Noppeney, 2016). Reliability-weighting traditionally supports late cross-modal integration of sensory inputs, whereby the most veridical, least variable, and statistically near-optimal multisensory percepts are accumulated for forming perceptual decisions. Hence, the likelihood of optimal behavioural responses for navigating our external environment are maximised, as the brain facilitates decisions based on integrated percepts providing salient, yet complementary, information, across multiple sensory modalities.

For example, Beauchamp, Pasalar, & Ro (2010) delivered touches to the hands of participants as they simultaneously watched a video of an artificial hand being touched. When the temporal synchrony of sensory information was manipulated (i.e., a touch occurred in one modality but not the other), they observed evidence of reliability-weighting through increased detection of touch stimuli weighted towards the more reliable modality. They analysed fMRI activity to determine the neural substrates that corroborated with increased stimulus detection, observing increased BOLD activity in the inferior parietal lobe (IPS), lateral occipital cortex (LOC), and/or somatosensory cortices, depending on the bias in visual-somatosensory reliability-weighting of incoming stimuli. Similarly, Boyle, Kayser & Kayser used task-based EEG to probe neural activity underlying mechanisms for multisensory integration and reliability-weighting of audiovisual information. They localised neural modulations at approximately 84ms post-stimulus onset over occipital and temporal regions, and approximately 120-204ms over occipital, temporal, and parietal regions of the brain respectively. Their results reaffirmed common behavioural neural origins for the integration and weighting of multisensory information for facilitating perceptual decisions that depended on the coherence of the audiovisual stimulus presentations, with coherence-induced modulations predicting enhanced neural processing beyond sensory cortices, and towards associative cortices thought to be implicated in sensory evidence accumulation for choice responses. Prominent regions linked to the reliability-weighting for multisensory perceptual decision-making includes the parietal cortex, which is thought to have a prominent role for integrating and accumulating unisensory inputs (Herding et al., 2019; Rohe & Noppeney, 2016; Sereno & Haung, 2014; Werner & Noppeney, 2010).

Finally, the dual integration hypothesis posits that unisensory information is integrated at both early sensory encoding and late decision formation stages. This consolidates a process of causal inference for determining whether or not multisensory information is supramodal in determining the temporal locus of integrating incoming sensory information (Aller & Noppeney, 2019; Cao et al., 2019; Gau & Noppeney, 2016; Kayser & Shams, 2015; Mercier & Cappe, 2020; Rohe et al., 2019; Rohe & Noppeney, 2016; Shams & Beierholm, 2010; Su, 2014). The previous research highlighted from Mercier & Cappe (2020) demonstrated evidence to support this hypothesis. Their study identified two temporally distinct neural processes underlying multisensory perceptual decision-making across both cue detection and cue categorisation paradigms. Importantly, decoding of EEG activity underlying unisensory 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 choice accuracy, and increased perceptual sensitivity towards multisensory cues) were concurrent with an acceleration of both processing stages, suggesting that the complementary information between unisensory cues benefitted both a faster integration of sensory information and consolidation of decisional evidence. Crucially, their findings implicate that where we observe correlates of multisensory integration in the brain benefitting sensory evidence accumulation, whether it is sensory cortical, or associative cortical, areas of the brain may depend on (a) the nature of the task paradigm applied to investigate the subsequent neurocognitive underpinnings and b) whether multisensory cues are cross-modally bound, in other words, thought to be perceived as originating from the same perceptual object (Bizley, Maddox, & Lee, 2016b). To this end, the process of causal inference supports the dual integration hypothesis. It suggests that repeated prior exposure to multisensory percepts optimises perceptual decision formation, as the brain integrates multisensory inputs that can be flexibly utilised to increase the likelihood of optimal choice responses. Hence, sensory inputs can be integrated early, with integrated percepts instead bound together, and then accumulated as decisional evidence, to causally infer previously formed choice responses, thus benefitting overt behavioural outcomes.

At a neurophysiological level, the statistical optimization of single neural responses to multisensory stimuli has been thoroughly documented (see Stein & Stanford, 2008, for a review). Accordingly, such findings have begun to be reconciled with computational modelling approaches to provide neural and/or cognitive bases for the mechanism of causal inference, coupled with reliability-weighting, enhancing multisensory perceptual decision-making behaviour (see Angelaki et al., 2009; Chandrasekaran, 2017, for reviews). These implicate that perceptual decision-making processes are recalibrated through repeated exposure to unisensory signal discrepancies. For example, Park & Kayser (2019) used MEG to investigate whole-brain activity during an audiovisual spatial localization task. They localized neural substrates for intra-trial integrative processes within temporal and parietal cortex regions, and trial-to-trial perceptual recalibrations to decision-making within overlapping medial superior parietal cortex. Similarly, Cao et al. (2019) provide further MEG evidence for functional connectivity between the two processes. Using an audiovisual rate categorisation task, they identified parietal-temporal regions and prefrontal cortex activity for speeded decisions to be made about the temporal rate of audiovisual pulse repetitions. Both studies provide evidence suggesting multisensory information causes an updating of integrative and perceptual decision-making processes, maximising the likelihood of elicited behavioural benefits within initiated actions. Computational approaches have yielded mechanistic interpretations for a role of multisensory evidence recalibration using models for Bayesian Integration and Bayesian Inference (Acerbi et al., 2018; Körding et al., 2007; Noppeney, 2021). Furthermore, findings have derived that prior expectations, formed from repeated exposure to multisensory percepts that causally infer the same object, benefit perceptual decision formation. For example, Ferrari &

Noppeney (2021) combined fMRI with a Bayesian model of causal inference to demonstrate that pre-stimulus anticipation and post-stimulus reporting of the spatiotemporal locus of auditory and visual stimuli modulated neural activity in the visual and/or posterior parietal cortices respectively. Similarly, Mahini et al. (2017) instructed participants to respond to spatial discrepancies in visual-tactile and visual-auditory-tactile stimuli to investigate optimal causal inference in discerning cue conflicts. Psychophysical analyses discerned a threshold for forming decisions that moved from integration of information to selection of cues of participants’ own discerned reliability, highlighting a weighting of information towards the modality that benefitted the accuracy of choices. These reaffirm that we process multisensory information to optimize perceptual decision-making through formulating prior probabilities of complementary sensory information. Therefore, causal inference operates on formed percepts with maximal prior probabilities of yielding salient information informing the guidance of perceptual and motor decisions (Filimon et al., 2013; Verdonck, Loossens, & Philiastides, 2022). To conclude, such computational frameworks reinforce the dual integration hypothesis.

The recent emergence of findings that empirically validate the dual integration hypothesis, by coupling reliability-weighting as a component of causal inference mechanisms, suggests that multisensory enhancements towards perceptual decision formation do not originate from a specific timeframe. Rather, the suggestion that multisensory benefits towards choice behaviour originate from one of earlier sensory encoding stages, or post-sensory stages as a product of decision dynamics, is a false dichotomy. However, a reconciliation of the two competing hypotheses, in which multisensory perceptual decision-making enhancements may arise from either an accumulation of unisensory evidence within independent modalities, or an accumulation of bound unisensory evidence, implicates that common neural and cognitive mechanisms determine the constituent benefits of multisensory integration towards perceptual decision formation.

This has important applications for applied fields of research investigating the underlying mechanisms of multisensory perceptual decision-making, in particular, the effects of natural ageing (de Dieuleveult et al., 2017; Mozolic et al., 2012). It has been extensively demonstrated, for example, that older adults exhibit impairments in the weighting of sensory information across sensory modalities (Guerreiro, Murphy, & Van-Gerven, 2013; McGovern et al., 2014), coupled with deficits to baseline (i.e., “bottom-up”) multisensory filtering that have been considered pivotal in disrupting higher-order (i.e., “top-down”) mechanisms for selective attentional control, for example, in postural stability (Redfern et al., 2001; 2009). Such findings attribute that ageing impacts earlier processes for encoding sensory information, arguing for the early integration hypothesis in which impaired multisensory representations in older adults degrades the outcome of decision behaviours. However, neuroimaging findings have countered this argument (Alain et al., 2022). For example, Peiffer et al. (2009) coupled fMRI with auditory and visual discrimination paradigms to assess the ability of auditory and visual sensory cortices to suppress neural responses in the alternate cortex, in which the suppression of neural responses in the deactivated sensory cortex is found to decrease distractibility in weighting unisensory information. They found discrepancies in the visual cortical regions for cross-modal deactivations, in which older adults benefitted from suppressing neural responses in dorsal regions, compared to younger adults, but not for ventral regions, in which younger adults benefitted from suppressing neural responses. Furthermore, these differences held when accounting for reductions in gray matter volume within the sensory cortices, further implying that age-related differences in brain structure and function cannot solely account for disruptions to earlier multisensory encoding. Furthermore, it has been extensively demonstrated that older adults exhibit preserved multisensory integration despite impacts on the projection of sensory cortices onto regions of the brain linked to combining multisensory signals; notably the superior colliculus (SC) and superior temporal sulcus (STS; Calvert & Thesen, 2004; Clemo et al., 2012). Whereas this demonstrates support for preserved multisensory integrative benefits due to later post-sensory processes (and thus empirical support for the late integration hypothesis), it contradicts behavioural research that found a general slowing of cognitive and motor processes as a function of age (Laurienti et al., 2006; Peiffer et al., 2007). Thus, in the context of the dual integration hypothesis, prospective future research needs to consider reconciling why older adults may exhibit preserved (and to an extent enhanced; Laurienti et al., 2006; Peiffer et al., 2007) multisensory integration within perceptual decision-making despite inherent declines in earlier sensory and later post-sensory decision dynamics relative to younger adults.

In conclusion, to fully understand the role of multisensory integration within perceptual decision formation, and how it benefits choice behaviours, it is necessary to establish the perceptual relevance of the integrative mechanism at different stages of perceptual decision formation.

1.3 Aims and Outline of the Thesis

As highlighted in this chapter, there is considerable research characterising the neural and cognitive underpinnings of unisensory perceptual decision-making. This formed the basis to formulate a framework that conceptualises multisensory perceptual decision-making, and therefore consolidates predictions for behavioural enhancements from the process of multisensory integration. The application of behaviourally-informed and/or neurally-informed modelling approaches has posited a serial, non-linear, hierarchical cascade of neural events underlying the cognitive processes for multisensory perceptual decision-making. Importantly, a mechanistic, or functional, insight into how perceptual decision formation is enhanced through the integration of multisensory information into unified percepts, and consolidation of percepts as choice evidence for optimising decisional behaviours (i.e., faster and/or more accurate decisions) has steadily been uncovered within perceptual decision-making literature. The persisting key question, however, concerns the spatiotemporal locus of the benefits from multisensory integration within the processes of perceptual decision-making. Specifically, *where*, *when*, and *how* in the brain we observe the benefits from combining complementary multisensory signals remains subject to debate (Bizley, Jones, & Town, 2016a; Bizley, Maddox, & Lee, 2016b; Otto, Dassy & Mamassian, 2013; Otto & Mamassian, 2017).

Perceptual decision-making in humans, both unisensory and multisensory, can be modulated by a variety of intrinsic and extrinsic (i.e., environmental) factors. In addition to incoming sensory information, these factors have the capacity to influence different cognitive and neural processing components within the perceptual decision formation process, thus affecting facilitated choice responses. We define these factors as *decision modulators*, or *decision neuromodulators*, depending on whether their influence is exerted on cognitive and/or neural perceptual decision-making mechanisms (Doya, 2008; Philiastides, Diaz, & Gherman, 2017; Siegel, Engel, & Donner, 2011). Examples of decision (neuro)modulators include reinforcement learning, which captures the potential effects of reward and punishment on value-based perceptual decision-making, (Fouragnan et al., 2015; 2017; Pisauro et al., 2017), confidence (Gherman & Philiastides, 2015; 2018; Boldt et al., 2019), and prior probability or pre-stimulus expectations (Diaz, Delis, & Philiastides, 2021; Kelly, Corbett, & O'Connell, 2021).

Furthering our understanding of how various decision (neuro)modulators influence the processes of perceptual decision formation, and how such influences can be accounted for within behaviourally-informed (i.e., for the underlying cognitive mechanisms) and/or neurally-informed modelling (i.e., for the underlying neural mechanisms) approaches of perceptual decision-making; notably sequential sampling model approaches, can not only benefit our understanding of the processes within multisensory perceptual decision-making, but where, when, and how multisensory benefits consequently emerge to enhance choice behaviours. Through gaining a functional insight into the exerted influence(s) of established decision (neuro)modulators on the cognitive processes and neural mechanisms of multisensory perceptual decision-making, we can contribute to the current understanding of where multisensory signals are combined to benefit perceptual decision-making, by examining whether modelled cognitive processes and/or neural signals are related to choice behavioural outcomes.

The work in this thesis aimed to provide insights into the cognitive and neural mechanisms underlying multisensory perceptual decision-making. We focused on audiovisual perceptual decision-making, comparing the cognitive and/or neural mechanisms for bimodal (i.e., auditory and visual) and unimodal (i.e., auditory or visual) stimuli. Specifically, it sought to investigate the constituent influences exerted by three decision modulators: (1) unisensory-formulated cross-modal associations (Chapter 2), (2) natural ageing (Chapter 3), and (3) perceptual learning (Chapter 4). To do so, we utilised a combination of EEG-based neuroimaging, 2AFC decision-making paradigms (recording single-trial RTs and binary choice responses), single-trial EEG decoding (using Linear Discriminant Analysis; LDA), and computational modelling (using the HDDM; Wiecki et al., 2013) to incorporate behaviourally-informed (for Chapter 3) and/or neurally-informed (for Chapters 2 and 4) approaches for provide novel insights into (a) the spatiotemporal locus of neural mechanisms underlying multisensory perceptual decision-making within the brain, (b) the cognitive processes mechanistically underlying identified spatiotemporal dynamics, and (c) the modulatory effects of our nominated decision modulators on the underlying neural mechanisms and/or cognitive processes.

Chapter 2

Neurocomputational Mechanisms underlying Cross-Modal Associations and their Influence on Perceptual Decisions[2]

2.1 Abstract

When exposed to complementary features of information across sensory modalities, our brains formulate cross-modal associations that bind features of stimuli presented separately to multiple modalities. For example, auditory pitch-visual size cross-modal 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 influence behavioural performance, i.e., choice accuracy and RTs, compared to non-preferential, or incongruent, cross-modal associations across multisensory decision-making paradigms. However, the neural mechanisms underpinning such influences in perceptual decision formation remain unclear, with discrepancies persisting as to whether associative congruency underlies ‘early’ sensory processing benefits, or ‘late’ post-sensory modulations in decision dynamics, due to the issue of mixed selectivity within neural signals when utilising experimental paradigms that present two or more unisensory features simultaneously. Here, we sought to address this issue and identify when perceptual improvements from associative congruency emerge in the brain during decision formation. Using a modified version of the Implicit Association Test (IAT), coupled with EEG, we measured the neural activity underlying the effect of auditory stimulus-driven pitch-size associations on perceptual decision formation. Since the IAT only presents one unisensory stimulus per trial, the issue of mixed selectivity within neural signals can be overcome. Behavioural results showed that participants responded significantly faster (i.e., lower RTs) during trials when auditory pitch was congruent, rather than incongruent, with its associative visual size counterpart. Using multivariate LDA, we characterised the spatiotemporal dynamics of EEG activity underpinning IAT performance, characterising an *Early* component (approximately 100-110ms post-stimulus onset), coinciding

---

[2] This work has been published as Bolam, J., Boyle, S.C., Ince, R.A.A., & Delis, I. (2022). Neurocomputational mechanisms underlying cross-modal associations and their influence on perceptual decisions. *NeuroImage*, *247*, 118841, 1-15.

with the time of maximal discrimination of the auditory stimuli, and a *Late* component (approximately 330-340ms post-stimulus onset) underlying IAT performance. In order to characterise the functional roles of these components in perceptual decision formation, we incorporated a neurally-informed HDDM, revealing that the Early component predicted increases in the duration of sensory-encoding processes for incongruent trials, and the Late component predicted decreases in response caution for congruent trials, requiring less sensory evidence needed to be accumulated to facilitate choice behaviours. Overall, our results provide a mechanistic insight into the contribution of ‘early’ sensory processing, as well as ‘late’ post-sensory neural representations of unisensory-driven associative congruency to perceptual decision formation.

2.2 Introduction

When exposed to complementary features of information across sensory modalities, the brain predominantly associates features of information in close spatial and/or temporal proximity. However, it has been demonstrated that the brain further embodies implicitly learnt *cross-modal associations* between arbitrary features of stimuli presented separately to multiple modalities (see Parise & Spence, 2013; Spence & Deroy, 2013; Spence, 2011, for detailed reviews). Cross-modal association(s) define(s) the brain’s tendency to systematically bind stimulus features based on their orthogonality across multiple modalities (Bizley, Maddox, & Lee, 2016b; Ernst & Bülthoff, 2004), in which features that correspond to the same object become redundantly associated, thus enabling the brain to exploit the correlation between informational cues when forming perceptual decisions from ambiguous, and often noisy, unisensory information (Bien et al., 2012; Glicksohn & Cohen, 2013). This is due to either an implicit lower-level statistical or structural correspondence across polar dimensions, or through an explicit higher-level degree of semantic coherence (see Parise & Spence, 2013; Spence, 2011, for detailed reviews).

A paradigmatic example demonstrated extensively across previous literature concerns cross-modal associations between auditory pitch and visual size (Bien et al., 2012; Evans & Tresiman, 2010; Gallace & Spence, 2006; Mondloch & Maurer, 2004; Parise & Spence, 2009; 2008; 2012). Other audiovisual cross-modal associations have been demonstrated with alternative statistical correspondences, such as between auditory pitch and visual brightness (Marks, 1987, Klapetek et al., 2012), auditory pitch and visual elevation (Jamal et al., 2018; McCormick et al., 2018, Zeljko et al., 2019), auditory pitch and visual lightness (Brunel et al., 2015, Zeljko et al., 2019), auditory pitch and visual shape (Köhler, 1929; Marks, 1987; Parise & Spence, 2012), and higher-order semantic coherence (Marks, 2004; Parise & Spence, 2013; Revill et al., 2014; Sadaghiani et al., 2009; Spence & Deroy, 2013; Spence, 2011).

Cross-modal associations have been shown to influence the consolidation of multisensory information when forming perceptual decisions (Bizley, Maddox, & Lee, 2016; Drugowitsch et al., 2014; Engel, Senkowski, & Schneider, 2012). 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*. Congruent auditory pitch-visual size associations, for example, implicitly map high-pitch tones with small-size objects, and low-pitch tones with large-size objects, whereas incongruent auditory pitch-visual size cross-modal associations map high-pitch tones with large-size objects, and low-pitch tones with small-size objects (Bien et al., 2012; Evans & Treisman, 2010; Marks, 1987; Shang & Styles, 2023)

The direction and magnitude of the effects of associative congruency depends on the type and context of the experimental paradigm used. Earlier research adopted *explicit* paradigms that used methodologies requiring participants to directly map features across multiple modalities based on their preferred associative congruency (Köhler, 1929, 1947; Davis; 1961; Sapir, 1929; Ramachandran & Hubbard, 2001a; 2001b; Stevens & Marks, 1965; Zigler, 1930). A seminal study by Köhler (1929), in particular, presented participants with two nonsensical words, *takete* and *baluma*, and two irregular shape outlines, a jagged/spikey one and a rounded/smooth one. When participants were explicitly instructed to match the words with the shapes, participants were more likely to map the word *takete* with the spiky/jagged shape outline and the word *baluma* with the rounded/smooth shape outline. Comparably, Steven and Marks (1965) instructed participants to adjust the magnitude of the brightness of a visual stimuli to subjectively match the magnitude of the loudness of an auditory stimuli and vice versa. They found at various levels of intensity; participants demonstrated consistently similar corroborations when assessing the cross-modal magnitude of stimulus energy. A constrained variant of this paradigm from Marks (1987; 1989) instead instructed participants to choose which of two presented stimuli, on a trial-by-trial basis, in a given sensory modality matched the target stimulus in the alternate sensory modality, observing equally similar corroborations across participants. Further methodologies for assessing explicit compatibility effects of associative congruency include eye tracking for preferential looking (Walker et al., 2010), semantic differential techniques (Albertazzi et al. 2020; Osgood, 1960; Oyama, Yamada, & Iwasawa, 1998), stimulus cueing (whereby the explicitly stated associative congruency of a cue preceding a target stimulus has been found to modulate reaction times of target stimulus categorisation; Chiou & Rich, 2012), speech vocalisation analyses (Parise & Pavani, 2011), and memory association in synaesthesia (a perceptual phenomenon whereby stimulation in one sensory modality leads to involuntary stimulation in a separate sensory modality, e.g., grapheme-colour synaesthesia; Ramachandran & Hubbard, 2001a; 2001b).

Recently, studies have adopted *implicit* paradigms that do not require participants to directly map features across multiple modalities based on their preferred associative congruency. These investigate how the associative congruency of presented stimulus features modulates perceptual decision formation without an observer’s introspection. Considering research investigating auditory pitch-visual size cross-modal associations, for example, Parise & Spence (2008) investigated the modulatory effects of auditory pitch-visual size associative congruency on the temporal ventriloquism effect (Morein-Zamir et al., 2003). They found that when asked to judge the temporal order of two different-sized visual stimuli (large and small grey circles), participants showed increased choice accuracy (i.e., higher sensitivity temporal order judgements; TOJs) when judging the temporal order of visual stimuli with congruent auditory tones preceding and following the display of each visual stimulus respectively. In a follow-up study, they further demonstrated that participants were able to judge the spatial discrepancy of an auditory stimuli less accurately (i.e. higher just noticeable difference discrimination thresholds; JNDs) when congruent with the visual stimulus presented (Parise & Spence, 2009). Similarly, Bien et al. (2012) presented participants with simultaneous visual (small-size/large-size circle) and auditory (high-pitch/low-pitch tone) stimulus features that differed in spatial location (left/right of a fixation cross). Participants were then instructed to make a forced-choice judgement of where the presented auditory stimulus spatially originated. They found that participants were less accurate in judging the spatial origin of the auditory stimulus when the stimuli were congruent (i.e. high-pitch tone/small-size circle, low-pitch tone/large-size circle) than incongruent (high-pitch tone/large-size circle, low-pitch tone/small-size circle), implicating that the spatial location of the visual stimulus “pulled” the auditory stimulus towards its location, suggesting a decisional bias of associative congruency that begins to modulate early sensory-perceptual bottom-up processes within multisensory decision-making.

In multisensory decision-making research, speeded classification paradigms, such as the 2AFC paradigm, have been used to demonstrate that the associative congruency of cross-modal associations benefits the consolidation of multisensory information when forming perceptual decisions (Bizley, Maddox, & Lee, 2016; Drugowitsch et al., 2014; Engel et al., 2012). In particular, enhanced behavioural performance in speeded classification paradigms has been evidenced by increased response speed (i.e., decreased reaction times; RTs; Kayser & Kayser, 2018; Laurienti et al., 2004; Silva et al., 2017), increased choice accuracy (Franzen et al., 2020; Kayser & Kayser, 2018; Kayser, Philiastides, & Kayser, 2017; Kim et al., 2008), and improved stimulus detection (Adam & Noppeney, 2014; Aller et al., 2015), when stimulus feature mappings are congruent versus incongruent, even if one stimulus feature was irrelevant to perceptual decision formation.

Considering research investigating auditory pitch-visual size cross-modal associations again, for example, Gallace & Spence (2006) used a visual discrimination paradigm, whereby participants were instructed to discriminate the size of visual stimuli (small/large light gray disks) while ignoring task-irrelevant auditory stimuli (high-pitch/low-pitch tones). They found that participants responded more rapidly (i.e., decreased RTs) when auditory stimulus pitch was congruent with the visual stimulus size than when incongruent, and when no auditory stimulus was presented. Similarly, Evans & Treisman (2010) used speeded classification paradigms to investigate cross-modal correspondence mappings of auditory pitch to four visual stimulus features (position, size, spatial frequency, and contrast). Comparisons of unisensory and bimodal simultaneous presentations of congruent and incongruent auditory and visual stimulus pairings revealed faster RTs and lower error rates (i.e., higher proportion of correct responses) for categorising congruent compared to incongruent auditory pitch-visual size pairings. Furthermore, these findings were consistent across a manipulation of task instructions, with associative congruency effects prevalent even when task instructions changed from directly categorising small/large-size circular gratings and high/low-pitch tones (“direct” task instructions) towards categorising whether auditory stimuli were sounds of a violin (high-pitch tone) or a piano (low-pitch tone; “indirect” task instructions), implying that associative congruency influences choice performance inherently and does not require relevant instructions to benefit perceptual decision formation. Comparably, Brunetti et al. (2017) reported faster (i.e., decreased RTs) and more accurate (i.e., higher proportion of correct responses) classifications of visual stimulus size (small/large black disks) in a visual classification paradigm while hearing concurrent high/low-pitched task-irrelevant sounds. In addition, when presented with an intermediate auditory-pitch tone on interleaved trials, they found that further evidence of faster classification (i.e., decreased RTs) of visual stimulus size when the intermediate auditory-pitch tone was sequentially congruent (i.e., lower/higher following the presentation of a high-pitch/low-pitch tone) than sequentially incongruent (i.e., higher/lower following the presentation of a low-pitch/high-pitch tone) with the auditory-pitch tone on the preceding trial, demonstrating relative flexibility in the nature of auditory pitch-visual size cross-modal associations in benefitting perceptual decision formation.

Interestingly, the associative congruency of cross-modal associations can modulate the consolidation of multisensory information when forming perceptual decisions when only one stimulus is presented on a single-trial basis. This has been demonstrated particularly in research that used the Implicit Association Test (IAT). The IAT is a 2AFC categorization task that measures implicit perceptual associations between two arbitrary stimulus features through the manipulation of stimulus feature-response key mappings (Greenwald et al., 1998). Using a block design, two stimulus features are assigned, or *mapped*, to the same response key in one block of trials, whereas in a separate block of trials, they are assigned to different response keys. Participants are then instructed to categorise, as quickly and as accurately as possible, which of the four stimulus features was presented using the correctly assigned response key. RTs and choice accuracy are collected as single-trial dependent variable measurements quantifying behavioural performance (and perceptual decision formation). The IAT assumes that the *associative congruency* of stimulus feature-response key mappings modulates behavioural performance, with perceptual choices faster (i.e., lower RTs) and/or more accurate (i.e. higher choice accuracy) when stimulus features are assigned to the same response key than when assigned to different response keys (i.e. higher RTs and/or lower choice accuracy). The IAT holds a symmetry between stimulus feature-response key mappings within blocks. Furthermore, at the beginning of each block of trials, participants only received new explicit instructions detailing which stimulus features are assigned to each response key, and no instructions to associate stimulus features assigned to the same response key. Hence, the outcome focuses on the implicit impact of associative congruency, reflecting both implicitly-formed associations and explicitly-attributed response mappings.

For example, Parise & Spence (2012) used the IAT to investigate the modulatory benefits of associative congruency for five types of cross-modal association (*takete-maluma;* Köhler, 1947, *mil-mal*; Sapir, 1929, auditory pitch-visual size, auditory pitch-visual angle, auditory pitch-visual shape). For all five types of cross-modal association, they showed decreased RTs for congruent, compared to incongruent, pairings when only one unisensory stimulus feature was presented per trial. Silva & Bellini-Leite (2020) replicated this finding utilising the IAT to investigate *bouba-kiki* cross-modal associations. Using a sine-wave stimuli, they found that when participants were trained to hear the sine-wave stimuli as speech, participants' speech-specific processing improved in an explicit cross-modal matching task, whereby behavioural benefits were observed underlying the mapping of the nonsensical words of *bouba* and *kiki* with their corresponding visual stimuli (curved shapes for *bouba*, jagged/sharp shapes for *kiki*). Moreover, these findings are consistent with those in studies that utilised the IAT to formulate implicit cross-modal associations with more ecological stimuli. For example, Demattè et al. (2006) found that when participants made speeded categorisations to two odours (strawberry and spearmint) and colour patches (pink and turquoise), participants showed decreased RTs and increased choice accuracy for odour-colour associations assigned to the same response key (strawberry-pink; spearmint-turquoise) than when assigned to different response keys. In addition, Roque et al. (2020) found decreased RTs, and hence faster response speed, for congruent pairings between visual bubble size and auditory pouring sounds than for their incongruent pairings. Importantly, these findings were all observed when only one stimulus feature was presented on a single-trial basis, and with no explicit reference or instructions as to how the stimulus features should be mapped. As such, these results implicate that the associative benefits of cross-modal associations emerge implicitly earlier in the perceptual decision formation, and not exclusively within later explicit decision-related processes.

However, when the benefits of associative congruency emerge within perceptual decision formation remains subject to extensive debate within cross-modal associative research (Spence & Deroy, 2013). Accordingly, investigating the underlying neural basis of cross-modal associations within perceptual decision formation has recently become a focus of human electrophysiology and neuroimaging research (Spence, 2011; Bizley, Jones, & Town, 2016). For example, Bien et al. (2012) used EEG to localise an effect of auditory pitch-visual size associative congruency across two identified Event-Related Potentials (ERPs) at approximately 250ms and approximately 300ms across parietal and frontal electrodes respectively. They then applied Transcranial Magnetic Stimulation (TMS) over right-lateralised parietal cortex; a region of the brain found to be responsible for the integration and processing of congruent multisensory inputs (Matsuhashi et al., 2004; Molholm et al., 2006; Pasalar, Ro, & Beauchamp, 2010). Using TMS, they observed that their identified ERPs were abolished, diminishing participants' behavioural performance. Comparably, Diaconescu et al. (2011) used Magnetoencephalography (MEG) to identify neural modulations of associative semantic congruency. They found that approximately 100ms post-stimulus onset, posterior parietal brain regions elicited increased amplitude modulations to congruent cross-modal stimuli irrespective of task instructions or higher-order semantic coherence, implicating an early effect of implicit associative congruency. Increased neural modulations in superior temporal and posterior cingulate cortices were identified between 200m and 400ms post-stimulus onset when participants were instructed to classify semantic categories, which extended earlier from 100ms to 400ms post-stimulus onset across parahippocampal, dorsomedial, and orbitofrontal cortices when participants were instructed to classify whether semantic pairings were congruent or not. Similarly, Kovic, Plunkett, & Westermann (2010) used EEG, and found significant differences in neural activity between congruent and incongruent learned label-object associations as early as 140ms across occipital regions of the brain. Furthermore, when participants were asked to identify mismatches to the learned label-object associations, detected mismatches evoked a modulation in activity between 340ms and 520ms across parietal regions.

The above studies have started to shed light on the neural underpinnings of associative congruency. Notably, they localised earlier perceptual origins of associative congruency, as well as later task-dependent modulations that affected decision formation. However, the neural mechanisms facilitating these behavioural enhancements remain less well understood. Discrepancies persist between such studies that sought to identify when perceptual improvements emerge in the brain during decision formation, with temporally localised effects of associative congruency reported at different early and late onsets of the perceptual decision formation process (Bizley, Jones, & Town, 2016). Hence, it is not clear whether such improvements represent ‘early’ sensory processing benefits, or ‘late’ post-sensory changes in decision dynamics. As such, we are yet to be provided with 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, as well as associative congruency, predominantly stems from the utilisation of experimental paradigms that present two or more unisensory features simultaneously, or in close spatial and/or temporal proximity. Previous research has associated multiple neural processes with the observed decision-making benefits as a result of multisensory information available, in particular (i) multisensory integration; integrating information across sensory modalities into unified percepts with significantly different bimodal neural responses (Angelaki et al., 2009; Calvert et al., 2004; Mercier & 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 & 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 recorded will display mixed selectivity representing a variety of sensory, decision-related, and task-relevant signals facilitated by multisensory paradigms

(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 or late post-sensory changes to decision dynamics during the formation of perceptual decisions. Thus, we cannot directly dissociate where and when perceptual improvements from associative congruency emerge in the brain during perceptual decision formation.

In this study, we sought to capitalise secondary analyses of a dataset that exploited the novelty of using a modified variant of the Implicit Association Test (IAT), demonstrated by Parise & Spence (2012), to formulate 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 using a modified variant of the IAT, coupled with EEG, we can record the neural activity underlying formulated auditory pitch-visual size associations, which is less likely to be affected by confounding activity, and therefore mixed selectivity, 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 analysed single-trial EEG activity using multivariate Linear Discriminant Analysis (LDA; Parra et al., 2002; 2005; Sajda et al., 2009; Philiastides & Sajda, 2006a; 2006b; Philiastides, Ratcliff, & Sajda, 2006; Philiastides et al., 2014). Finally, to dissect the constituent processes underlying the effects of pitch-driven associations on perceptual decision formation, we adopted a neurally-informed cognitive modelling approach, that fits computational models to both behavioural and neuroimaging data (Turner et al., 2013; 2016; 2017; Franzen et al., 2020; Delis et al., 2018; Diaz et al., 2017; Kayser & Shams, 2015). Thus, 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 & Cappe, 2020).

In the present study, we employ 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. Thus, we aim to ultimately characterise the mechanistic effect of cross-modal associations on perceptual decision formation.

2.3 Materials and Methodology[3]

2.3.1 Participants

20 participants (male = 7, female = 13; age range = 19-32) were recruited using the University of Glasgow Subject Pool and received £6/hour (UK Sterling) for their participation. All participants were right-handed, reported normal hearing, normal or corrected-to-normal vision, and no history of neurological conditions. The study was approved by the ethics committee of the College of Science and Engineering at the University of Glasgow (CSE 300130001) and was conducted in accordance with the Declaration of Helsinki (World Medical Association, 2013).

---

[3] The study was undertaken at the Institute of Neuroscience and Psychology at the University of Glasgow by Dr Stephanie Boyle and Dr Robin A.A. Ince. Subsections 2.3.1 *Participants*, 2.3.2 *Stimuli*, 2.3.2 *Implicit Association Test*, 2.3.4 *Procedure*, and 2.3.6 *EEG Recording and Preprocessing* outline contents of the experiment and the datasets collected. All secondary analyses was conducted at the School of Biomedical Sciences at the University of Leeds. For further information, see page ii *Intellectual Property and Publication Statements*.

2.3.2 Stimuli

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

2.3.3 Implicit Association Test

This study used a modified variant of the IAT, adapted from Parise & Spence (2012), to formulate auditory pitch-visual size cross-modal associations (Figure 2.1a). In this variant, on each block, one auditory (high-pitch tone/low-pitch tone) and one visual (small-size circle/large-size circle) stimulus feature were assigned to one of two response keys (left/right response keys). Participants were 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 (Figure 2.1a, *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 (Figure 2.1a, *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 & Spence, 2006; Evans & Treisman, 2010; Parise & 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 hours (~2 hours for EEG setup/cleanup, ~1.5 hours 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 stimulus features 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 previous studies (Parise & Spence, 2012; Roque, Lafraire, & Auvray, 2020).

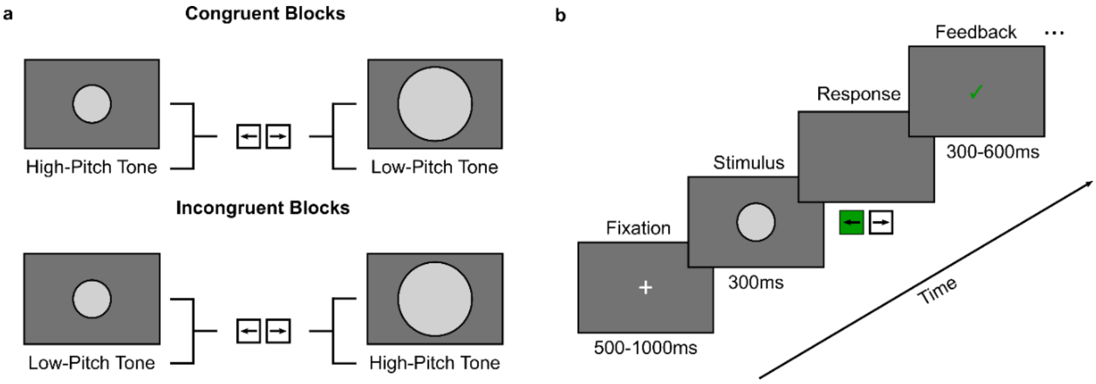


Figure 2.1 Implicit Association Test. a, Participants were presented with one unisensory stimulus feature (auditory high/low-pitch tone; visual small/large-size circle) per trial, and asked to categorise which stimulus feature, within that modality, was presented as quickly and as accurately as possible using the correctly assigned response key (left/right response key). Auditory congruency was manipulated by switching the stimulus feature-response key mappings across blocks (*top*, congruent block mappings; *bottom*, incongruent block mappings). **b,** Each trial started with a fixation cross presented centrally on-screen for a randomized 500-100ms period, before one of four stimulus features (see *Stimuli* section) were selected randomly and presented for 300ms. Feedback was provided after each response for a randomised 300-600ms period.

2.3.4 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 2.2.3 *Implicit Association Test* section). Participants were given as much time as they needed to memorize the instructions for the upcoming block. Figure 2.1b illustrates the procedure on a single-trial basis. Each trial started with a fixation cross presented centrally on-screen for a randomised period (uniform distribution from 500 to 1000ms). Then, one of the four stimuli (see 2.3.2 *Stimuli* section) were selected randomly and presented for 300ms. Participants were instructed to categorize, as quickly and as accurately as possible, the presented stimulus feature using the left and right keyboard response keys, as defined by the instructions given for that specific block (see 2.3.3 *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 300ms to 600ms). In total, participants completed 8 blocks (4 blocks each for the congruency of stimulus feature-response key mappings presented in a random order) for a total of 1280 trials (160 trials per block; 40 trials for each stimulus feature).

2.3.5 Statistical Analysis of Behavioural Data

For each participant, median RTs (calculated in milliseconds; ms) and choice accuracy (calculated as the proportion of correct choices over all trials) were used as dependent variable measurements of behavioural performance. These were calculated separately for two independent variables: (1) stimulus feature (Auditory: High-Pitch Tone/Low-Pitch Tone; Visual: Small-Size Circle/Large-Size Circle), and (2) 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.

For separate auditory and visual stimuli, trials with RTs more or less than the median RT +/- 2.5 Median Absolute Deviations (MADs) were then excluded from further analyses, with these RTs attributed to outliers corresponding to “fast guesses” or attentional lapses during testing (Whelan, 2008). This pre-processing criterion was selected as previous research has demonstrated that MADs are a more robust measurement of central dispersion than standard deviation (Leys et al., 2013). Overall, 1442 trials (Auditory: 830 trials; Visual 612 trials) were excluded from further analysis, leaving a total of 9697 trials for auditory stimuli and 9939 trials for visual stimuli. Next, an Anderson-Darling test was first used to assess if RTs were normally distributed (Anderson & Darling, 1954; Nelson, 1998; Stephens, 1974). This is because it is commonly observed in research that RT distributions tend to be not normally distributed, with histograms prominently illustrating a positive right skew (Marmolejo-Ramos et al., 2015; Whelan, 2008).

As expected, the assumption of normality was found to be violated for RTs over the full pre-processed behavioural dataset ($A^2 = 245.39$, $p < 0.001$). 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 feature (high-pitch tone/low-pitch tone) and congruency (congruent/incongruent) respectively (Rosenthal, Cooper, & Hedges, 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 tone/low-pitch tone). For Mann-Whitney U Paired testing analysing RTs of correct and incorrect choices across congruency and 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 stimulus feature conditions This enabled us to analyse the effects of both congruent/incongruent stimulus feature-response key mapping conditions, and high-pitch tone/low-pitch 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 from both behavioural analyses 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 (R Core Team, 2022).

2.3.6 EEG Recording and Pre-processing

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 seconds, from -0.5 to 1.5 seconds 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 & Siegel, 2013; Keren, Yuval, Greenberg, & Deouell, 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.

2.3.7 EEG Signal Analysis – Linear Discriminant Analysis

We applied single-trial multivariate Linear Discriminant Analysis (LDA; Parra et al., 2002; 2005; Philiastides & Sajda, 2006a; 2006b; Philiastides, Ratcliff, & Sajda, 2006; Philiastides et al., 2014; Sajda et al., 2009) to extract EEG components discriminating between (a) congruent and incongruent trials, and (b) high-pitch tone and low-pitch tone 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 congruency and stimulus feature conditions for auditory stimulus features only. All time windows had a width of 50ms, with the window centre t shifted from -100ms to 800ms, relative to auditory stimulus-onset, in 5ms 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 congruency and stimulus feature 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. In discriminating the two stimulus feature categories, the classifier maps negative and positive discriminant component amplitudes to high-pitch tone and low-pitch tone 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. Similarly, larger negative values indicate a higher likelihood of categorizing high-pitch tones, and larger positive values indicate a higher likelihood of categorizing low-pitch tones, with values near zero reflecting less discriminative component amplitudes.

We quantified classification performance of our classifiers for each time window using the area under a receiver operating characteristic (ROC) curve (Green & Swets, 1966), referred to as an A_z value, using a leave-one-out cross-validation procedure (Gherman & Philiastides, 2015; Philiastides & Sajda, 2006a; 2006b). To determine group significance thresholds for discriminator performance, we implemented a permutation test, whereby congruency and stimulus feature 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 equation (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 5ms 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).

2.3.8 HDDM – 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 DDM (Forstmann, Ratcliff, & Wagenamkers, 2016; Ratcliff et al., 2015; 2016; Ratcliff & 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*).

2.3.9 HDDM – 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. HDDM uses Markov-Chain Monte Carlo (MCMC) 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 & 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, 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

& 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 & 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 (Diaz et al., 2017; Delis et al., 2018, Frank et al., 2015; Franzen et al., 2020; Mercier & Cappe, 2020; Tremel & 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 correspond 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 & Childers, 2015). For each model, we ran 5 separate Markov chains with 11000 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 25000 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 & 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.

2.3.10 HDDM – 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 (see section 2.4.1 *Behavioural Results*), 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:

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

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

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

where $\left|y_{early}^{max}\right|$ and $\left|y_{late}^{max}\right|$ 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 ~ 110ms; Late ~ 340ms; see Figure 2.3b). Coefficients $\alpha_1, \beta_1, \gamma_1$ and $\alpha_2, \beta_2, \gamma_2$ weight the slope of each parameter by the absolute values of $\left|y_{early}^{max}\right|$ and $\left|y_{late}^{max}\right|$ respectively, with intercepts $\alpha_0, \beta_0, \gamma_0$, 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 $\left|y_{early}^{max}\right|$ and $|y_{late}^{max}|$ values, and incongruent trials were predominantly categorised by positive $\left|y_{early}^{max}\right|$ and $|y_{late}^{max}|$ values respectively (see 2.3.7 *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 2.4.1 *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., 2021). 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 towards 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 predictive effects of our regression coefficients on changes to estimated posterior parameters favoured by our hypotheses.

2.4 Results

2.4.1 Behavioural Results

2.4.1.1 Reaction Time

Participants responded faster in auditory trials with congruent compared to incongruent stimulus feature-response key mappings (Figure 2.2a, Congruent: median = 624ms post-stimulus offset; Incongruent: 657ms post-stimulus offset). Wilcoxon Signed-Rank Testing determined this finding to be statistically significant (Z = -2.135, p = 0.033, effect size = -0.447, Wilcoxon Signed-Rank Testing). This result held for both correct (Figure 2.2c, Congruent/Correct: median = 608ms post-stimulus offset; Incongruent/Correct: median = 639ms post-stimulus offset, Z = -6.333, p < 0.001, effect size = -0.090, Mann-Whitney U testing) and incorrect trials separately (Figure 2.2c, Congruent/Incorrect: median = 573ms post-stimulus offset; Incongruent/Incorrect: median = 617ms post-stimulus offset, Z = -2.890, p = 0.004, effect size = -0.092, Mann-Whitney U testing). Furthermore, RTs were significantly longer for correct compared to incorrect responses for congruent stimulus feature-response key mappings (Figure 2.2c, Correct: median = 608ms post-stimulus offset; Incorrect: median = 573ms post-stimulus offset, Z = -2.793, p = 0.005, effect size = -0.030, Mann-Whitney U Testing), but not for incongruent stimulus feature-response key mappings (Figure 2.2c, Correct: median = 639ms post-stimulus offset; Incorrect: median = 617ms post-stimulus offset, Z = -1.159, p = 0.247, effect size = -0.017, Mann-Whitney U Testing). We found no significant effect of stimulus feature on median RTs (Figure 2.2d, High Pitch Tone: median = 632ms; Low Pitch Tone: median = 639ms, Z = -0.788, p = 0.430, effect size = -0.176, Wilcoxon Signed-Rank Testing). There was also no significant effect when testing correct (Figure 2.2f, High Pitch Tone/Correct: median = 622ms post-stimulus offset; Low Pitch Tone/Correct = 621ms post-stimulus offset, Z = -0.146, p = 0.884, effect size = 0.001, Mann-Whitney U Testing) or incorrect trials separately (Figure 2.2f, High Pitch Tone/Incorrect: median = 592ms post-stimulus offset; Low Pitch Tone/Incorrect: median = 589ms post-stimulus offset, Z = -0.719, p = 0.472, effect size = -0.025, Mann-Whitney U Testing). Furthermore, we found no significant difference in RT between correct and incorrect responses for either high-pitch tones (Figure 3.2f, Correct: median = 622ms post-stimulus offset; Incorrect: median = 592ms post-stimulus offset, Z = -1.443, p = 0.149, effect size = -0.021, Mann Whitney-U Testing), or low-pitch tones (Figure 2.2f, Correct: median = 621ms post-stimulus offset; Incorrect: median = 589ms post-stimulus offset, Z = -1.627, p = 0.052, effect size = -0.023, Mann Whitney-U Testing).

We found no significant effect of associative congruency on median RTs for visual stimuli (Congruent: median = 579ms post-stimulus offset; Incongruent: median = 608ms post-stimulus offset; Z = 1.199, p = 0.231, effect size = -0.268, Wilcoxon Signed-Rank Testing). We further found no significant effect of visual stimulus feature on median RTs (Small-Size Circle: median = 619ms post-stimulus offset; Large-Size Circle: median = 599ms post-stimulus offset, Z = -0.974, p = 0.330, effect size = -0.218, Wilcoxon Signed-Rank Testing).

2.4.1.2 Choice Accuracy

Regarding choice accuracy, participants had a slightly, but not significantly, higher proportion of correct responses for auditory trials with congruent compared to incongruent stimulus feature-response key mappings (Figure 2.2b, 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 (Figure 2.2e, High Pitch Tone: proportion correct = 0.915; Low Pitch Tone: proportion correct = 0.918, Z = -0.055, p = 0.956, effect size = -0.012, Wilcoxon Signed-Rank Testing).

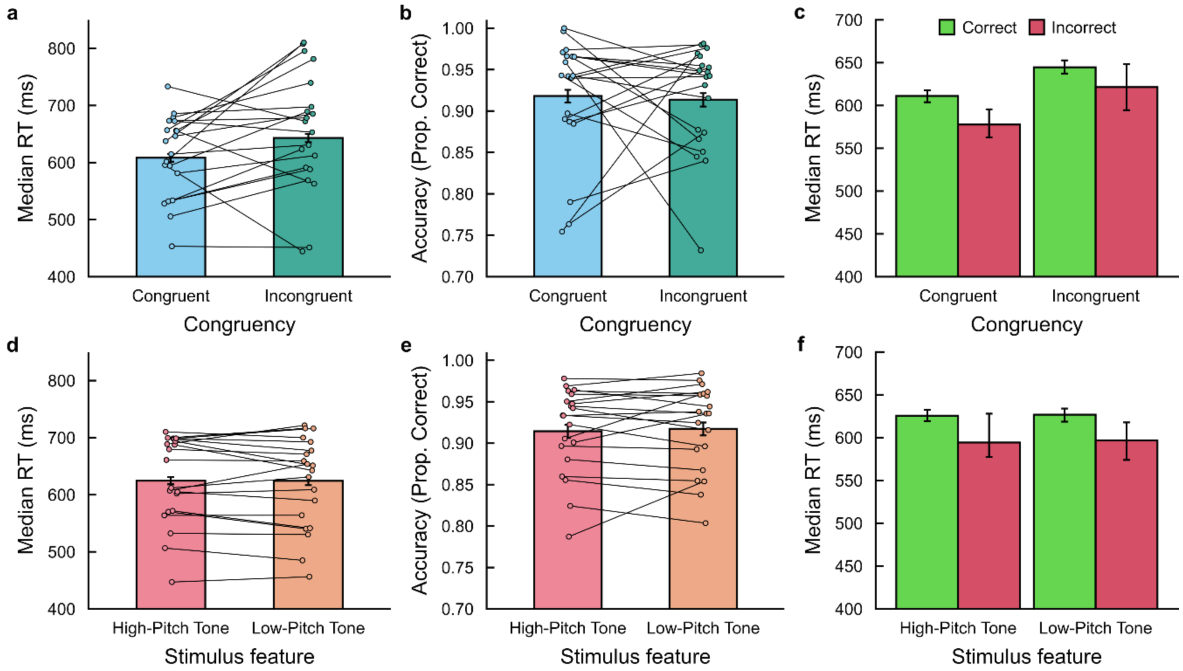


Figure 2.2 Behavioural Performance. *Left.* **a, d,** Median RTs and **b, e,** choice accuracy (proportion of correct responses) for condition (bars) and participants (scatter points) for **a**, **b**, Congruency (congruent/incongruent) and **d**, **e**, Stimulus Feature (High Pitch Tones/Low Pitch tones). *Right.* Median RTs for correct and incorrect RT for **c**, Congruency (congruent/incongruent) and **f**, Stimulus Feature (High-Pitch Tones/Low-Pitch Tones). For all graphs, 95% Confidence Intervals (CIs) were computed using 1000 bootstrapping random sampling iterations to estimate the distribution of average performance measurements. Note that behavioural performance here was quantified with an initial preprocessing criteria excluding all RTs between 300ms and 1200ms, yielding similar median RTs for congruency and stimulus feature outcomes.

In addition, 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 (Congruent: proportion correct = 0.958; Incongruent: proportion correct = 0.957, Z = -0.423, p-value = 0.673, effect size = -0.094. Wilcoxon Signed-Rank Testing). There was also no significant effect of visual stimulus feature on choice accuracy (Small-Size Circle: proportion correct = 0.956; Large-Size Circle: proportion correct = 0.960; Z = -0.905, p-value = 0.365, effect-size = -0.202, Wilcoxon Signed-Rank Testing).

2.4.1.3 Interim Summary of Behavioural Results

To summarise, 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.

2.4.2 EEG Signal Analysis Results

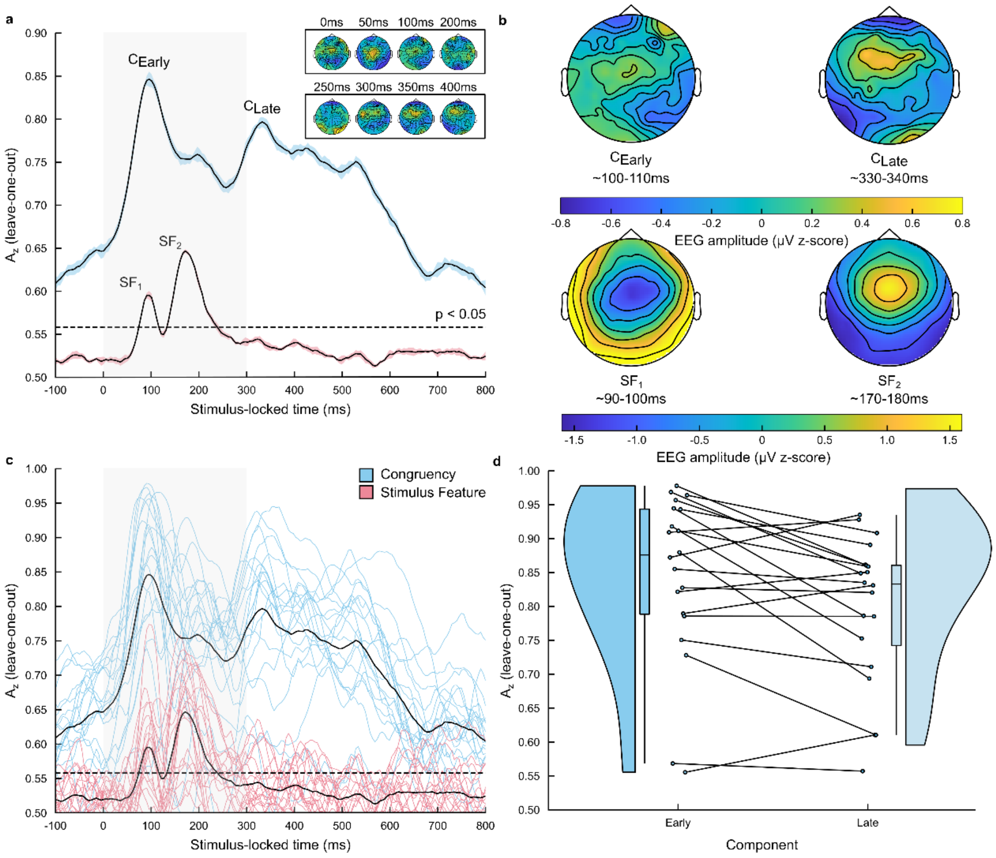


Figure 2.3 Multivariate Linear Discriminant Analysis Group Results. a, Mean multivariate discriminator performance (A_z), quantified by leave-one-out trial cross-validation procedure, during outcome discrimination of stimulus-locked EEG responses, as a function of congruency (congruent-vs-incongruent; blue) and stimulus feature (high-pitch tones-vs-low-pitch tones; red) 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 0ms (post-stimulus onset) to 300ms (post-stimulus offset). **a,** *Top Right*. Scalp topographies at representative time windows corresponding to EEG

components from 0-400ms in 50ms time increments. **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. **c,** Participants' mean discriminator performance (A_z) obtained from a leave-one-out cross-validation procedure, during stimulus feature (red) and congruency (blue) 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 polygons for **a, c,** denote the presentation of auditory stimuli, from 0ms (post-stimulus onset) to 300ms (post-stimulus offset). **d,** Participants' mean discriminator performance (A_z) for the Early and Late congruency-discriminating EEG components.

Next, we analyzed 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 50ms, 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 2.3.7 *EEG Signal Analysis - Linear Discriminant Analysis* section). 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 respectively, 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., AUC-ROC; 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, leading 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 0-600ms, above the significance level estimated from our permutation test. Specifically, discriminator performance within this range was characterized by two temporally specific components (Figures 2.3a and 2.3b; C_{Early}: mean peak time = 100-110ms, A_z value = 0.846; C_{Late}: mean peak time = 330-340ms, Az value = 0.797). These components were consistent across participants (see Figure 3.4c for the A_z curves and Figure 3.3d for the maximum A_z values of each participant). We then computed the corresponding scalp topographies, obtained using the forward model, correlating between peak discriminant output and EEG data (Figure 3.3b, averaged over a 50ms time window centered 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 which presented either 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 (Figure 2.3b; SF_1: mean peak time = 90-100ms, A_z value = 0.595; SF_2: mean peak time = 170-180ms, A_z 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.

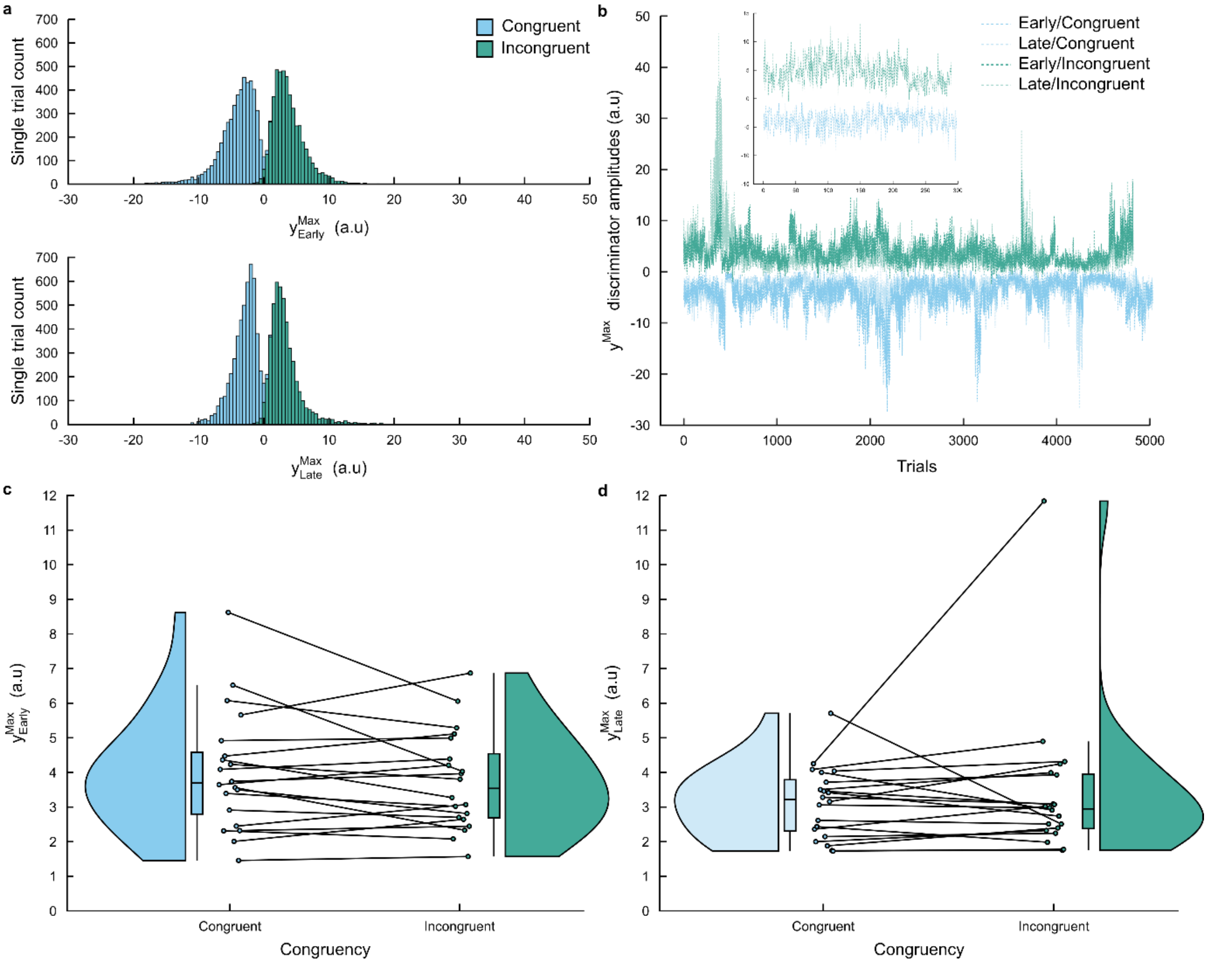


Figure 2.4 Multivariate Discriminator Amplitude Participant Results. a, Single-trial discriminator amplitudes (y) for Congruent and Incongruent component amplitudes illustrated as histograms for the *Early* (*top*) and *Late* (*bottom*) EEG components respectively. Negative values indicate neural evidence for associative congruency whereas positive values indicate neural evidence for associative incongruency. **b,** Single-trial discriminator amplitudes (y) for *Early* and *Late* component windows for Congruent and Incongruent trials. *Above*. Single-trial discriminator amplitudes from a representative subject. **c, d,** Absolute values of our single-trial discriminator amplitudes (y)s for Congruent and Incongruent component amplitudes for the *Early* (**c**) and *Late* (**d**) components.

The covariation between the strength of the A_z values for *Early* (i.e., C_{Early}) and *Late* (i.e., C_{Late}) components was then assessed using a covariation matrix. The variance within the Early and Late components respectively was consistently low (*Early*: 0.015; *Late*: 0.012), as was the covariance across N = 20 participants (0.011). The low variance demonstrates a consistency across participants to display peak component activity for temporally distinct neural processes. Furthermore, a positive covariance value indicates that two variables tend to simultaneously increase or decrease, suggesting that the strength of the *Early* component predicts the strength of the *Late* component, albeit weakly. This initial covariance analysis implies that the onset of the *Early* component may partially influence the onset of the *Late* component. In addition, partial correlations were applied to assess the degree of association between *Early* (i.e., C_{Early}) and *Late* (i.e., C_{Late}) components assuming the effect of the peak onset of the stimulus feature components (i.e., SF_1 and/or SF_2) was controlled. Therefore, we could observe if there is an influence of the peaks for discriminating auditory stimulus feature (i.e., high-pitch tones versus low-pitch tones) on the onset of the peaks for discriminating associative congruency. A partial correlation was run to determine the relationship between a participants' C_{Early} and C_{Late} component peaks whilst controlling for SF_1 and SF_2 component peaks. There was a moderate positive partial correlation between C_{Early} and C_{Late} component peaks whilst controlling for SF_1 and SF_2 component peaks, which was statistically significant (r(16) = 0.746, N = 20, p < 0.001). However, zero-order correlations (i.e., excluding for controlling of SF_1 and SF_2 component peaks) showed that there was a statistically significant, moderate, positive correlation between C_{Early} and C_{Late} component peaks (r(18) = 0.791, N = 20, p < 0..001), indicating that SF_1 and SF_2 component peaks had very little influence in controlling for the relationship between C_{Early} and C_{Late} component peaks. We reaffirmed these findings by then assessing the relationship between an individual's C_{Early} and C_{Late} component peaks whilst controlling for SF_1 or SF_2 component peaks separately. We found moderate positive partial correlations between C_{Early} and C_{Late} component peaks whilst controlling for SF_1 and SF2 component peaks separately, both of which were statistically significant (SF_1: r(17) = 0.725, N = 20, p < 0.001; SF_2: r(17) = 0.788, N = 20, p < 0.001). Zero-order correlations confirmed that this significant moderate positive correlation between C_{Early} and C_{Late} component peaks resided, indicating that neither stimulus feature component had a profound influence in controlling for the relationship between C_{Early} and C_{Late} component peaks (SF_1: r(18) = 0.791, N = 20, p < 0.001; SF_2: r(18) = 0.791, N = 20, p < 0.001). Together, these findings imply that the auditory stimulus feature encoded in the associative congruency of stimulus feature-response key mappings did not profoundly influence the variance nor the correlation between the temporal peaks of our identified components reflecting *Early* and *Late* mechanisms underlying the benefits of associative congruency towards perceptual decision formation.

2.4.3 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 versus incongruent trials. To achieve this, we used a neurally-informed variant of the HDDM (Wiecki et al., 2013, see Figure 2.5a for a graphical illustration and sections 2.3.8 *HDDM – Description,* 2.3.9 *HDDM – Fitting, and* 2.2.10 *HDDM – EEG Regressors* for modelling details). As previously mentioned, the HDDM is a Bayesian implementation of the well-known DDM, used for characterizing perceptual decision formation in 2AFC paradigms (Ratcliff & McKoon, 2008).

We extracted the maximum single-trial discriminator amplitudes ($\left|y_{early}^{max}\right|$ and $|y_{late}^{max}|$) 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 Figure 2.4a for histograms of y_{early}^{max} and y_{late}^{max} 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 points 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 $\alpha_1, \beta_1, \gamma_1$ and $\alpha_2, \beta_2, \gamma_2$ for $\left|y_{early}^{max}\right|$ and $|y_{late}^{max}|$ 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 Figures 2.4c and 2.4d for the average $|y_{early}^{max}|$ and $|y_{late}^{max}|$ of each participant).

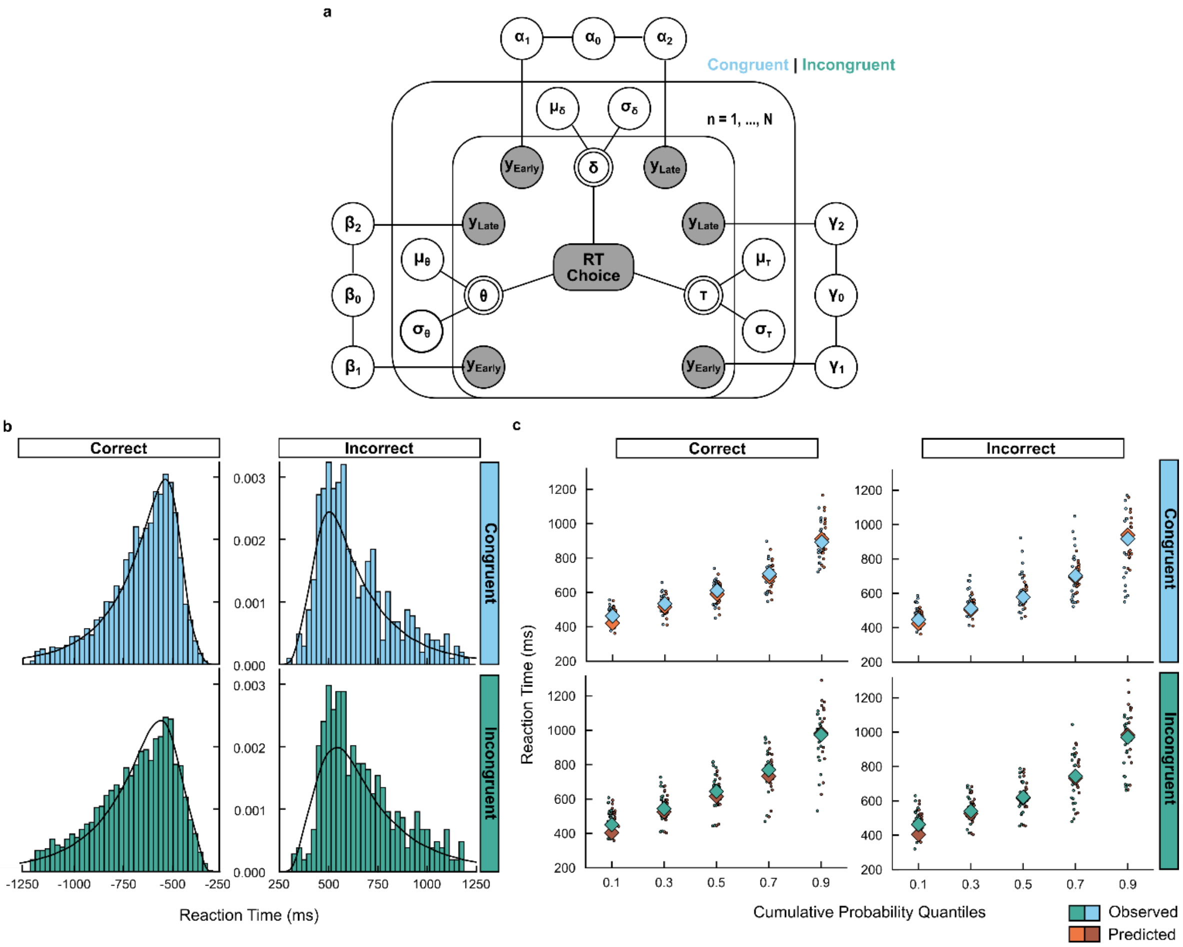


Figure 2.5 Neurally-Informed Cognitive Modelling – Posterior Predictive Checks. a, Graphical representation illustrating the Bayesian hierarchical framework for estimating neurally-informed HDDM parameters. Round nodes represent continuous random variables, and shaded nodes represent recorded or computed signals, i.e., single-trial behavioural data (RTs and Choice) and EEG component discriminator amplitudes (ys). 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 , …, N) and the outer plate is over congruency conditions (Congruent | Incongruent). Parameters are modeled 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, c,** Posterior predictive checks of the Neurally-informed HDDM fitting to participant and group behavioural data. Modelling fit to behavioural data was assessed using **b,** histogram and neurally-informed HDDM model fits for RT distributions split across congruency conditions, and **c,** 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, i.e. single-trial behavioural data (RTs), and predicted RTs, i.e. simulated RTs from HDDM posterior predictive estimates. Diamonds represent group averages and circles represent single-participant values.

We found a good fit of the behavioural data (i.e., choice accuracy and RTs) from our proposed neurally-informed HDDM (Figures 2.5b and 2.5c). 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. Late: $P(\gamma_2^{Congruent}$ < 0) = 0.62, log-odds = -2.712; $P(\gamma_2^{Incongruent}$ > 0) = 0.936; log odds = 2.690; Figure 2.6c). 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 (approximately 100-110ms post-stimulus), 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 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 = -1.186; Figure 3.6b). 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 (Figure 2.6b).

Finally, we did not find prevalent evidence that single-trial amplitudes of either the *Early* or *Late* component were predictive of increases in drift rate parameter estimates for congruent or incongruent trials, since they did not exceed the hypothetical log-odds proportion of our false positive rate suggesting predictive effects (Early: P($\alpha_1^{Congruent}$ > 0) = 0.930, log-odds = 2.587; P($\alpha_1^{Incongruent}$ > 0) = 0.387, log-odds = -0.461. Late: $P(\alpha_2^{Congruent}$ > 0) = 0.0.852, log-odds = 1.754; P($\alpha_2^{Incongruent}$ > 0) = 0.0.853, log-odds = -1.757; Figure 2.6a). This result argues against our hypotheses implicating higher sensory evidence in congruent compared to incongruent trials for either the *Early* and *Late* component, and does not mechanistically characterize shorter RTs in congruent compared to incongruent trials.

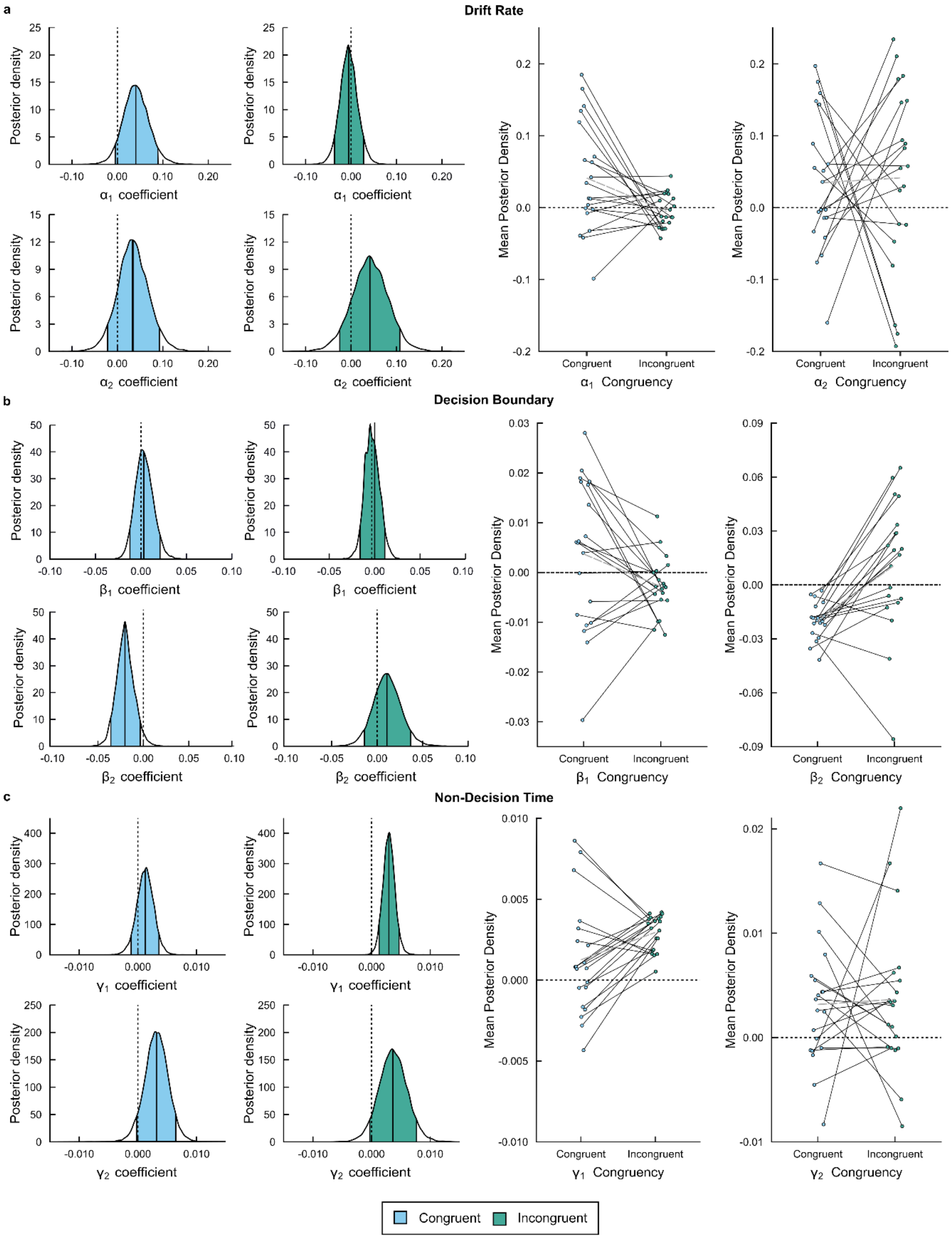


Figure 2.6 Neurally-Informed Cognitive Modelling - Results. Group-level (*left*) and participant-level (*right*) posterior density distributions of estimated regression coefficients for **a,** drift rate (α's), **b,** decision boundary (β's) and **c,** non-decision time (γ's) for *Early* (*top*) and *Late* (*bottom*) EEG component discriminator amplitudes. For group-level posterior density distributions, all regression coefficients are derived from the neurally-informed HDDM, including N = 20 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. For participant-level posterior distributions, points denote the mean point estimates for each subject, and the dashed line denotes the zero point.

2.5 Discussion

In this work, we used single-trial multivariate LDA 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 modified variant of the Implicit Association Test (Parise & Spence, 2012), we showed significant behavioural improvements as a result of associative congruency, whereby participants responded faster to congruent than incongruent stimulus feature-response key mappings (Figure 2.2a). Additionally, we showed that participants were slower for correct compared to incorrect responses for congruent stimulus feature-response key mappings only. Our multivariate discriminant analysis on the underlying EEG signals revealed neural information discriminating the congruency of stimulus feature-response key mappings predominantly within a 0-600ms post-stimulus onset window, including the 300ms period for stimulus feature presentations and (approximately) after which for perceptual decision formation up to median choice responses. Within this 0-600ms time window, we characterised two temporally-distinct EEG components carrying congruency-relevant information on a single-trial basis: first, an *Early* component, peaking at approximately 100-110ms post-stimulus onset, and a *Late* component, peaking at approximately 330-340ms post-stimulus onset. Using neurally-informed cognitive modelling, we linked these prominent neural correlates of associative congruency (for auditory pitch stimulus feature-response key mappings) with the corresponding behavioural observations for forming perceptual decisions. We thus associated the observed shorter RTs in congruent trials with (a) an increase in the duration for processing and encoding sensory information modulated by the *Early* component during incongruent trials, and (b) a decrease in the quantity of post-sensory evidence that needs to be consolidated and accumulated to facilitate a perceptual choice modulated by the *Late* component in congruent trials, albeit not to the detriment of either choice accuracy or slowing of responses for congruent trials.

Our behavioural results further consolidate the existence of auditory pitch-visual size cross-modal associations that have been reported in previous research (Bien et al., 2012; Brunetti et al., 2017; Evans & Treisman, 2010; Gallace & Spence, 2006; Marks, Ben-Artzi, & Lakatos, 2003; Parise & Spence, 2008; 2009; 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 & Spence (2012), who reported faster RTs for congruent compared to incongruent trials for five auditory-visual stimulus combinations, including frequency-pitch and object-size. Importantly, as the IAT only presents one unisensory stimulus feature per trial, results in both auditory and visual modalities were equally relevant in forming perceptual decisions, and no explicit instructions were given to participants to preferentially map frequency-pitch and object-size, our results can be interpreted as evidence of a more genuine effect of stimulus feature compatibility in benefitting perceptual decision formation, and not from general multisensory benefits arising from simultaneously presented stimulus features, notably multisensory integration (Angelaki et al., 2009; Calvert et al., 2004; Mercier & Cappe, 2020) and/or a form of selective attention/attention-dividing (Bien et al., 2012; Choi et al., 2018; Gallace & Spence, 2006; Marks., 2004).

Whether the nature of associative congruency towards benefitting perceptual decision formation is absolute (i.e., automatic) or relative remains a persistent debate in cross-modal correspondence research (Moors & De Houwer, 2006; Spence, 2019). Traditionally, the prevalent generalization was that congruent cross-modal associations are inherently automatic in benefitting rapid perceptual decision formation, with such benefits arising due to pure bottom-up perceptual processing with specific one-to-one cross-modal mapping between particular stimulus features across multiple sensory modalities, for example, a particular pitch frequency and size range of an audiovisual object. However, this has been questioned in recent studies. Across audiovisual cross-modal associative research, contradictory findings persist concerning the replicability of the benefits of associative congruency towards decision formation, with advantages in speeded classification paradigms not always exclusively reported (e.g., auditory pitch-visual spatial location; Heron et al., 2012; auditory pitch-visual elevation; Chiou & Rich, 2012; auditory/tactile amplitude frequency-visual spatial frequency; Orchard-Mills; Alais, & Van der Burg, 2013). In addition, such findings further demonstrate that the benefits of associative congruency are not unsusceptible to strategic top-down cognitive modulations towards multimodal perceptual decision formation, notably the amodal properties of the presented stimuli, the modality characteristics determining how incoming features are consolidated, and processes determining how the integration of features underlies the goal-orientated behaviours of the observer itself. In consideration (see Welch & Warren, 1980; Chen & Spence, 2017, for reviews), it should be emphasized that the benefits of associative congruency observed in our study should be considered relative in nature, and not inherently automatic in enhancing perceptual decision formation.

Specifically, it is the variation within blocks of trials, and subsequent trial-by-trial contrasts between high frequency and low frequency pitch tones, which influences behavioural performance, and not necessarily the absolute pitch frequency of the auditory tones presented. Our results complement findings from Brunetti et al. (2017; see section 2.2 *Introduction*), as well as from Getz & Kubovy (2018), who utilised a novel paradigm whereby they manipulated the associative congruency and associative compatibility of cross-modal associations between auditory pitch and five visual features (object angularity, brightness, height, size, and spatial frequency). Manipulations of associative compatibility were defined according to whether participants were explicitly instructed to pair congruent audiovisual or incongruent audiovisual endpoints. Their findings implicated a consistent advantage of associative congruency across all five pairings of decreased RTs for correct decisions (replicating findings from both explicit and implicit speeded classification paradigms; see section 2.2 *Introduction*), but which were negated (albeit subject to the audiovisual cross-modal association under investigation) when associative compatibility explicitly instructed participants to map non-preferential audiovisual stimulus features, demonstrating evidence that the benefits of associative congruency towards perceptual decision formation involve both bottom-up and top-down processing, and the degree of which both are involved modulates the consequential choice benefits. Given in our paradigm participants were only instructed as to which response key should be selected to categorise the presented auditory or visual stimulus feature (without explicit instructions to map compatible auditory-pitch visual-size cross-modal associations according to research consensus), and our participant-level behavioural, neuroimaging, and modelling results illustrate not all participants were susceptible to benefits from associative congruency from one stimulus feature presentation per trial (see sections 2.4.1 *Behavioural Results*, 2.4.2 *EEG Signal Analysis Results*, and 2.4.3 *Neurally-Informed Cognitive Modelling Results*), we recommend operationalising manipulations to this paradigm that can boost top-down processing influences for future research in order to dissociate whether unisensory-formulated cross-modal associations are biased towards one end of this absolute-relative associative benefits spectrum. Comparably, analysing the effects of statistical learning of transitional probabilities within unisensory-formulated cross-modal associations could be assessed to ascertain whether participants begin to exploit the benefits of associative congruency across modalities (Glickson & Cohen, 2013). In particular, statistical learning could be enhanced or reduced, depending on if decision-making performance benefits between and within blocks of different associative congruency levels when mapping unisensory-led stimulus feature-response key mappings. Future research should consider this to determine if the decision-making benefits of associative congruency begin to emerge (i.e., *relative* cross-modal associative benefits), or remain stable (i.e., *automatic* cross-modal associative benefits) during IAT participation.

We further provide neuroimaging evidence demonstrating a robust modulation to neural activity by the associative congruency of auditory-driven stimulus feature-response key mappings. Most importantly, as the IAT only presents one unisensory stimulus feature per trial, the resultant modulations are minimised from confounding neural activity attributed to further multisensory decision-making mechanisms, notably multisensory integration (Franzen et al., 2020; Mercier & Cappe, 2020), and a form of selective-attention/attention-dividing (Bien et al., 2012; Marks, 2004) between stimulus features presented simultaneously, or within close spatial and/or temporal proximity. To examine neural activity specifically related to the behavioural benefits of auditory stimulus-driven cross-modal associative congruency, we applied multivariate LDA to decode congruent versus incongruent stimulus feature-response key mappings. The application of multivariate LDA 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-lateralised centroparietal, and left-lateralized occipital electrodes, with the consequential spatiotemporal trajectory between the components characterised by a gradual emergence toward fronto-central regions of the brain from 0-400ms post-stimulus onset.

Crucially, the first component (C_{Early}: approximately 100-110ms post-stimulus onset) arises near simultaneously with the prominent components for encoding auditory stimuli (i.e. SF_1 and SF_2: ~90-100ms and ~170-180ms post-stimulus onset respectively), whereby the early latency onset of the discrimination of auditory stimulus-driven associative congruency, reflecting higher neural evidence between congruent versus incongruent mappings, coincides with our results revealing an increase in discrimination of the frequency of the presented auditory stimulus feature itself. This implicates an overlapping mapping of perceptual priors of auditory stimulus-driven pitch-size associations that systematically influences early sensory encoding/processing. Based on our observations of the behavioural results (see section 2.4.1 *Behavioural Results*), we contend that the benefits of associative congruency modulate neural activity due to a form of perceptual feedback, or trial-by-trial perceptual recalibration (Boyle, Kayser, & Kayser, 2017; Kayser & Kayser, 2018; Park & Kayser, 2019). This results in an inducing of mappings through repeated exposure (i.e., on a trial-to-trial basis) between preferential pitch frequency and object size within the early processing of sensory information across multiple modalities during the perceptual decision formation process.

Previous research has demonstrated that repeated exposure to complementary stimulus features shapes their multisensory composition, thus informing implicit preferences to congruent mappings (Habets et al., 2017). For example, Kayser & Kayser (2018) used an audiovisual flanker paradigm and EEG to investigate current and serial trial-by-trial interactions of associative congruency in discriminating visual motion coherence and the resultant modulations in underlying neural activity. They found that participants elicited higher choice accuracy and faster RTs when an accompanying acoustic stimulus mimicked motion in the same coherent direction of a visual random dot display as opposed to opposite mimicked motion. Intriguingly, they found clustered Region of Interest (ROI) modulations of the effect of current-trial associative congruency over sensory specific areas of the occipital lobe at a latency of approximately 300ms, followed by clustered ROI modulations of the effect of serial trial-by-trial associative congruency over right ventral frontal lobe. Together, their results demonstrated that neural representations of preferential audiovisual mappings are serially dependent on perceived associative congruency formulated from previous experience, whereby neural representations of encoded multisensory information in modality specific regions on the brain can be hierarchically modulated to benefit perceptual decision formation independently from paradigm demands, which consolidate the formulated decision itself.

Similarly, Park & Kayser (2019) investigated the neural mechanisms underlying multisensory integration and trial-by-trial perceptual recalibration effects using MEG. Using an audiovisual ventriloquist paradigm and single-trial multivariate classification analysis, they found that participants were susceptible to the ventriloquist effect (VE) and ventriloquist after-effect (AVE) whereby a congruent visual stimulus biased the perceived location of its auditory stimulus counterpart, demonstrating the decisional influence of associative congruency through audiovisual spatial discrepancy. Underpinning these behavioural findings were significant neural correlates localised at approximately 80ms in the cingulum and prenucleus; temporally shifting towards inferior/superior parietal areas at approximately 220ms, and approximately 160ms in the left occipital-parietal regions for encoding congruent visual information (for auditory stimulus presentations) and auditory information (for visual stimulus presentations) respectively. Their MEG analysis further elucidated a discrimination of neural activity underlying their behavioural responses, with previous congruent auditory and visual stimuli influencing overlapping neural activity as early as 80ms in the right parietal cortex, as well as frontal and temporal regions. These results led the researchers to contend that parietal regions represent relevant neural information of associative multisensory congruency, which affects the neural encoding of the currently perceived unisensory stimuli, thus elucidating a perceptually-relevant encoding role of medial parietal regions that consolidates congruent multisensory stimuli to guide adaptive decision-making responses. Further research has also demonstrated that multisensory interaction enhancements during decision formation can occur at very short latencies (Boyle, Kayser, & Kayser, 2017; Cappe et al., 2010; Foxe et al., 2000; 2002; Foxe & Schroeder, 2005; Molholm et al., 2002; 2006; Sperdin et al., 2009). Such findings reaffirm our assumption that associative congruency shapes multisensory decision formation through trial-by-trial perceptual mechanisms, thus improving either or both the speed and accuracy of choice responses. Our results are complementary in this respect, whereby the early modulation we observed further suggests such enhancements are not exclusively multisensory, since on each trial only a single unisensory stimulus was presented. Consequently, we contend that the early onset of our results underlying the benefits to decision formation (i.e. reduced RTs for congruent stimulus feature-response key mappings) implies 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 cross-modal associations may automatically influence early sensory encoding. Cross-modal associations reflect a naturally occurring mapping between stimulus features presented to multiple sensory modalities (Parise, Knorre, & Ernst, 2014; Parise & 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 priors shaped in the statistics of our external 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 mapping. This interpretation is supported by the observation that our discriminator's performance for congruency exceeded the significance level prior to auditory stimulus feature presentation (i.e. $A_z > 0.05$ at 0ms, see Figures 2.3 and 2.4). 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 & Rahnev (2017) present psychophysical evidence to implicate the effects of pre-stimulus anticipation to support this interpretation. In their study, participants made forced choices as to whether the overall orientation of a series of Gabor patches were horizontally orientated clockwise or anti-clockwise. They reported that participants' response bias were significantly affected by post-stimulus cues (presented after stimulus presentations), and not pre-stimulus cues (presented before stimulus presentations, which offered predictions indicative of whether the overall patch orientation was likely to be clockwise or anti-clockwise. They further found that neither pre-stimulus or post-stimulus cues affected stimulus sensitivity, leading to a conclusion that post-stimulus anticipation altered decision criterion mechanisms, but not mechanisms for categorising the sensory signal itself. The contention of an influence of existing priors, or pre-stimulus anticipation, further contributes to the longstanding debate in the field concerning the degree of automaticity of cross-modal associations (Chen & Spence, 2017; Spence & Deroy, 2013; Getz & Kubovy 2018), reaffirming associative congruency benefits involving both perceptual bottom-up and modulatory top-down processes if there are contextual cues that can aid early perceptual processes in improving perceptual decision formation (Ten Oever et al., 2014; Vroomen & Gelder, 2000).

The onset of the second component (C_{Late}: approximately 330-340ms post-stimulus onset) further suggests that the formulation of cross-modal associations and benefits from associative congruency 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 & Sajda, 2006a; 2006b; 2007), arising approximately 300-500ms post-stimulus across centro-parietal electrodes, reflecting neural activity for accumulating evidence to facilitate a choice. Previous studies by Mercier & Cappe (2020) and Schaffhauser, Boubenec, & Mamassian (2021) have further attributed that the CPP indexes the accumulation of sensory evidence for multisensory decision-making and is distinct from the encoding of sensory information earlier during the perceptual decision formation process. In our study, the decoded *Late* component highly resembled the spatiotemporal characteristics of this indexed neural signature, with a positive discriminative topography emerging across similar centro-parietal regions of the brain. Given we further observed higher neural evidence for discriminating associative congruency as late as ~600ms, we contend that the congruency of cross-modal associations for accumulating sensory evidence at a further decisional stage is important to consider, supporting studies demonstrating the CPP for both unisensory and multisensory decision-making, and thus providing a consistent mechanism for benefitting perceptual decision formation. However, we did not observe highly predictive increases in drift rate for congruent trials in our neurally-informed modelling results. Thus, further research should enquire as to the extent cross-modal binding expedites the rate of sensory evidence accumulation uptake, and whether it is dependent on alternate factors such as the type of cross-modal association under investigation and experimental paradigm demands.

In consideration of the possible mechanisms underlying these two EEG components, we contend that cross-modal associations and behavioural enhancements arising from associative congruency may benefit from being consolidated within a predictive coding framework as a mechanism benefitting choices for multisensory perceptual decision-making (Shi & Burr, 2015; Talsma, 2015). Predictive coding frameworks generally model a process of predicting sensory inputs based on the context and representativeness of information from our external environment. Strong evidence that applies a predictive coding framework stems from studies that implement Bayesian interpretations of the effects of existing priors (Huang & Rao, 2011; McGovern et al., 2016; Tong et al., 2020), using an probabilistic causal inference for using optimising choice formations from multisensory stimuli (Cao et al., 2019; Kayser & Shams, 2015). Such Bayesian theories have implicated that cross-modal associative congruency strengthens stimulus feature binding during multisensory integration (Parise & Spence, 2013), demonstrating the pronounced effect of associating priors for benefitting consequent decision formation (Acerbi et al., 2018; Gau & Noppeney, 2016; Rohe Ehlis, & Noppeney., 2019; Rohe & Noppeney, 2015a; 2015b; 2016). Our interpretations posit further support for predictive coding implementations, whereby repeated exposure to auditory pitch-visual size mappings, formulated from a single unisensory stimulus on a single-trial basis, could relate to some existing underlying binding (i.e. mapping) of the perceptual priors between high/low-pitch tones and small/large-size objects respectively. Hence, the early sensory benefits observed from associative congruency may be influenced by newly formed priors of auditory pitch-visual size associations (and perhaps from existing naturalistic priors), with top-down processing signalling to early sensory regions of the brain providing feedback that embeds the environmental prior.

Previous studies have used DDMs to study multisensory perceptual decision-making (Delis et al., 2018; Franzen et al., 2020; Kayser et al., 2017; Mercier & 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. Importantly, our findings suggest that key mechanistic insights can be elicited by coupling models of perceptual decision formation with neuroimaging data. Our results highlighted that 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.

Chapter 3

Characterising Age-Related Impacts in Multisensory Perceptual Decision-Making Processes: A Hierarchical Drift Diffusion Model Analysis

3.1 Abstract

Natural ageing has been demonstrated to diversely impact multisensory perceptual decision-making dynamics. Previous perceptual decision-making research has outlined that older adults display enhanced benefits from perceptually utilising multisensory versus unisensory information within the formation of decisions, coupled with SATO-induced deficits towards decisional speed and accuracy. Despite hypotheses outlining why older adults display preserved multisensory integrative benefits, a mechanistic insight reconciling why older adults may exhibit preserved multisensory integrative benefits, yet display generalised perceptual deficits, relative to younger adults, remains inconclusive. To address this limitation, we sought to investigate age-related impacts on the behavioural indices of multisensory perceptual decision-making. We employed an online variant of a well-established object categorisation paradigm (i.e., audiovisual face-versus-car categorisation paradigm; Franzen et al., 2020), to assess how complementary audiovisual (AV) information benefits perceptual decision formation, compared to visual (V) or auditory (A) information alone, between younger adults (Mean ± Standard Deviation = 27.95 ± 5.82 years) and older adults (Mean ± Standard Deviation = 60.96 ± 10.35 years), recording single-trial RTs and binary responses as metrics of decision-making performance. HDDM was fitted to participants' RTs and binary responses in order to probe age-related impacts on the latent underlying processes of multisensory decision formation. Behavioural results outlined that whereas older adults were generally slower and less accurate relative to younger adults across all trial types, they exhibited no significant difference in RTs between AV versus V trial types, implicating preserved benefits of multisensory integration towards perceptual decision formation. HDDM demonstrated parsimonious fittings for characterising these behavioural discrepancies between YAs and OAs. Notably we found slower rates of sensory evidence accumulation (i.e., ↓ drift rates) for OAs across all sensory trial types, coupled with (1) higher rates of sensory evidence accumulation (i.e., ↑ drift rates) for OAs between AV versus V trial types irrespective of stimulus difficulty, coupled with (2) increased response caution (i.e., ↑ decision boundaries) between AV versus V trial

types, and (3) decreased non-decisional processing duration (i.e., ↓ non-decision times) between AV versus V trial types for stimuli of increased difficulty respectively. Together, our findings provide novel insights into reconciling age-related impacts on multisensory decision-making using HDDM. We implicate that OAs display an increased reliance on multisensory integration to preserve benefits towards decisional speed, therefore compensating for a generalised slowing in consolidating modality-specific unisensory stimuli, yet display increased response caution when unisensory stimuli are more difficult to reconcile. Overall, our findings demonstrate how computational modelling can reconcile contrasting hypotheses of age-related changes in processes underlying multisensory perceptual decision-making behaviour.

3.2 Introduction

As we naturally age, we observe significant lifelong alterations towards our sensory systems, as well as a diverse variety of psychological functions, across the adult lifespan (Cullum et al., 2000; Deary et al., 2009; Hedden & Gabrieli, 2004; Paraskevoudi, Balcı, & Vatakis, 2018; Van Hooren et al., 2007; Zaninotto et al., 2018). Generally, such alterations are characterised by contrasting trajectories of gradual decline and stability underpinning sensation, perception, cognition, and/or movement control (see Alais et al., 2010; Dully, McGovern, & O’Connell, 2018, for detailed reviews). Salient age-related declines in vision, for example, include deficits in visual acuity (Elliott 1987; Elliott, Whitaker, & MacVeigh, 1990; Ward et al., 2018), perception of colour (Paramei, 2012), depth (Normal et al., 2009; 2013), and motion (Bennett, Sekuler, & Sekular, 2007; Billino & Pilz, 2019). Salient age-related declines in hearing, for example, include deficits in pure-tone hearing thresholds (Brant & Fozard, 1990; Lee et al., 2005), frequency and/or intensity discrimination (Freigang et al., 2011), and perception of speed (Taitelbaum-Sweard & Fostick, 2016). Such age-related decrements impact performance levels in experimental paradigms investigating, for example, information processing speed (Salthouse, 1996; 2009), episodic and/or working memory (De Beni & Palladino, 2004), task switching (Wasylyshyn, Verhaeghen, & Sliwinski, 2011), executive function (Rhodes, 2004; Veríssimo et al., 2022), and motor response latency (Birren & Fisher, 1995; Falkenstein et al., 2006). Underpinning such performance outcomes are widespread alterations in brain structure (Pareek, Raliabandi, & Roy, 2018) and brain activity (Rieck, Baracchini, & Grady, 2021), suggesting that the interplay between early sensory and later cognitive processes is affected (Diaconsecu et al., 2013; Dully et al., 2018). Moreover, ageing can have a profound negative impact on everyday living, as it has been shown that ageing compromises of multifaceted alterations to physiological, psychological, and social health and wellbeing (Reetz et al., 2012). Therefore, it is an important field of research to investigate how ageing impacts everyday functioning.

In order to understand why older adults exhibit such alterations to everyday functioning, researchers have sought to pinpoint core processes, affected by ageing, that traverse across multiple domains of psychological functioning (Dully, McGovern, & O'Connell, 2018, Park & Reuter-Lorenz, 2009; Stern, 2002). One fundamental process concerns perceptual decision-making, since most (if not all) behavioural paradigms require a translation of sensory, or sensory-motor, information into appropriate actions. However, unravelling the latent mechanisms impacted by ageing within perceptual decision-making necessitates exploring how multisensory integration is affected by the ageing process, given daily activities require processing information presented to multiple sensory modalities for perceptual decision formation (see Jones & Noppeney, 2021; Mozolic et al., 2012, for reviews). Consequently, unravelling the mechanisms by which multisensory perceptual decision-making varies not only from individual to individual, but with ageing, benefits our understanding of how behavioural attributes of optimal decision-making rely on a consolidation of information across multiple sensory modalities, as well as prior effects of repeated exposure, known to drive adaptive trial-by-trial recalibrations of multisensory perceptual decision-making, across the adult lifespan (Park & Kayser, 2020). Furthermore, in consideration of the profound negative impacts that ageing can have in everyday living, it is important to consider the core processes of multisensory perceptual decision-making; notably multisensory integration, that may be impaired. For example, impaired multisensory integration has been linked with predictions of higher incidences of falling (Zhang et al., 2020; Mahoney, Cotton, & Verghese, 2018), of decreased mobility (Bates & Wolbers, 2014), and of increased risk of pre-dementia stages such as Mild Cognitive Impairment (MCI; Mahoney & Verghese, 2020) in the elderly population.

Previous research provided mixed results on the differences in general multisensory processing capabilities between younger and older adults. Stine, Wingfield, & Myers (1990), for example, found that younger adults, but not older adults, demonstrated improvements in episodic memory (i.e., increased choice accuracy) for audiovisual presentations of television news compared to auditory presentations alone. Similarly, Dobreva, O'Neill, & Paige (2012) investigated spatial memory through ocular fixations of younger and older adults in a paradigm instructing them to localise auditory, visual (both unimodal) and audiovisual (bimodal) targets using a manual laser pointer. They reported that older adults were influenced more strongly by irrelevant distractors in all unimodal and bimodal target conditions, as they demonstrated greatly impaired localisation accuracy and precision in auditory (through vertical overshoots), visual (through horizontal overshoots) and audiovisual (through both horizontal and vertical overshoots) target conditions, implicating modality-dependent deterioration in the multisensory processing of audiovisual spatial objects and a greater susceptibility to biases in processing complementary sensory information during multisensory integration within specific modalities. Finally, Cienkowski & Carney (2002) and Hernández et al. (2019) both reported age-related increases in susceptibility towards McGurk and Sound-Induced Flash illusions respectively, whereby older adults were more likely to report fusions of auditory and visual information when compared to young adults, thus implicating alterations to reliability-weighting of multimodal information for perceptual decision formation. The general conclusion across these results indicates that multisensory processing is affected by ageing, hence resulting in increased variability in perceptual decision formation, characterised by deficits to choice behaviours. Despite this indication, however, it is difficult to interpret how ageing alters multisensory processing capabilities within perceptual decision formation; specifically, how it alters common underlying cognitive mechanisms, since these studies used different complex paradigms and analysis methodologies (see De Dieuleveult et al., 2017; Mozolic et al., 2012, for reviews).

The application of speeded paradigms, or response time enhancement paradigms (e.g., 2AFC paradigms), can contribute to narrowing the gap in understanding age-related changes in multisensory perceptual decision-making. Speeded paradigms assess the degree to which multisensory stimulus presentations, that is, the near-synchronous presentation of two or more unisensory stimuli benefits decision formation, often through quantifying differences in the decisional speed (i.e., RTs) and/or accuracy. Recent research using speeded target detection paradigms have demonstrated that, contrary to findings from non-speeded paradigms, older adults show not only preserved multisensory response facilitation, but enhanced multisensory response facilitation in addition, whereby older adults display, and hence benefit more from stimuli presented near-simultaneously across two or more modalities, suggesting multisensory integration is not necessarily impaired (Jones & Noppeney, 2021; Mahoney et al., 2011; Mozolic et al., 2012). Laurienti et al. (2006), for example, investigated age-related differences in speeded classification of colours, whereby participants had to categorise the colour of visual (coloured discs on a computer screen), auditory (soundbites of colour word verbalizations), and audiovisual stimuli in a 2AFC paradigm as either red or blue. They reported that whilst older adults displayed a generalised slowing of RTs across all conditions, they demonstrated significantly larger differences in RTs between multisensory and unisensory conditions, compared to younger adults. This demonstrates an age-related multisensory enhancement towards perceptual decision formation. Moreover, the observed enhancement (i.e., decreased RTs) for older adults matched and sometimes individually exceeded the faster of the two unisensory conditions for younger adults. Findings from studies by Peiffer et al. (2007) and Diederich et al. (2008) reaffirmed the findings from Laurienti et al. (2006), whereby older adults exhibited greater supra-threshold choice RTs to multisensory stimuli than younger adults, despite a generalised slowing of RTs during the facilitation of decisions.

Together, these findings highlighted that, despite observable deficits to domains of cognitive function (e.g., increased RTs, hence slower decision formation), multisensory integration is not fully impaired. Rather, it can remain preserved, and sometimes enhanced to some extent, given the observations of greater multisensory benefits towards perceptual decision-making behaviour (Bucur et al., 2005; Diaconescu et al., 2013; Gondan & Minakata, 2016). However, there remains a degree of uncertainty about whether older adults benefit more from multisensory integration within perceptual decision formation. The aforementioned study from Diederich et al. (2008), for example, employed a saccadic RT paradigm and a Time-Window-of-Integration (TWIN) model to assess RT patterns cross-modally. Their findings inferred a slowing of peripheral sensory processing in elderly participants. This was further coupled with wider temporal binding windows (TBWs) for integrating multisensory information when compared to younger participants.

This has been empirically validated in later studies (Basharat et al., 2019; Bedard & Barnett-Cowan, 2016), suggesting that multisensory integration was not enhanced, as TWIN modelling did not demonstrate the wider TBWs fully compensating for age-induced peripheral sensory slowing.

Several possible hypotheses have been posited to explain the basis of an age-related preservation in multisensory perceptual decision-making response facilitation (Dieuleveult et al., 2017; Freiher et al., 2013). For example, the *general slowing hypothesis* (Birren & Fisher, 1995; Cerella, 1985; 1990; De Dieuleveult et al., 2017 Mozolic et al., 2012; Salthouse, 1996; 2000) outlines that older adults exhibit slower cognitive and sensorimotor processing for unisensory signals. Therefore, the formation of unisensory perceptual decisions increases in cognitive demand, as less coherent (i.e., "noisy") unisensory signals become more difficult to consolidate, impairing perceptual decision formation (DeLoss et al., 2013; Eusop, Sebban, & Piette, 2001; Guerreiro et al., 2014; 2015). The processing of multisensory signals, however, is less difficult to consolidate, because the unified percepts provide redundant information about the same multimodal object (i.e., same stimulus properties presented to different sensory modalities). This proposes that enhanced multisensory gains in perceptual decision-making may be an artefact of increased proportional differences within the speed of sensory information consolidation between younger and older adults. In particular, it is the differences in cognitive demand for multisensory and unisensory information processing. Despite this explanation, general sensorimotor and cognitive slowing during ageing has not been fully attributed to enhanced multisensory integration, as it cannot fully explain the acceleration of RTs or increased choice accuracy. For example, Laurienti et al. (2006) utilised a *post hoc* adjustment for comparing multisensory and unisensory RTs in their two-choice audiovisual discrimination experiment. This involved log-transforming RTs to equate for age-related differences in the speed of decision classification, even when unisensory trials were the same in stimulus categorisation difficulty for younger and older adults. Their results suggested that older adults still exhibited enhancements in processing multisensory information for perceptual decision-making than younger adults after log-transforming RTs, even when stimulus categorising difficulty was equivalent. However, comparisons of RTs on unisensory trials in the study by Peiffer et al. (2007) were found to be the same for younger and older adults when their analysis fixed the effect of stimulus categorisation difficulty. Thus, it can be argued that the general slowing hypothesis cannot fully capture the benefits of multisensory integration remaining intact, and to an extent enhanced, in older adults compared to younger adults (Murray et al., 2018; Yordanova et al., 2004).

An additional hypothesis concerns the governing principle of inverse effectiveness for multisensory integration (Holmes, 2009; Meredith & Stein, 1986a; 1986b; Stein & Stafford, 2008). As previously outlined in Chapter 1 (see subsection 1.2.1 *Multisensory Integration*), the principle of inverse effectiveness outlines that the magnitude of multisensory enhancements increases when the effectiveness of processing individual sensory stimuli decreases. Therefore, less salient unisensory stimuli are more likely to be integrated, and more salient unisensory stimuli less likely to be integrated, in order to benefit perceptual decision formation. In consideration of research demonstrating age-related functional deficits in sensory systems, multisensory enhancements towards perceptual decision formation are likely to remain preserved, or subsequently increased, in older adults due to reduced acuity in individual senses. This would result in increased multisensory benefits towards choice selection as the stimulus coherence of unisensory stimuli remains naturally degraded (Laurienti et al., 2006; Peiffer et al., 2007). Behavioural findings from Mahoney et al. (2011) empirically validate this notion. They presented younger (mean age = 19.17 years) and older (mean age = 76.44 years) participants with unisensory and paired auditory, visual, and somatosensory stimuli, instructing participants to depress a foot pedal as quickly and as accurately as possible in a simple detection paradigm. They observed faster RTs in all multisensory conditions across both age groups. Interestingly, however, they found that older adults exhibited a greater RT gain when responding to visual-somatosensory stimuli compared to younger adults. This implicates a multisensory response facilitative benefit when integrating visual and somatosensory information, thus benefitting the formation of perceptual decisions. On the contrary, younger adults exhibited a greater RT gain for auditory-visual and auditory-somatosensory stimuli compared to older adults. As such, it can be argued that the principle of inverse effectiveness is either not generalisable for multisensory integration across all modalities, or not the only process to ubiquitously determine multisensory integrative enhancements towards perceptual decision formation.

Contrastingly, further studies have demonstrated that in scenarios with degrading stimulus effectiveness, either due to natural deficits in sensory systems or reduced coherence in super-threshold unisensory stimuli, the principle of inverse effectiveness is not always evident. For example, Tye-Murray et al. (2007) used closed-set Build-A-Sentence (BAS) and City University of New York (CUNY) Sentence tests to investigate age-related effects on audiovisual speech perception. The assessment of scores between unisensory (auditory or visual) and multisensory (audiovisual) conditions did not implicate enhanced speech recognition in older adults, nor multisensory benefits under poor signal clarity (through degrading image and SNR contrasts for visual and auditory stimuli respectively). However, stimulus coherence (i.e., “noise”) was manipulated in the presence of continuous background auditory non-speech. This makes it difficult to discern if this had a confounding effect on older adult’s speech recognition capabilities. Furthermore, it implicates further existing hypotheses for why we see preserved multisensory integration in older adults. Notably, deficits in top-down attentional control (Mozolic et al., 2008a; 2008b), and increased noise at “baseline” levels (de Dieuleveult et al., 2017; Mozolic et al., 2012) when processing unisensory stimuli in perceptual decision formation. Age-related impacts on selective attention are important to consider, since this is a top-down mechanism allowing for a sustained focus on important properties of unisensory signals presented near-simultaneously (Talmsa et al., 2010). Importantly, it has been extensively demonstrated that older adults display deficits in selective attention, whereby they are prone to distractions of irrelevant stimuli presented within attended sensory modalities (Naveh-Benjamin et al., 2005) and across unattended sensory modalities Dobreva et al., 2012; Poliakoff et al., 2006). Younger adults, however, do not display deficits in selective attention, with comparable behavioural and neuroimaging findings demonstrating dividing attention across multiple specific modalities modulates the likelihood of multisensory integration (Rienäcker et al., 2020; Yang et al., 2020). However, previous studies contrast with these findings, implicating that older adults can, in fact, successfully engage selective attention within and across sensory modalities to preserve multisensory integration, therefore retaining enhanced benefits to perceptual decision formation. (Hugenschmidt et al., 2009a; Jones et al., 2019; Mishra et al., 2014) Because of this, it is crucial to consider the interplay between top-down attentional and bottom-up perceptual mechanisms, notably, the hypothesis that older adults display increases in baseline sensory noise within unisensory modalities. Thus, age-related increases in distractibility arise not from deficits to top-down selective attention or attention dividing capabilities, but the adequate means of filtering irrelevant unisensory stimuli (Hugenschmidt et al., 2009b).

The outlined hypotheses: the general slowing hypothesis, increased inverse effectiveness, deficits in attentional control, and increased noise in baseline levels of unisensory processing, offer explanations theorising how older adults display preserved, and sometimes enhanced, multisensory integration within perceptual decision formation. Whilst their explanations are plausible, they are not fully generalisable in solely explaining the variability in age-related effects on multisensory perceptual decision-making behaviour. Furthermore, it is important to highlight that task-dependent manipulations result in age-related increases in response latencies during multisensory decision-making performance. Thus, the core cognitive mechanisms, and their key computations, underpinning age-related impacts on multisensory perceptual decision-making remain subject to debate.

To overcome this, computational modelling can be used to parse the psychological processes underpinning age-related impacts on the behavioural indices of (multisensory) perceptual decision-making (see Dully et al., 2018, for review). As previously outlined in Chapter 1 (see section 1.1. *Unisensory Perceptual Decision-Making*), a highly influential set of computational models for perceptual decision formation concerns sequential sampling models (O’Connell et al., 2018), in particular, the DDM (Forstmann, Ratcliff, & Wagenmakers, 2016; Ratcliff et al., 2016; Ratcliff & McKoon, 2008; Ratcliff & Smith, 2004; Ratcliff, Smith, & McKoon, 2015; Wiecki et al., 2013). The DDM decomposes behavioural data into latent cognitive parameters that capture the perceptual decision formation process, such as the standard of sensory evidence accumulated in the decision process (i.e., *drift rate*), the quantity of evidence required to trigger a decision (i.e., *decision boundary*), and the duration of processes not attributable to sensory evidence accumulation such as sensory encoding and motor response latency (i.e., *non-decision time*). By decomposing a behavioural dataset, i.e., RTs and choice accuracy, into DDM parameters that capture response latencies, age-related processes that drive changes in choice behaviour can be inferred.

Notably, researchers have highlighted three consistent trends that have emerged from DDM applications to perceptual decision-making studies investigating ageing (O’Govern et al., 2018; Dully et al., 2018). First, an age-dependent increase in decision boundaries, thought to be indicative of increases in response caution, and an adoption of more cautious decision policies (Ratcliff et al., 2001; 2006a; McKoon & Ratcliff, 2013; Scheib, Stoll, & Randerath, 2023; Servant & Evans, 2020; Starns & Ratcliff, 2010; Thapar et al., 2003). Second, an age-dependent increase in the duration of non-decisional processes (non-decision time), thought to be indicative of delays in sensory encoding and/or motor response production latency, due to age-related degradations in sensorimotor processes (Ratcliff, et al., 2004a; 2004b; 2006b; von Krause et al., 2020). Third, a highly task-dependent impact of individual-related ageing effects on the average rate of sensory evidence accumulation (drift rate; Ratcliff, Thapar, & McKoon, 2006; Theisen et al., 2021).

Research applying computational modelling, particularly sequential sampling models, to probe age-related differences in perceptual decision-making behaviour has proven invaluable in gaining a mechanistic insight into the psychologically meaningful latent parameters affected. However, a major limitation concerns that their application has been limited to modelling unisensory perceptual decision-making, and not multisensory perceptual decision-making. To our knowledge, no studies have directly addressed whether age-related differences in multisensory perceptual decision-making using similar approaches. Thus, the extent to which model parameter differences accurately reflect modulations to the underlying cognitive computations in multisensory perceptual decision-making remains unclear. Thus, it remains relatively unprobed a unified view of the age-related processes driving changes in the indices of multisensory perceptual decision-making (i.e., RTs and choice accuracy), and in particular, age-related changes in the processes of multisensory integration within perceptual decision formation.

In the present study, we coupled single-trial measurements of multisensory perceptual decision-making behaviour, i.e., RTs and choice accuracy, recorded from an internet-based (i.e., online) variant of an audiovisual object categorisation paradigm (Franzen et al., 2020) with Hierarchical Drift Diffusion Modelling (Wiecki et al., 2013), in order to (a) assess the effects of natural ageing on the behavioural indices of multisensory perceptual decision-making, and (b) to dissect the constituent processes underlying identified age-related modulations of multisensory perceptual decision-making, thus gaining a mechanistic insight into internal cognitive mechanisms that are either degraded or preserved in older adults. By utilizing this experimental paradigm, we could examine the extent to which natural ageing influences the consolidation of complementary unisensory information for perceptual decision formation in a multisensory context. Therefore, we could observe whether the consolidation of congruent audiovisual information improves the ability to formulate perceptual decisions when compared to auditory or visual information alone.

3.3 Materials and Methodology

3.3.1 Participants

Participants were recruited using the Gorilla Experiment Builder research platform (www.gorilla.sc; Anwyl-Irvine et al., 2020; 2021). An *a priori* power analysis was conducted using G*Power version 3.1.9.7 (Faul et al., 2007; 2009; Mayr et al., 2007) to determine the minimum sample size required to test the study hypotheses. Results indicated that for a fixed linear multiple regression model (with R^2 deviation from zero) with three predictors, a minimum sample size of 176 participants was required to achieve 95% power for detecting an effect size of 0.1, at a significance criterion of $\alpha = 0.05$.

Therefore, 212 participants (male = 105, female = 107; mean age = 43.52, standard deviation = 18.46, age range = 18.08 – 86.83) were selected after completing the full experiment between December 2021 and June 2022, receiving a £10 Amazon Voucher as payment. All participants provided informed consent prior to participation, self-reported normal hearing, normal/corrected-to-normal vision, and no history of neurodevelopmental conditions that may affect sensory processing, cognitive function, or movement capabilities. In addition, this sample was selected after being screened to ensure they demonstrated an honest commitment to completing the full experiment (see sections 3.3.3 *Experimental Paradigm* and 3.3.4 *Procedure and Statistical Analysis of Behavioural Data* respectively). This study was approved by the Research Ethics Committees of the College of Business, Law and Social Sciences at Nottingham Trent University (BLSS

REC 2021/45) and the Faculty of Biological Sciences at the University of Leeds (BIOSCI 19-021). It was conducted in accordance with the Declaration of Helsinki (World Medical Association, 2013).

3.3.2 Stimuli

We used a set of 36 grayscale images – 18 images of faces and 18 images of cars (image size: 512 x 512 pixels; bit depth: 8 bits per pixel) – adapted from previous experiments (Diaz et al., 2017; Franzen et al., 2020; Philiastides & Sajda, 2006a; 2006b; 2007; Philiastides, Ratcliff, & Sajda, 2006). The original face images were selected from the Face Database of the Max Plank Institute of Biological Cybernetics (Troje & Bülthoff, 1996) and the car images were sourced from the Internet. All images were equated for spatial contrast, frequency, luminance, and the numbers of frontal and size views (maximum of ±45°), with initial backgrounds removed and replaced with uniform gray backgrounds. They all had identical magnitude spectra (average magnitude spectrum of all images in the database), with their corresponding phase spectra manipulated using the weighted mean phase technique (Blanz & Vetter, 1999; Dakin et al., 2002). This alters the phase coherence of the images and characterises the stimulus coherence of visual stimuli (i.e. the amount of visual evidence available). Two levels of stimulus coherence were used for this study (37.5% and 32.5%, high and low levels of stimulus coherence respectively), since previous findings have found that these are known to yield performance spanning the psychophysical threshold (Philiastides & Sajda, 2006a; 2006b, Philiastides, Ratcliff, & Sajda, 2006). All images were displayed on a white background (RGB [255 255 255]), using the PsychoPy Software (version 1.82.01; Peirce et al., 2019), for a consistent duration of 300ms.

We further used a set of 36 sounds – 18 sounds of human speech and 18 car/street-related sounds – adapted from Franzen et al. (2020). These sounds were sourced from the Internet. No copyright restrictions were in place and modification of the sound files was permitted. These were sampled at a rate of 22.05 kHz and stored as .wav files. A 10ms cosine on/off ramp was added using MATLAB (Version 2015b, The Mathworks, 2015, Natick, Massachusetts) to reduce the effects of sudden sound onsets, with all sounds normalised by their Standard Deviation (SD). Subsequently, the amplitude of these normalised sounds was reduced by 80%, therefore lowering their intensity. Sounds were embedded in Gaussian white noise, and the relative amplitude of the sounds and noise was manipulated to create two different levels of relative signal-to-noise ratio, corresponding to two levels of stimulus coherence of auditory stimuli (0% and 25%, high and low levels of stimulus coherence respectively). The resulting noisy speech and car-related sounds were presented binaurally for a consistent duration of 300ms.

3.3.3 Experimental Paradigm and Procedure

We employed a modified variant of the audiovisual face-vs-car categorisation paradigm adapted from Franzen et al. (2020). This is a 2AFC paradigm that requires participants to categorise, on a single-trial basis, whether a face or a car is embedded in a presented stimulus. The presented stimuli consisted of the following: (a) images of faces and cars (visual stimuli: V), (b) sounds of human speech or car/street-related sounds (auditory stimuli: A), or (c) simultaneously presented and congruent (i.e., matching) images and sounds of faces/human speech and cars/car or street-related sounds (audiovisual stimuli: AV). All stimuli were presented briefly (i.e., 300ms) and in a pseudorandomised sequence. Two levels of stimulus phase coherence were used to vary the amount of sensory evidence in presented images and sounds (i.e., noise; High Coherence: HC; Low Coherence: LC; see 3.3.2 *Stimuli* section), therefore manipulating the difficulty of stimulus categorisation. Participants were instructed to indicate their categorised decision via a standard keyboard button press as quickly and accurately as possible, with RTs and choice accuracy collected as single-trial dependent variable measurements quantifying behavioural performance (and perceptual decision formation).

The experimental paradigm was prepared using the Gorilla Experiment Builder research platform (www.gorilla.sc; Anwyl-Irvine et al., 2020; 2021), and available to complete through an online URL. It could only be completed on a standard desktop computer or portable laptop. Prior to completing the experimental paradigm, participants were presented with an ethical consent form, study information, and instructions for preparation prior to completing the paradigm. The instructions first specified that participants should position themselves in a quiet environment, and to use headphones, or a sound system set at an appropriate volume, in order to adequately hear auditory stimuli. Then, they explicitly specified the instructions of the paradigm itself, whereby they would be shown a series of quick and distorted stimuli consisting of images only (V trial type), sounds only (A trial type), and images and sounds together of faces/human speech or cars/street-related noises (AV trial type), and asked to categorise the presented stimulus as either a face or a car, using the j and k standard keyboard buttons, as quickly and as accurately as possible, with further instructions to position their left index and middle fingers over the j and k keys respectively. Notably, they were informed that audiovisual stimuli would always be matching (i.e., congruent) and that images and sounds would neither be mismatching, and instructed to refrain from categorising images and sounds individually in these trials. They were further informed that if they were unsure in their decision to guess to the best of their capabilities, since they had a maximum time limit to indicate a response, and that they would receive visual feedback following each response.

Figure 3.1 illustrates the procedure on a single-trial basis. Each trial started with a black fixation cross (RGB: [0 0 0]) presented centrally on-screen for 1000ms. Then, one of three stimuli (i.e., images only, sounds only, simultaneously presented images and sounds) were presented for 300ms. Auditory stimuli were accompanied by an image of a speakerphone on screen to immediately indicate to the participant that the presented stimulus is a sound. Participants were then instructed to categorise, as quickly and as accurately as possible, the presented stimuli using the j and k response keys for face and car stimuli respectively. The response deadline was set at 3000ms. Feedback was presented centrally for 1000ms for two possible outcomes: (a) a tick in green for correct responses (RGB: [3 129 3]), or (b) a cross in red for incorrect responses (RGB: [129 3 3]). In total, we presented 216 trials (divided equally between the two response categories, (i.e., 108 face trials; 108 car trials), and the two levels of stimulus coherence, (i.e., 108 HC trials; 108 LC trials), presented in three blocks of 72 trials each and divided equally between the stimuli (i.e., 24 V trials; 24 A trials; 24 AV trials), totalling 72 trials for each stimuli respectively, with a 60 second rest period between blocks. The entire experiment lasted approximately 20-25 minutes.

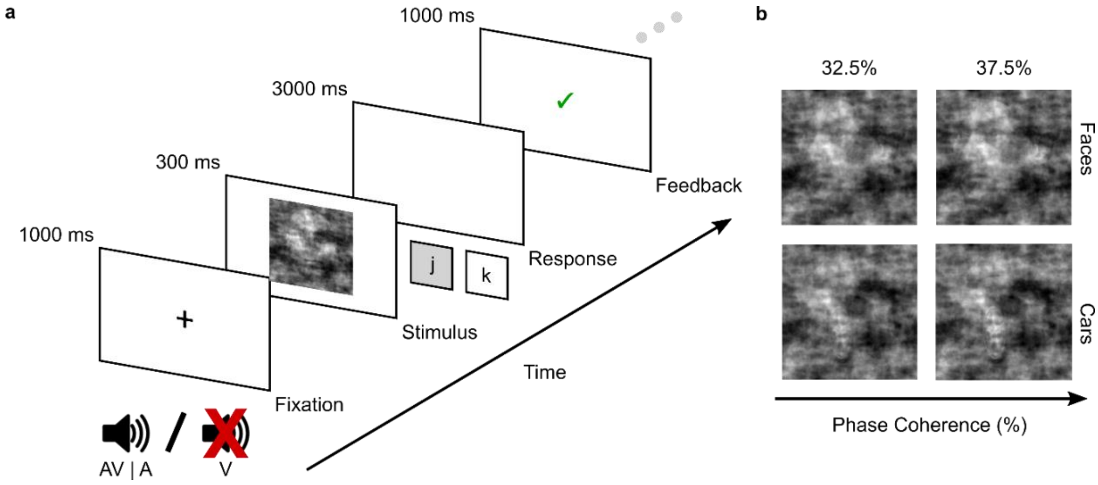


Figure 3.1 Experimental Paradigm. a, Schematic representation of the experimental paradigm on a single-trial basis. Participants were instructed to categorise noisy representations of faces and cars. First, a fixation cross was presented centrally on-screen for 1000ms. Then, a brief stimulus, which was either an image (V), a sound (A), or a simultaneously presented image and sound (AV), was presented for 300ms, and followed by a delay period of a maximum of 3000ms during which participants were instructed to indicate their response with a keyboard button press (*j* or *k* keyboard button presses for face/car categorisations respectively). Following their response, feedback was then presented for 1000ms (a green tick or a red cross, for a correct or incorrect response, respectively), which preceded an inter-stimulus interval (ISI) of 500ms (illustrated by the ellipsis). **b**, Sample face (*top*) and car (*bottom*) images at the two levels of stimulus phase coherence used in the experimental paradigm (Low Coherence: 32.5%; High Coherence: 37.5%).

3.3.4 Statistical Analysis of Behavioural Data

For each participant, RTs (calculated in milliseconds) and choice accuracy (calculated as a binary variable of correct and incorrect responses) were collected as single-trial dependent variable measurements quantifying behavioural performance for three categorical independent variables: (1) sensory trial type (three levels: V | A | AV), (2) stimulus coherence (two levels: HC | LC), and (3) age range (Younger Adults, YA: aged 18-40 | Older Adults: aged 40+). The categorical split of age range was calculated to set a near-even split of participants for further statistical analyses (YA: N = 112; male = 53, female = 59; mean age = 27.95, standard deviation = 5.82; OA: N = 100; male = 46, female = 54; mean age = 60.96, standard deviation = 10.35), and is justified by previous multisensory research that used similar age range splits (Lee et al., 2018; Stevenson et al., 2015).

All participants were screened to ensure they demonstrated an honest commitment towards completing the full experiment to the best of their capabilities. Specifically, we ensured that participants did not attempt to complete the experiment more than once, that they had not procured timed-out responses (i.e., a maximum RT of 3000ms) in all or most trials, and that they had proportions of correct responses greater than 50% across all sensory trial types (demonstrating behavioural performance above a baseline chance level, i.e., guesses). Trials with RTs more or less than the median RT +/- 2.5 Median Absolute Deviations (MADs) were then excluded from further analyses, with these RTs attributed to outliers corresponding to "fast guesses" or attentional lapses during testing (Whelan, 2008). This pre-processing criterion was selected as previous research has demonstrated that MADs are a more robust measurement of central dispersion than standard deviation (Leys et al., 2013). Overall, 4027 trials were excluded from an initial 45792 trials, leaving 41765 trials for further analyses. Next, an Anderson-Darling test was used to assess if RTs were normally distributed (Anderson & Darling, 1954; Nelson, 1998; Stephens, 1974). As previously mentioned in Chapter 2 (see section 2.3.5 *Statistical Analysis of Behavioural Data*), this was used because it is commonly observed in research that RT distributions tend to be not normally distributed, with histograms prominently illustrating a positive right skew (Marmolejo-Ramos et al., 2015; Whelan, 2008). As expected, the assumption of normality was found to be violated for RTs over the full pre-processed behavioural dataset ($A^2 = 347.71$, $p < 0.001$).

Therefore, our main statistical analysis therefore quantified participants' behavioural performance using Generalised Linear Mixed-Effects Models (GLMMs), which were applied using the *lme4* package in RStudio (Bates et al., 2014; R Core Team, 2022).. GLMMs are an extension of linear mixed-effects models (LMMs) and are considered preferable to use over conventional repeated-measures (M)ANOVA statistical analyses, due to their principled methodologies of modelling non-spherical error variance and heteroscedasticity (Bono, Alarcón, & Blanca, 2021; Jaeger, 2008). In particular, random effects structures can be incorporated into the design of a GLMM to account for inter-individual and inter-predictor variability around population-level average effects, therefore increasing statistical power. In addition, GLMMs permit for the mixing of categorical and continuous variables in the statistical analysis of outcome variables, which themselves may be categorical or continuous, and can flexibly accommodate different types of outcome distributions through the application of a variety of link functions (Aarts et al., 2014; Baayern et al., 2008).

Our GLMM analyses included main effects, two-way interactions, and three way-interactions of the three predictor variables: sensory trial type (three levels: Visual, V; Auditory, A; Audiovisual, AV), stimulus coherence (two levels: High Coherence, HC; Low Coherence, LC), and age range (two levels: Older Adults, OA; Younger Adults; YA), along with by-participant random slopes and random intercepts for sensory modality and age range main effects. Random correlations were excluded for all GLMMs. This random effects structure was justified by our experimental design and adopted to ensure parsimonious fits of our GLMMs to the behavioural dataset. The predictor sensory trial type (i.e., three levels) was entered using mean-centred backwards difference coding, whereas the predictors stimulus coherence and age range were entered in mean-centred form (i.e., deviation coding). By using mean-centred contrast coding schemes, small imbalances in trial numbers between each predictor's levels (and their interactions) can be accounted for. All GLMMs were fit using a *bobyqa* optimizer to ensure model convergence. *Post Hoc* likelihood-ratio (χ^2) model comparisons were used to quantify the predictive power and significance of all main effects and interactions in our GLMM analyses. These *Post Hoc* likelihood-ratio (χ^2) model comparisons compared full models (i.e., models including main effects, their two-way interactions, and random effects) to reduced models that excluded the main predictor or two-way interaction in question.

3.3.5 HDDM – Description

We fit participants' behavioural performance (i.e., RTs and choice accuracy) with Hierarchical Drift Diffusion Models (HDDMs; Wiecki et al., 2013). Similar to traditional Drift Diffusion Models (DDMs; Ratcliff et al., 2015; 2016; Forstmann et al., 2016; Ratcliff & McKoon, 2008; Ratcliff, 1978), HDDMs assume a continuous and stochastic process of sensory evidence accumulation over time, towards one of two decision boundaries, corresponding to two choice alternatives (e.g., correct/incorrect responses; j/k response keys). For each decisional process, the HDDM returns estimates of four parameters that define the scopes of the internal components of perceptual decision formation: (1) the rate of evidence accumulation (*drift rate*), (2) the distance between the two decisional boundaries that quantifies the amount of evidence to facilitate one particular choice alternative (*decision boundary*), (3) the duration of non-decisional processes, that is, the time taken for processes that are not part of the evidence accumulation process, such as stimulus encoding and motor-response production latency (*non-decision time*), and (4) possible *a priori bias* towards one of the two choice alternatives (*starting point*).

As previously outlined in Chapter 2 (see sections 2.3.8 *HDDM - Description* and 2.3.9 *HDDM – Fitting* respectively), HDDMs are extensions of traditional DDM (Ratcliff et al., 2015; 2016; Forstmann et al., 2016; Ratcliff & McKoon, 2008; Ratcliff, 1978) that use Bayesian hierarchical frameworks and MCMC sampling to iteratively adjust DDM parameters to maximise the summed log-likelihood of predicted mean RTs and choice accuracy. Specifically, prior probability distributions of the model parameters are simultaneously estimated at individual participant and group levels under the assumption that individual participant level parameters are constrained within the group level distributions, and vice versa. These prior distributions are then updated on the basis of the likelihood of the data given the model, to yield posterior probability distributions of the sampled parameters. The use of Bayesian hierarchical frameworks, and specifically HDDMs, allows for several benefits relative to traditional non-hierarchical DDM analyses. In particular, the assumption that individual participant samples are randomly drawn from a constrained group level distribution yields posterior parameter estimates with higher stability and robustness. In addition, they can quantify uncertainty through posterior distributions for estimated parameters, rather than converging on the most likely value for each parameter (Gelman, 2003; Navarro & Fuss, 2009; Vadekerchkove et al., 2011; Wiecki et al., 2013).

3.3.6 HDDM – Fitting

To fit HDDMs to participants' single-trial behavioural performance measurements (i.e., RTs and choice accuracy), and estimate the internal components of perceptual decision formation, 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. We used a process referred to as accuracy-coding, which fits HDDMs to RT distributions that assume the upper and lower decision boundaries correspond to correct and incorrect choices respectively. Eight

HDDM accuracy-coded variants were fit to our behavioural dataset. Seven variants sampled posterior parameter estimates for combinations of drift rate (δ), decision boundary (θ), and non-decision time (τ) for the conditional dependencies (levels) of all three of our independent variables (sensory trial type, stimulus coherence, and age range), whereas one variant was fit to our behavioural dataset that did not allow parameters to vary by our conditional dependences. Starting point (z) was set as the midpoint between the two decision boundaries for all variants, since stimuli were presented in a pseudorandomised order in the experimental paradigm, thereby considerably reducing the likelihood of an *a priori* bias towards either choice alternative. In addition, we fixed the trial-to-trial variabilities of each parameter to 0, since previous research has found that these can improve parameter estimates for drift rate (δ), decision boundary (θ), and non-decision time (τ; Lerche & Voss, 2016).

In total, we sampled estimated posterior distributions for a maximum of 12 drift rate (δ), decision boundary (θ), and non-decision time (τ) parameters across all conditional dependencies for the three independent variables (sensory trial type: Visual, V; Auditory, A; Audiovisual, AV; stimulus coherence: High Coherence, HC; Low Coherence, LC; Age Range: Older Adults, OA; Younger Adults; YA) as follows:

$$\delta = \chi_{i,j} \sim F\left(\delta^{\substack{V,\ A,\ AV \\ OA,\ YA}}_{HC,\ LC}\right)$$

$$\theta = \chi_{i,j} \sim F\left(\theta^{\substack{V,\ A.\ AV \\ OA,\ YA}}_{HC,\ LC}\right)$$

$$\tau = \chi_{i,j} \sim F\left(\tau^{\substack{V,\ A,\ AV \\ OA,\ YA}}_{HC,\ LC}\right)$$

where $\chi_{i,j}$ represents the observed behavioural data (i.e. RTs and choice accuracy) for participant *i* and trial *j*, and $F(\dots)$ represents the DDM Wiener likelihood function (as formulated by Navarro & Fuss, 2009) to simultaneously sample individual participant and group level parameters.

For each HDDM variant, we ran 5 separate Markov chains with 11000 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 25000 remaining samples for each modelling variant, which constituted the probability distributions for each estimated parameter, allowing us to compute individual parameter estimates for participants and condition categories in each variant. To ensure Markov Chain convergence, we computed Gelman-Rubin $\hat{R}$ statistics between chains (Gelman & 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. Finally, for comparisons between each variant, we used the Deviance Information Criterion (DIC), a measure popular for assessing and comparing relative hierarchical model fits to behavioural datasets, particularly those that use MCMC sampling (Spiegelhalter et al., 2002). The DIC evaluates the trade-off between a modelling variant’s goodness-of-fit and complexity (i.e., number of parameters). We selected the modelling variant with the lowest DIC, which favours the model with the highest likelihood of a good fit to data for the least degrees of freedom.

3.3.7 HDDM – Hypothesis Testing

As previously outlined in Chapter 2 (see sections 2.3.9 *HDDM – Fitting* and 2.3.10 *HDDM – EEG Regressors*), Bayesian hierarchical modelling frameworks naturally 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). Consequently, null-hypothesis significance testing approaches commonly utilised in frequentist approaches are not recommended. Therefore, in order to analyse the predictive power of each posterior distribution for each parameter (and its conditional dependencies), we used Bayesian hypothesis testing. Specifically, we compared the proportion of posterior distributions between two of the same parameters, complete with their conditional dependences, that did not overlap in their sampled parameter estimations, returning a value between 0 and 1, depending on the hypothesis tested. To determine the prevalence of true positive results, implicating strong predictive effects in the difference between posterior distributions, we calculated the log posterior odds proportion of a hypothetical proportion corresponding to a false-positive rate of $\alpha = 0.05$ (i.e., a 95% true-positive threshold, similar to the utilisation of p-values in null-hypothesis significance testing approaches; Ince et al., 2021). Non-overlap proportions greater than the hypothetical log-odds proportion of our false positive rate (which is equal to 2.944) suggests highly predictive effects in the non-overlap of parameter posterior distributions favoured by our hypotheses.

We sought to use the results of our statistical analyses to inform the fitting of the HDDM to our behavioural data (i.e., single-trial RTs and binary responses). Tables 3.1, 3.2, and 3.3 outline the hypotheses across all conditional dependencies. Generally, in line with the results of our RT statistical analyses (see section 3.4.1 *Behavioural Results*), in which we observed (1) significant decreases in RTs across levels of multisensory (i.e., AV trial types) and unisensory (i.e., V and A trial types) sensory trial type, stimulus coherence (i.e., HC versus LC trial types), and age range (i.e., YAs versus OAs), (2) significant increases in choice accuracy across levels of multisensory (i.e., AV) and unisensory (i.e., V and A) sensory trial type, stimulus coherence (i.e., HC versus LC trial types), and age range (i.e., YAs versus OAs), and (3) discrepancies in significance in two-way interactions between sensory trial type and age range for RTs (coupled with non-significant two-way interactions between sensory trial type and age range for choice accuracy), we hypothesised that there would be strong predictive differences (i.e., prevalent non-overlaps between proportions of posterior distributions) in drift rate and non-decision time parameter estimates.

In general, we tested hypotheses outlining that drift rate parameter estimates would predictively increase across sensory trial types; in which AV trial types would have higher drift rate parameter estimates versus V trial types ($\delta_{AV} > \delta_{V}$) and A trial types ($\delta_{AV} > \delta_{A}$), and V trial types would have higher drift rate estimates versus A trial types ($\delta_{V} > \delta_{A}$), predictively decrease across levels of stimulus coherence, whereby HC trial types would have higher drift rate parameter estimates versus LC trial types ($\delta_{HC} > \delta_{LC}$), and predictively decrease across age ranges, whereby younger adults would have higher drift rate parameter estimates versus older adults ($\delta_{YA} > \delta_{OA}$), and (2) non-decision time parameter estimates would predictively decrease across sensory trial types, whereby AV trial types would have lower non-decision time estimates versus V trial types ($\tau_{AV} < \tau_{V}$) and versus A trial types ($\tau_{AV} < \tau_{A}$), and V trial types would have lower non-decision time estimates versus A trial types ($\tau_{V} < \tau_{A}$), predictively increase across levels of stimulus coherence, whereby HC trial types would have lower non-decision time estimates versus LC trial types ($\tau_{HC} < \tau_{LC}$), and predictively increase across age ranges, whereby older adults would have higher non-decision time parameter estimates versus younger adults ($\tau_{YA} < \tau_{OA}$).

Conversely, in line with the results of our choice accuracy statistical analyses, in which we report significant effects of choice accuracy decreasing across levels of sensory trial type, stimulus coherence, and age ranges, but no significant two-way interaction between sensory trial type and age range we hypothesised that there would be strong predictive differences (i.e., prevalent non-overlaps between proportions of posterior distributions) in decision boundary estimates. In particular, we tested the following bidirectional hypotheses: a) coupled with significant decreases in RTs, decision boundaries would predictively decrease across sensory modalities, whereby AV trials would have lower decision boundary parameter estimates versus V trial types ($\theta_{AV} < \theta_{V}$), and versus A trial types ($\theta_{AV} < \theta_{A}$), and V trial types would have lower decision boundary estimates versus A trial types ($\theta_{V} < \theta_{A}$), predictively increase across levels of stimulus coherence, whereby HC trial types would have lower decision boundary estimates versus LC trial types ($\theta_{HC} < \theta_{LC}$), and predictively increase across age ranges, whereby younger adults would have lower decision boundary parameter estimates versus older adults ($\theta_{YA} < \theta_{OA}$).

3.4 Results

3.4.1 Behavioural Results

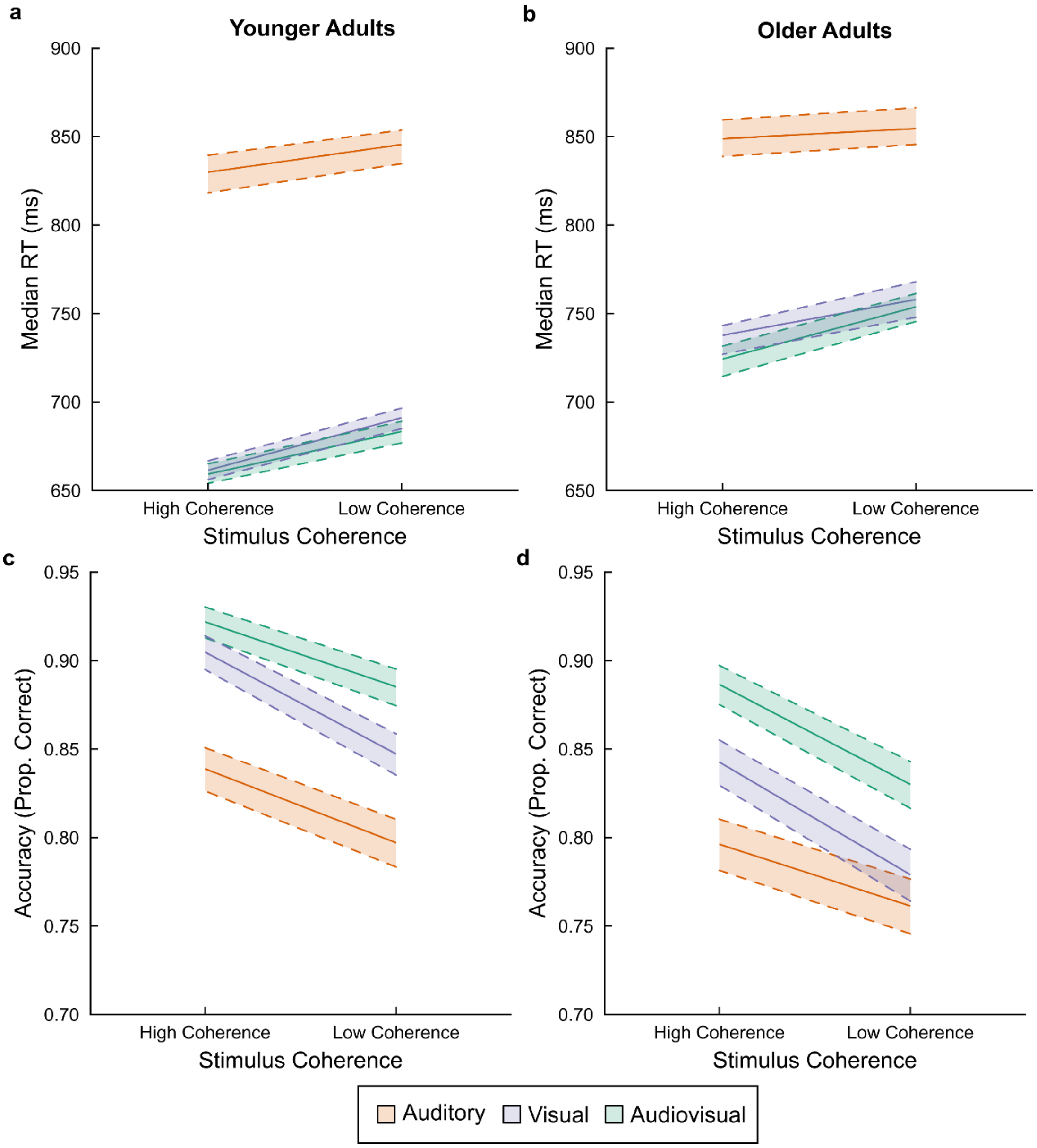


Figure 3.2 Behavioural Performance. a, **b**, Median RTs (*top*) and **c**, **d**, choice accuracy (proportion of correct responses; *bottom*) for Younger Adults (*left*) and Older Adults (*right*) across the two levels of stimulus coherence (high/low coherence) and as a function of visual (purple), auditory (orange), and audiovisual (green) trials. Shaded regions indicate 95% Confidence Intervals (CIs), which were calculated using 1000 bootstrapping random sampling interactions to estimate the distribution of group average performance measurements.

Generalised Linear Mixed-Effects Models (GLMMs) and *post hoc* likelihood-ratio (χ^2) model comparisons were used to analyse RTs and choice accuracy (using *gamma* and *binomial* logit models for RTs and choice accuracy respectively) as a function of sensory modality, stimulus coherence, and age range, as well as their subsequent two-way interactions.

3.4.1.1 Reaction Time GLMM Analyses Results

Our RT analyses demonstrated a significant main effect of sensory trial type (Figures 3.2a and 3.2b; GLMM: $\chi^2 = 140.78$, df = 2, $p < 0.001$); in which participants responded faster to AV versus V trial types ($Median_{AV} = 697$ms post-stimulus offset, $Median_{V} = 704$ms post-stimulus offset; GLMM (AV versus V): $\chi^2 = 10.14$, df = 1, $p < 0.001$), AV versus A trial types ($Median_{AV} = 696$ms post-stimulus offset, $Median_{A} = 844$ms post-stimulus offset; GLMM (AV versus A): $\chi^2 = 96.27$, df = 1, $p < 0.001$), and during V versus A trial types ($Median_{V} = 704$ms post-stimulus offset, $Median_{A} = 844$ms post-stimulus offset; GLMM (V versus A): $\chi^2 = 67.68$, df = 1, $p < 0.001$). In addition, they demonstrated a significant main effect of stimulus coherence, in which participants responded faster to HC versus LC trial types ($Median_{HC} = 727$ms post-stimulus offset, $Median_{LC} = 749$ms post-stimulus offset; GLMM: $\chi^2 = 145.84$, df = 1, $p < 0.001$). A significant two-way interaction was further found between sensory trial type and stimulus coherence (GLMM: $\chi^2 = 18.60$, df = 2, $p < 0.001$), in which participants responded faster to HC versus LC trial types between AV versus A trial types ($Median_{AV \mid HC} = 683$ms post-stimulus offset; $Median_{A \mid HC} = 838$ms post-stimulus offset; $Median_{AV \mid LC} = 713$ms post-stimulus offset; $Median_{A \mid LC} = 850$ms post-stimulus offset; GLMM (AV versus A): $\chi^2 = 15.50$, df = 1, $p < 0.001$), and between V versus A trial types ($Median_{V \mid HC} = 691$ms post-stimulus offset; $Median_{A \mid HC} = 838$ms post-stimulus offset; $Median_{V \mid LC} = 717$ms post-stimulus offset; $Median_{A \mid LC} = 850$ms post-stimulus offset; GLMM (V versus A): $\chi^2 = 10.12$, df = 1, $p = 0.001$), but not between AV versus V trial types ($Median_{AV \mid HC} = 683$ms post-stimulus offset; $Median_{V \mid HC} = 691$ms post-stimulus offset; $Median_{AV \mid LC} = 713$ms post-stimulus offset; $Median_{V \mid LC} = 717$ms post-stimulus offset; GLMM (AV versus V): $\chi^2 = 0.72$, df = 1, $p = 0.395$). Together, these results highlight larger decreases in RTs with AV trial types, relative to V and A trial types, with a further decrease in RTs with the amount of complementary V evidence salience (i.e., stimulus difficulty), but not for A evidence salience, benefitting decision speed.

Our analyses further demonstrated a significant main effect of age, in which YAs responded faster than OAs (Figures 3.2a and 3.2b; $Median_{YA} = 708$ms post-stimulus offset, Median$_{OA}$ = 774ms post-stimulus offset; GLMM: χ^2 = 18.13, df = 1, p < 0.001). Moreover, a significant two-way interaction was found between sensory trial type and age for RTs (χ^2 = 33.10, df = 2, p < 0.001). In particular, we found significant age-related differences in RTs between AV versus A trial types, whereby YAs demonstrated a larger decrease in RTs for AV versus A trial types (Median$_{AV \mid OA}$ = 738ms post-stimulus offset; Median$_{A \mid OA}$ = 852ms; Median$_{AV \mid YA}$ = 670ms post-stimulus offset; Median$_{A \mid YA}$ = 837ms post-stimulus offset; χ^2 = 5.01, df = 1, p < 0.025) and between V and A trials (Median$_{V \mid OA}$ = 745ms post-stimulus offset; Median$_{A \mid OA}$ = 852ms post-stimulus offset; Median$_{V \mid YA}$ = 676ms post-stimulus offset; Median$_{A \mid YA}$ = 837ms post-stimulus offset; GLMM (AV versus A): χ^2 = 9.58, df = 1, p = 0.025), but not between AV versus V trial types (Median$_{AV \mid OA}$ = 738ms post-stimulus offset; Median$_{V \mid OA}$ = 745ms post-stimulus offset; Median$_{AV \mid YA}$ = 670ms post-stimulus offset; Median$_{V \mid YA}$ = 676ms post-stimulus offset; χ^2 = 1.31, df = 1, p = 0.254). To reaffirm these significant trends and account for an age-related generalised slowing, log-transformed RTs were further analysed (Cerella, 1995; Diederich et al., 2008; Laurenti, 2006; Peiffer et al., 2007; Salthouse, 1996). Application of an ANOVA to compare AV trial types versus V and A trial types, between OAs versus YA, demonstrated significant main effects of sensory trial type ($F_{2,\,41759}$ = 1179.00, p < 0.001) and age range ($F_{1,\,41759}$ = 394.12, p < 0.001), and a sensory trial type by age range interaction ($F_{2,\,41759}$ = 53.09, p < 0.001). *Post Hoc* comparisons using Tukey HSD testing between log-transformed AV versus V trial type RTs for OAs versus YAs further indicated that OAs, but not YAs, had significantly greater differences between AV versus V trial types (OAs: Mean Difference$_{Log\text{-}RTs}$ = - 0.016 [-0.030 -0.003], p = -0.008; YAs: Mean Difference$_{Log\text{-}RTs}$ = - 0.005 [-0.018 0.007], p = 0.820).

Finally, our analyses did not find a significant two-way interaction between stimulus coherence and age range (Median$_{HC \mid OA}$ = 763ms post-stimulus offset; Median$_{LC \mid OA}$ = 784ms post-stimulus offset; Median$_{HC \mid YA}$ = 695ms post-stimulus offset; Median$_{LC \mid YA}$ = 721ms post-stimulus offset; χ^2 = 1.97, df = 1, p = 0.161), nor a significant three-way interaction (Median$_{V \mid HC \mid OA}$ = 738ms post-stimulus offset; Median$_{A \mid HC \mid OA}$ = 849ms post-stimulus offset; Median$_{AV \mid HC \mid OA}$ = 724ms post-stimulus offset; Median$_{V \mid LC \mid OA}$ = 758ms post-stimulus offset; Median$_{A \mid LC \mid OA}$ = 855ms post-stimulus offset; Median$_{AV \mid LC \mid OA}$ = 754ms post-stimulus offset; Median$_{V \mid HC \mid YA}$ = 662ms post-stimulus offset; Median$_{A \mid HC \mid YA}$ = 830ms post-stimulus offset; Median$_{AV \mid HC \mid YA}$ = 659ms; Median$_{V \mid LC \mid YA}$ =

691ms post-stimulus offset; $\text{Median}_{A \mid LC \mid YA}$ = 846ms post-stimulus offset; $\text{Median}_{AV \mid LC \mid YA}$ = 683ms; χ^2 = 2.93, df = 2, p = 0.232).

3.4.1.2 Choice Accuracy GLMM Analyses Results

Our choice accuracy analyses demonstrated significant main effects of sensory trial type (GLMM: χ^2 = 166.00, df = 2, p < 0.001), in which participants responded more accurately to AV versus V trial types ($\text{Proportion Correct}_{AV}$ = 0.885; $\text{Proportion Correct}_{V}$ = 0.851; GLMM (AV versus V): χ^2 = 70.70 df = 1, p < 0.001), AV versus A trial types ($\text{Proportion Correct}_{AV}$ = 0.885; $\text{Proportion Correct}_{A}$ = 0.798; χ^2 = 113.82, df = 1, p < 0.001; GLMM (AV versus A): χ^2 = 113.82, df = 1, p < 0.001), and V versus A trial types ($\text{Proportion Correct}_{V}$ = 0.851; $\text{Proportion Correct}_{A}$ = 0.798; GLMM (V versus A): χ^2 = 21.32, df = 1, p < 0.001). In addition, they demonstrated a significant main effect of stimulus coherence, in which participants performed more accurately in HC versus LC trial types ($\text{Proportion Correct}_{HC}$ = 0.869; $\text{Proportion Correct}_{LC}$ = 0.823; GLMM: χ^2 = 189.00, df = 1, p < 0.001). A significant two-way interaction was further found between sensory trial type and stimulus coherence (GLMM: (χ^2 = 18.40, df = 2, p < 0.001), in which participants performed more accurately to HC versus LC trial types between AV versus A trial types ($\text{Proportion Correct}_{AV \mid HC}$ = 0.907; $\text{Proportion Correct}_{A \mid HC}$ = 0.815; $\text{Proportion Correct}_{AV \mid LC}$ = 0.863; $\text{Proportion Correct}_{A \mid LC}$ = 0.780; GLMM (AV versus A): χ^2 = 10.37, df = 1, p = 0.001), and between V versus A trial types ($\text{Proportion Correct}_{V \mid HC}$ = 0.0.880; $\text{Proportion Correct}_{A \mid HC}$ = 0.815; $\text{Proportion Correct}_{V \mid LC}$ = 0.821; $\text{Proportion Correct}_{A \mid LC}$ = 0.780; GLMM (V versus A): χ^2 = 21.32, df = 1, p < 0.001), but not between AV versus V trial types ($\text{Proportion Correct}_{AV \mid HC}$ = 0.907; $\text{Proportion Correct}_{V \mid HC}$ = 0.0.880; $\text{Proportion Correct}_{AV \mid LC}$ = 0.863; $\text{Proportion Correct}_{V|LC}$ = 0.821; GLMM (AV versus V): χ^2 = 0.21, df = 1, p = 0.643). Together, these results highlight larger decreases in choice accuracy with AV trial types, relative to V and A trial types, with a further decrease with the amount of complementary A evidence salience (i.e., stimulus difficulty), but not for V evidence salience, benefitting decision accuracy.

Our analyses further demonstrated a significant main effect of age, in which YAs performed more accurately than OAs ($\text{Proportion Correct}_{YA}$ = 0.870; $\text{Proportion Correct}_{OA}$ = 0.819; χ^2 = 23.80, df = 1, p < 0.001). However, no significant interaction was found between sensory trial type and age range ($\text{Proportion Correct}_{AV \mid OA}$ = 0.862; $\text{Proportion Correct}_{A \mid OA}$ = 0.772; Proportion

Correct$_{V | OA}$ = 0.819; Proportion Correct$_{AV | YA}$ = 0.906; Proportion Correct$_{A | YA}$ = 0.821; Proportion Correct$_{V | YA}$ = 0.879; GLMM: χ^2 = 2.12, df = 2, p = 0.346) nor stimulus coherence and age range (Proportion Correct$_{HC | OA}$ = 0.843; Proportion Correct$_{LC | OA}$ = 0.794; Proportion Correct$_{HC | YA}$ = 0.892; Proportion Correct$_{LC | YA}$ = 0.847; GLMM: χ^2 = 1.62, df = 1, p = 0.204). In addition, no significant three-way interaction was found between sensory trial type, stimulus coherence, and age range (Proportion Correct$_{V | HC | OA}$ = 0.850; Proportion Correct$_{A | HC | OA}$ = 0.787; Proportion Correct$_{AV | HC | OA}$ = 0.889; Proportion Correct$_{V | LC | OA}$ = 0.787; Proportion Correct$_{A | LC | OA}$ = 0.758; Proportion Correct$_{AV | LC | OA}$ = 0.835; Proportion Correct$_{V | HC | YA}$ = 0.906; Proportion Correct$_{A | HC | YA}$ = 0.840; Proportion Correct$_{AV | HC | YA}$ = 0.924; Proportion Correct$_{V | LC | YA}$ = 0.851; Proportion Correct$_{A | LC | YA}$ = 0.800; Proportion Correct$_{AV | LC | YA}$ = 0.877; χ^2 = 0.92, df = 2, p = 0.632).

3.4.1.3 Interim Summary of Behavioural Results

We observed general age-related declines in decision speed (i.e., increased RTs) and accuracy (i.e., decreased proportions of correct responses). In addition, we saw age-related differences in RTs between multisensory (i.e., AV) versus unisensory (i.e., V and A trial types), albeit with different trajectories in RT difference. Specifically, whereas YAs displayed a greater multisensory benefit in RT differences between AV versus A trial types (i.e., larger AV-A RT Differences; coupled with a greater benefit in RT differences between V versus A trial types; larger V–A RT Differences) relative to OAs, OAs displayed a greater multisensory benefit in log-transformed RT differences between AV versus V trial types (i.e., larger AV-V RT Differences), irrespective of the salience of complementary sensory evidence (i.e., stimulus coherence), with non-transformed RT differences suggesting no significant two-way interaction, indicating a preservation of multisensory integrative benefits in consolidating complementary A information. Regarding choice accuracy, however, no age-related differences between multisensory (i.e., AV trial types) and unisensory (i.e., V and A trial types) were found, implying a preservation of choice accuracy not impacted by ageing. In addition, no further effects seemed to be elicited suggesting generalisable multisensory deficits towards choice accuracy as a function of either categorical or chronological age between AV versus V trial types, implicating an age-related preservation of the benefits towards choice accuracy.

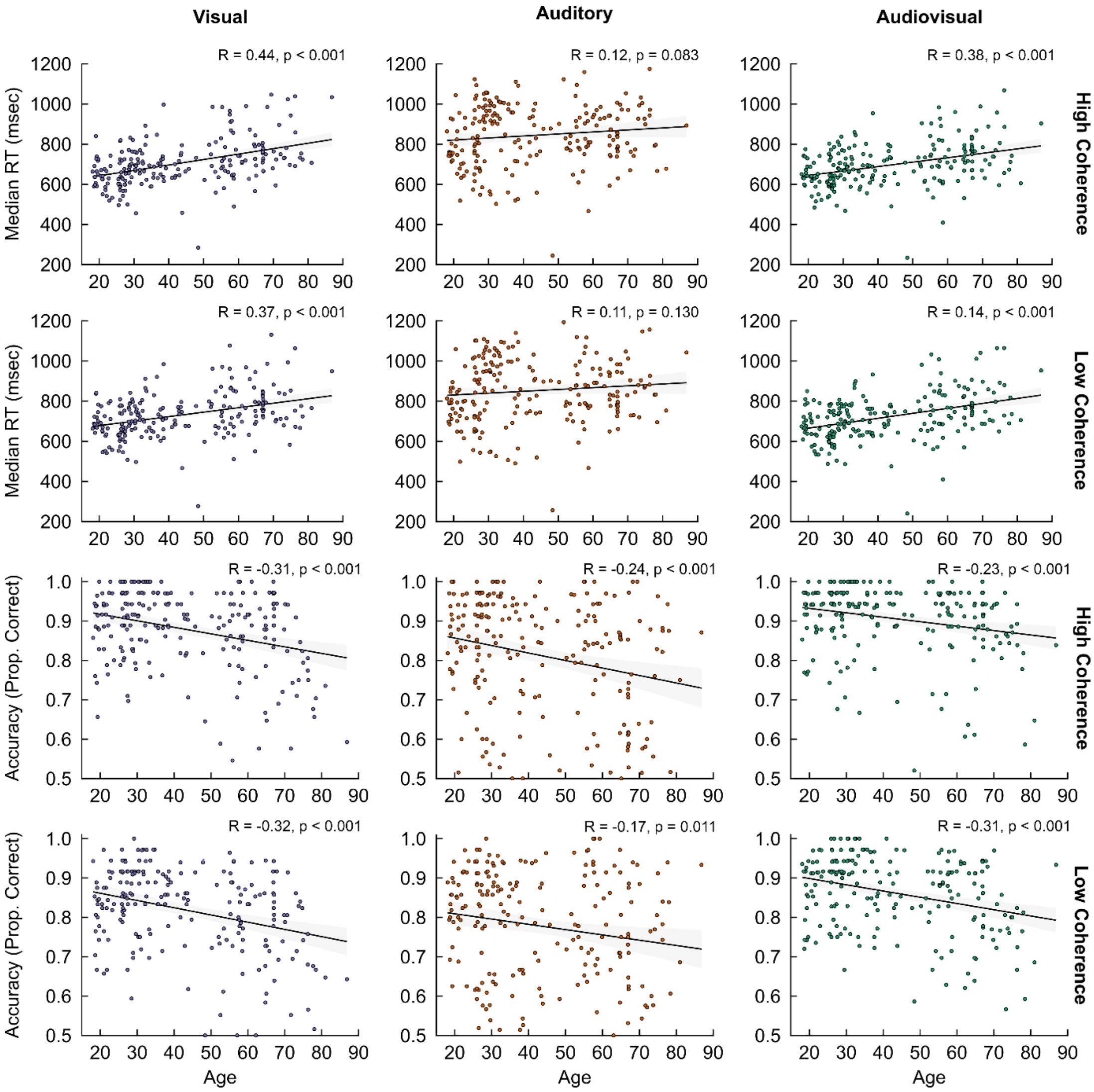
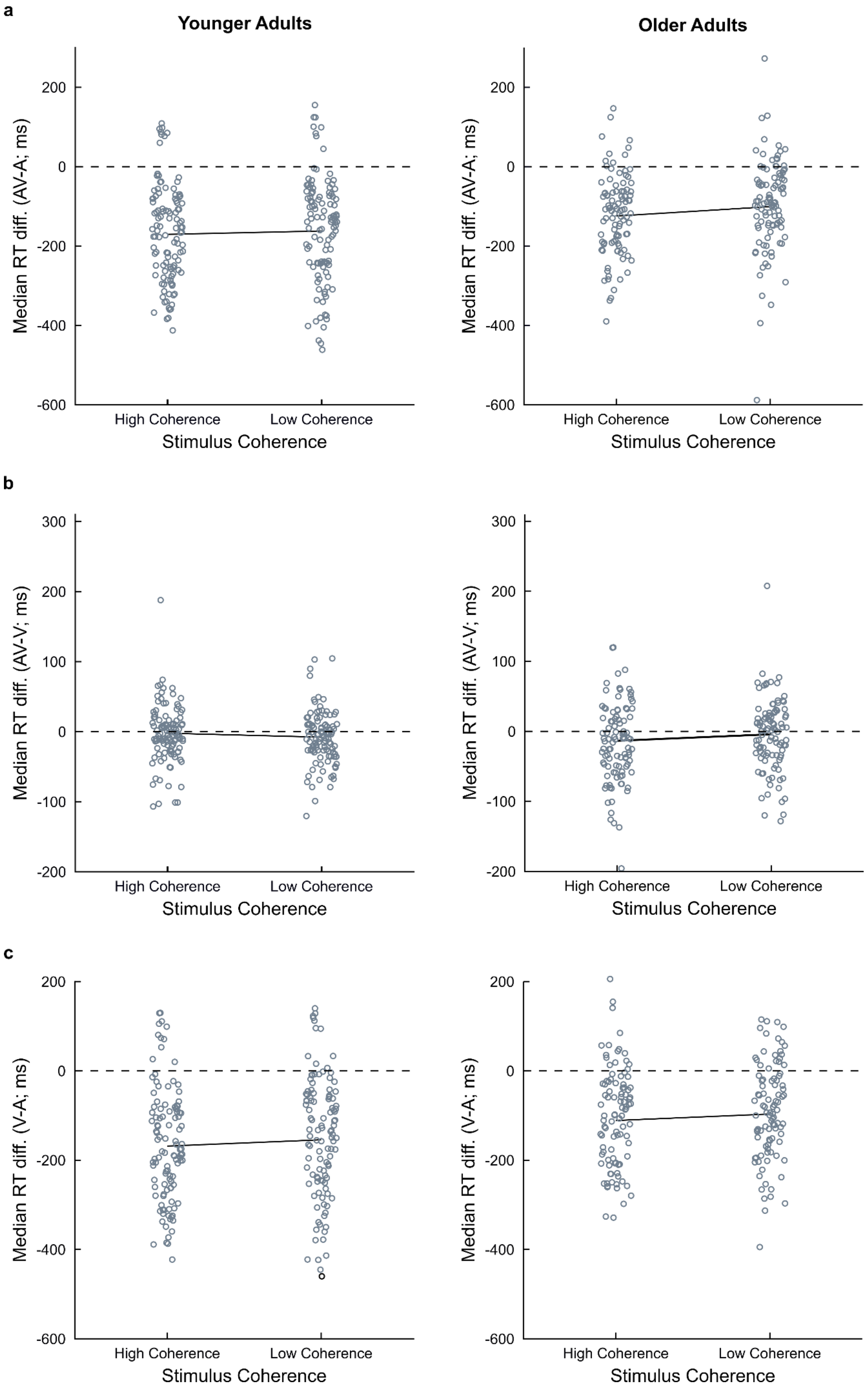


Figure 3.3 Behavioural Performance Correlations. Pearson correlations of age with median RTs (*top*) and choice accuracy (proportion of correct responses; *bottom*) across the two levels of stimulus coherence (high/low coherence) as a function of visual (*left*), auditory (*middle*), and audiovisual (*right*) trial types. Pearson Correlation Coefficients (R) and p-values are shown for each correlation. Shaded regions indicate 95% Confidence Intervals (CIs).


a
Younger Adults
Older Adults
Median RT diff. (AV-A; ms)
200
0
-200
-400
-600
High Coherence
Low Coherence
Stimulus Coherence
b
Median RT diff. (AV-V; ms)
300
200
100
0
-100
-200
c
Median RT diff. (V-A; ms)

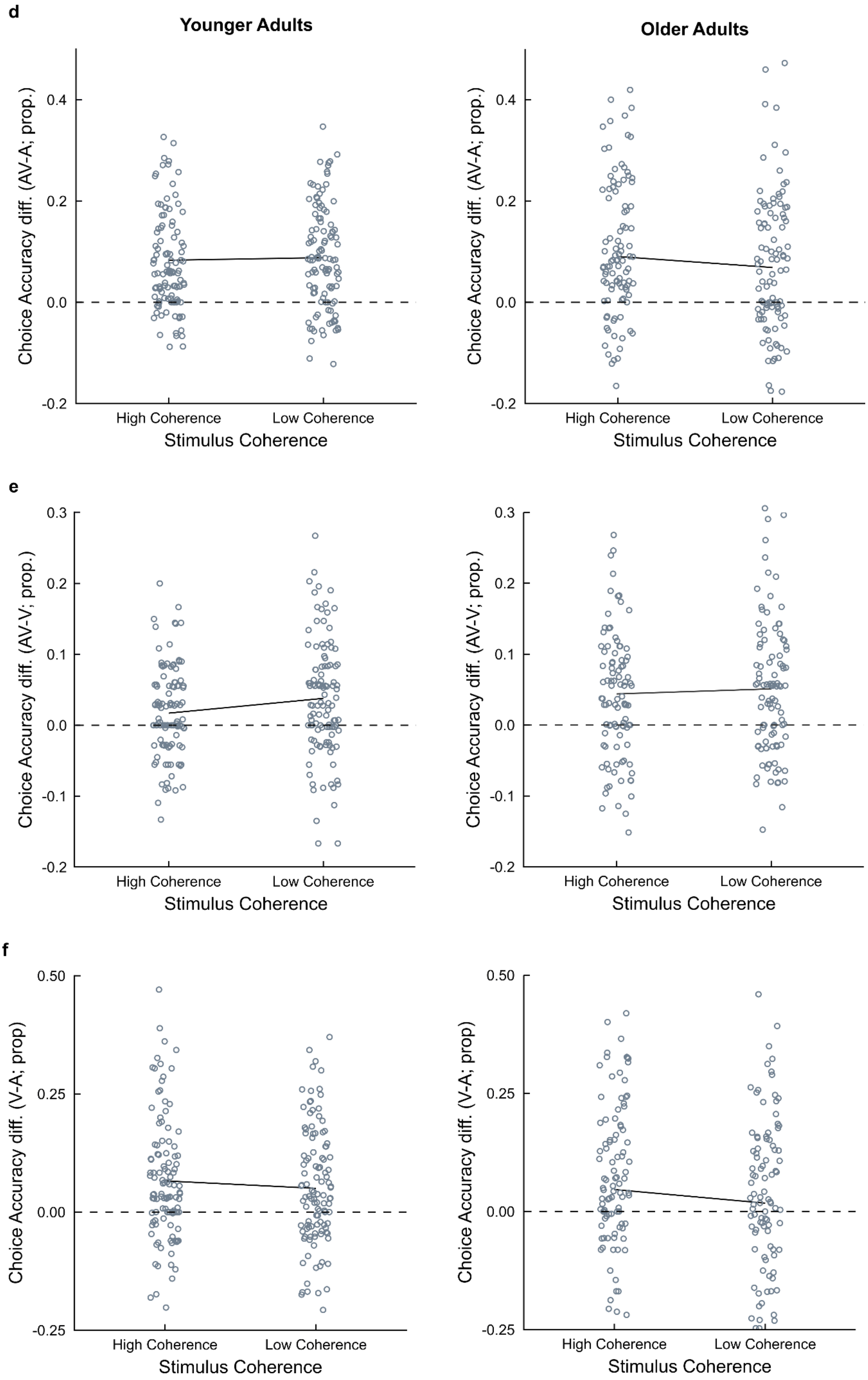


Figure 3.4 Behavioural Performance Differences. Individual participant behavioural performance changes (AV-A; *top*, AV-V*; middle*, V-A; *bottom*)

for **a**, **b**, **c**, median RTs and **d**, **e**, **f**, choice accuracy (proportion of correct responses) for Younger Adults (*left*) and Older Adults (*right*) across the two levels of stimulus coherence (high coherence/low coherence). Solid black lines denote the group averages (calculated across N = 212 participants), and dashed grey lines denote the boundaries for no behavioural performance differences.

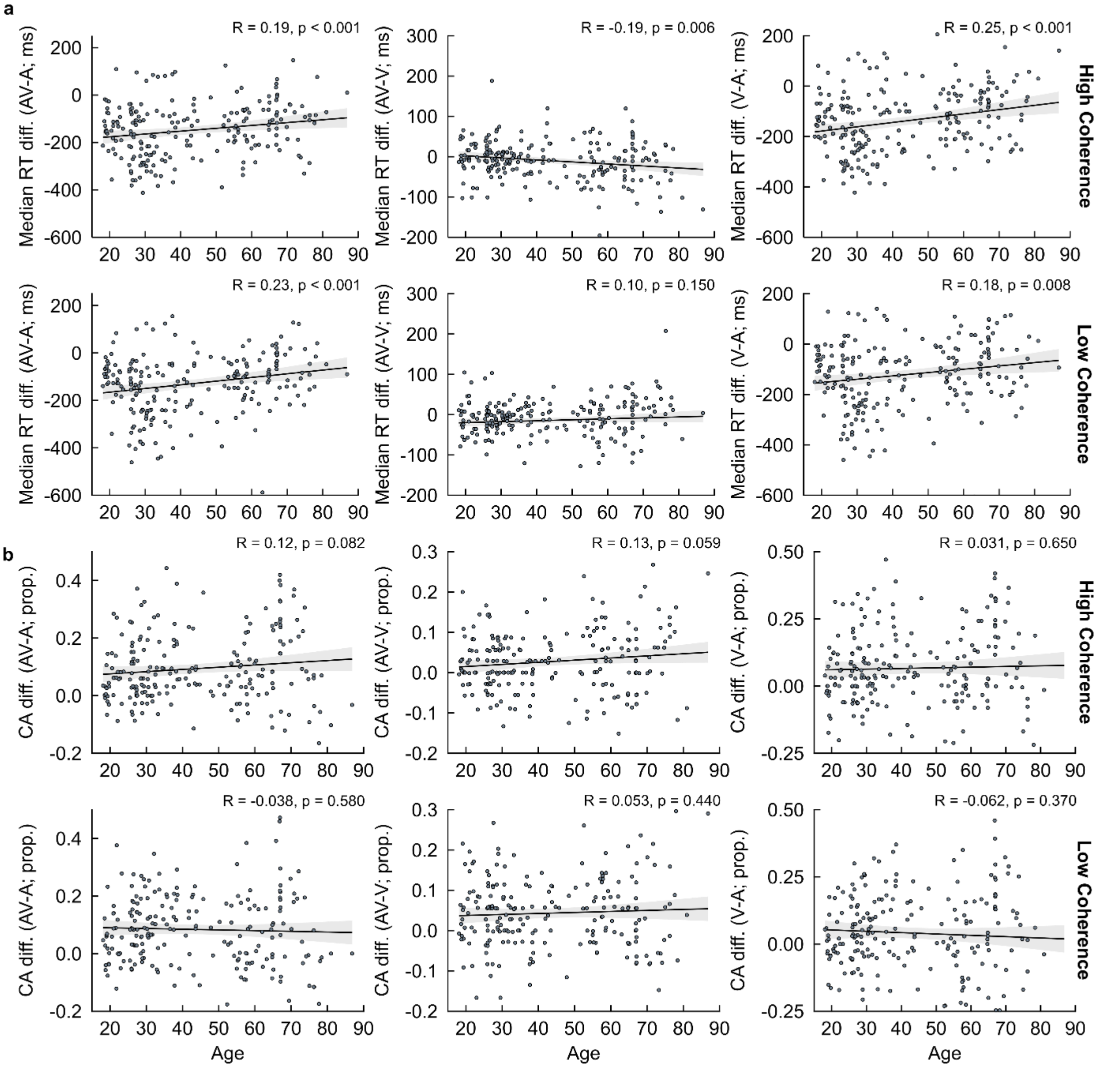
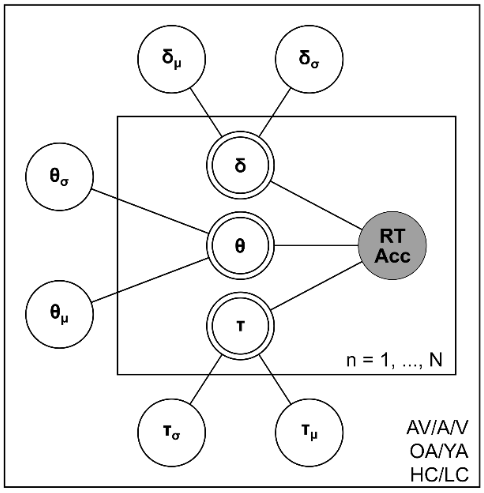


Figure 3.5 Performance Correlational Differences With Age Range
Individual Pearson's correlations of age with **a**, median RT differences and **b**, choice accuracy (proportion of correct responses) differences between AV and A trial types (*left*), between AV and V trial types (*middle*), between and V and A trials types, across the two levels of stimulus coherence (high/low coherence). Pearson Correlation Coefficients (R) and p-values are shown for each correlation. Shaded regions indicate 95% Confidence Intervals (CIs).

3.4.2 HDDM – Model Convergence

For all HDDM analyses, MCMC sampling routines (Gelman & Rubin, 1992) ran for 5 chains of 11000 iterations, each with a burn-in period of 1000 iterations and thinning factor of 2. For each variant, model convergence was assessed by examination of the posterior samples and the computed Gelman-Rubin $\hat{R}$ convergence diagnostics between chains for simultaneously estimated participant and group-level parameters. As previously outlined, Gelman-Rubin $\hat{R}$ convergence diagnostics are a measure of convergence among multiple MCMC chains and indicate that posterior density estimates and have converged to a fixed estimate. A $\hat{R}$ convergence diagnostic between 0.98 and 1.02, and closer to 1, indicates that chains with different starting values have converged to the same posterior estimate. This was confirmed for all eight modelling variants, with successful convergence confirmed by MCMC errors of all parameters for each variant smaller than 0.01.

a



b

Modelling Variant	Parameters Free To Vary	DIC
1	-	19568.012
2	δ	16889.597
3	θ	17075.272
4	τ	7404.423
5	δ, θ	15300.031
6	δ, τ	6184.432
7	θ, τ	6849.456
8	**δ, θ, τ**	**5719.890**

Figure 3.6 HDDM – Framework and Convergence Results. a, Graphical representation illustrating the Bayesian hierarchical framework for estimating behavioural HDDM parameters. Round nodes represent continuous random variables, with shaded nodes representing recorded variables, i.e., single-trial behavioural performance data (RTs and choice accuracy). 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, …, N) and the outer plate is over conditional dependencies for sensory trial type (Audiovisual; AV, Auditory; A, Visual; V), age range (Older Adults; OA, Younger Adults, YA), and stimulus coherence (High Coherence; HC, Low Coherence; LC). Combinations of drift rate (δ), decision boundary (θ), and non-decision time (τ) parameters are modelled as random variables with inferred means μ and variances σ, and are constrained by inferred estimates over all conditional dependencies. **b**, Deviance Information Criterion (DIC) scores and fixed parameters (i.e., free to vary across all conditional dependences) for all modelling variants tested. Shaded row denotes the best fitting Bayesian hierarchical modelling framework (i.e., lowest DIC score) towards our behavioural dataset.

After assessing modelling convergence, we performed a quantitative comparison of all variants by computing each variant's associated Deviance Information Criterion (DIC) score. The DIC assesses the goodness-of-fit for modelling variants to the applied dataset and penalises for its overall complexity (i.e., number of parameters). Modelling variants with a lower DIC score are to be preferred to those with a higher DIC, indicating the most parsimonious explanation of the dataset. Figure 3.5a outlines the DICs for each modelling variant. For our HDDM analysis, the modelling variant that best described the data (i.e., the model with the lowest DIC score) was the three-parameter model (Model 8) that sampled drift rate (δ), decision boundary (α), and non-decision time (τ) parameters for each participant (Model 8, $DIC_{\substack{\alpha\\\tau}}^{\delta} =$ 5719.890). In addition, a difference in DICs greater than 10 indicates substantial evidence that the model with the lower DIC is a better fit (Anderson & Burnham, 2004; Burnham & Anderson, 1998; Spiegelhalter et al., 2002). Because the difference between the modelling variant with the lowest DIC (Model 8, $DIC_{\substack{\alpha\\\tau}}^{\delta} =$ 5719.890) and the modelling variant with the second lowest DIC (Model 6, $DIC_{\tau}^{\delta} =$ 6184.432) exceeds 10 ($\Delta DIC = -464.542$), we consider this substantial evidence that Model 8 should be considered the most parsimonious account of the dataset.

Finally, to ensure there were no systematic discrepancies between the empirical behavioural dataset and the posited model, we performed a posterior predictive check, simulating a behavioural dataset with the fitted model, and comparing with the empirical dataset to illustrate that the posited model was a good fit. Figure 3.5 illustrates that a simulated behavioural dataset based on the best fitted modelling variant (i.e. Model 8, $DIC_{\substack{\alpha\\\tau}}^{\delta} =$ 5719.890) was consistent with the empirical behavioural dataset, namely that the empirical behavioural dataset metrics were within the 90% highest density interval (HDI) of the distributions and quantiles of simulated behavioural dataset metrics (Turkkan & Pham-Gia, 1993). Therefore, further analyses focused on this modelling variant.

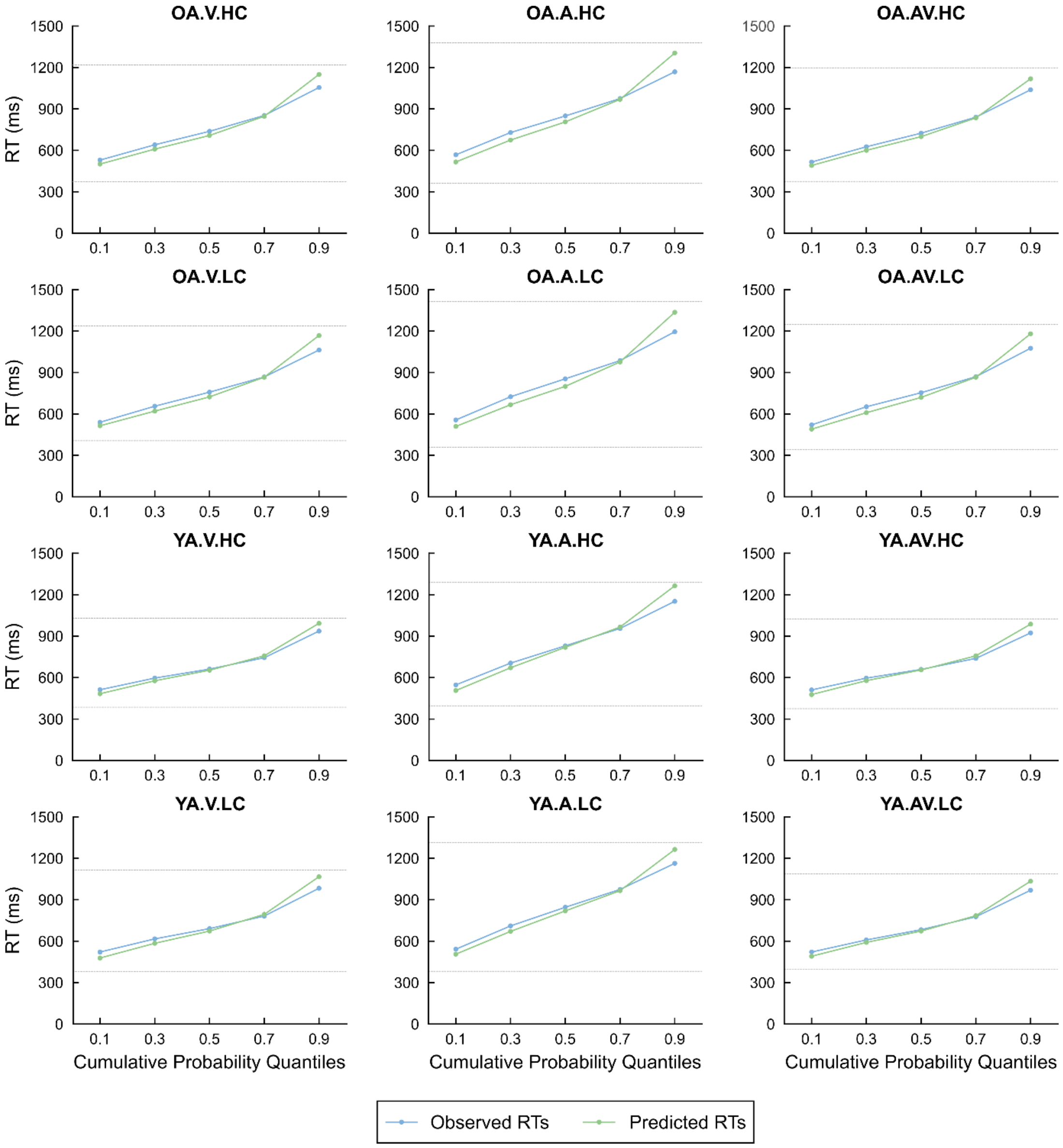


Figure 3.7 HDDM – Highest Density Region Intervals. Highest (posterior) Density Regions (HDRs) of the best fitting HDDM for the behavioural dataset. Modelling fit to the behavioural dataset was assessed using a cumulative quantile-probability plot, showing quantiles of RT distributions split across all conditional dependencies. Cumulative probability quantiles are plotted along the x-axis for observed RTs (blue) and predicted RTs (green), i.e., simulated RTs from HDDM posterior predictive estimates. Dashed grey lines on the y-axis denote the credible intervals representing the HDRs at a 90% confidence level.

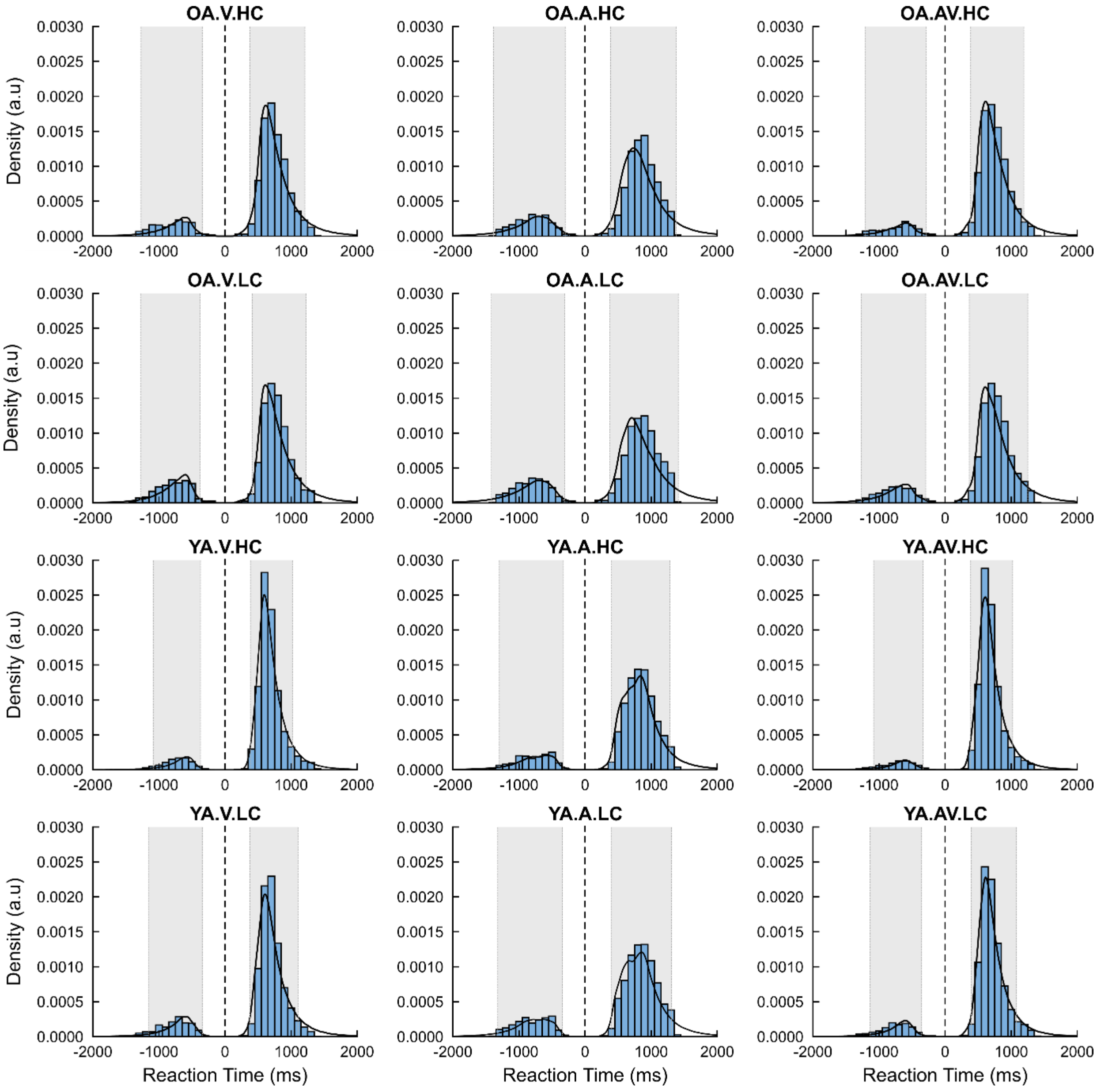


Figure 3.8 HDDM – Posterior Predictive Checks. Posterior predictive checks of the best fitting HDDM for the behavioural dataset. Modelling fit to the behavioural dataset was assessed using combined histogram and density plots for RT distribution splits across all conditional dependences and accuracy of choice responses (positive and negative RT histogram/density plot distributions for correct and incorrect choice responses respectively). Histograms (in blue) and density plots (in black) denote observed and predicted RTs respectively. Dashed grey lines on the x-axis denote the credible intervals representing the HDRs at a 90% confidence level for correct and incorrect choice responses.

3.4.3 HDDM – Behaviourally-Informed Cognitive Modelling Results

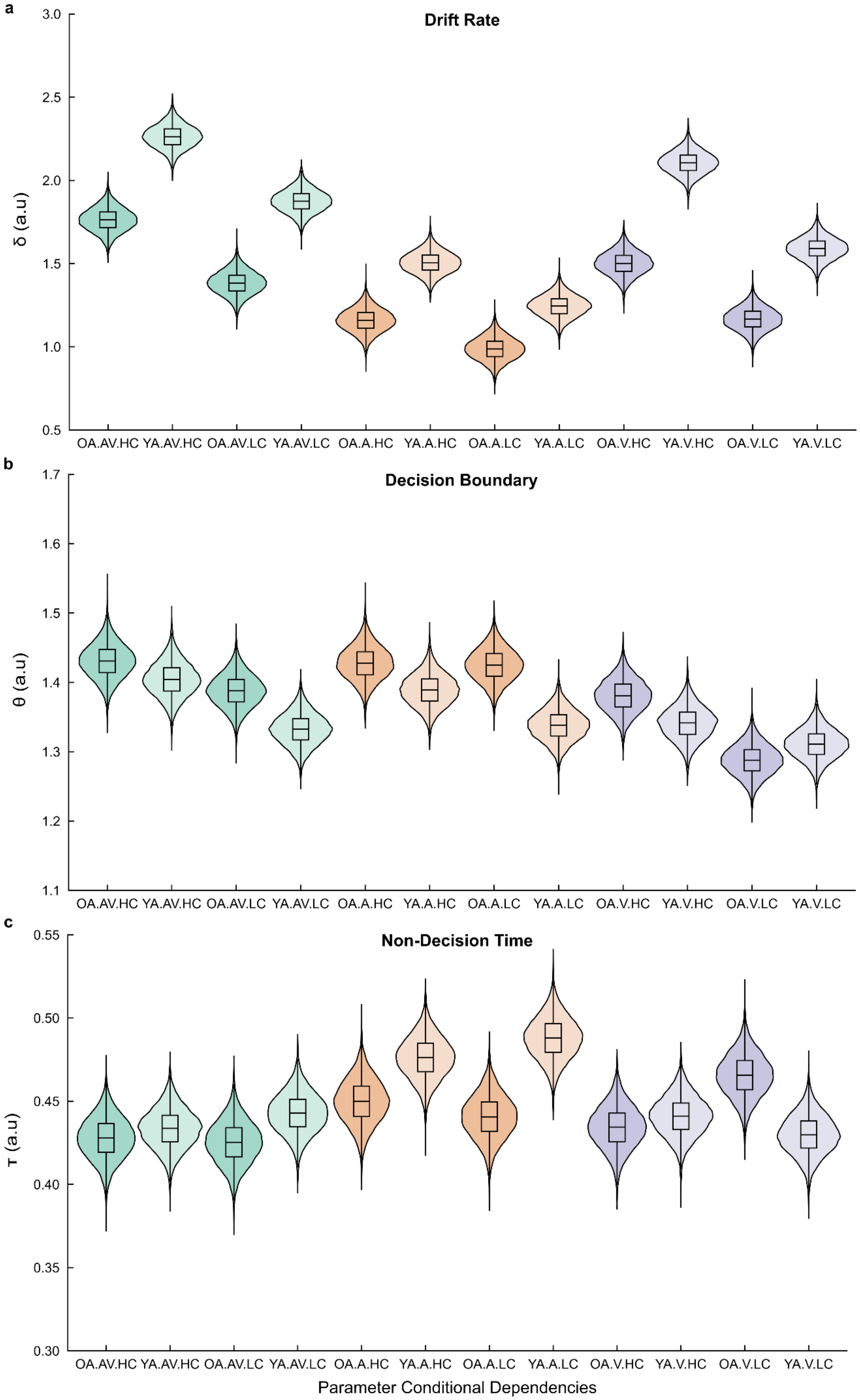


Figure 3.9 HDDM – Results. Group posterior density boxplots and violin plots for **a**, drift rate (δ), **b**, decision boundary (θ), and **c**, non-decision time (τ) parameters across all conditional dependencies (sensory trial type: Visual, V; Auditory, A; Audiovisual, AV; stimulus coherence: High Coherence, HC; Low Coherence, LC; age range: Older Adults, OA; Younger Adults, YA). All posterior density plots are derived from the behavioural HDDM, including N = 212 participants and 41765 trials.

Tables 3.1, 3.2, and 3.3 outline the hypotheses testing results across all conditional dependencies. We provide summaries of the hypothesis testing for each of our predictor variables (Age Range, Sensory Modality, Stimulus Coherence) below.

3.4.3.1 Age Range

Drift rate (δ) estimations were higher for YAs versus OAs across all sensory trial type and stimulus coherence conditional dependencies (Visual/High Coherence: P($\delta_{YA} > \delta_{OA}$) = > 0.999, log-odds = > 6.660; Visual/Low Coherence: P($\delta_{YA} > \delta_{OA}$) = > 0.999, log-odds = > 6.660; Auditory/High Coherence: P($\delta_{YA} > \delta_{OA}$) = > 0.999, log-odds = > 6.660; Auditory/Low Coherence: P($\delta_{YA} > \delta_{OA}$) = 0.996, log-odds = 5.488; Audiovisual/High Coherence: P($\delta_{YA} > \delta_{OA}$) = > 0.999, log-odds = > 6.660; Audiovisual/Low Coherence: P($\delta_{YA} > \delta_{OA}$) = > 0.999, log-odds = > 6.660). These results are consistent with the significantly lower RTs and higher proportions of correct responses observed between YAs and OAs across all sensory trial types (Figures 9, 10, and 11). Since the drift rate parameter captures the average rate of sensory evidence accumulation, these results implicate that younger adults are able to accumulate sensory evidence at a faster rate than older adults.

Decision boundary (θ) estimations were lower for YAs versus OAs, but only for AV and A trial types of LC (Visual/High Coherence: P($\theta_{YA} < \theta_{OA}$) = 0.879, log-odds = 1.980; Visual/Low Coherence: P($\theta_{YA} < \theta_{OA}$) = 0.229, log-odds = -1.212; Auditory/High Coherence: P($\theta_{YA} < \theta_{OA}$) = 0.875, log-odds = 1.949; Auditory/Low Coherence: P($\theta_{YA} < \theta_{OA}$) = 0.996, log-odds = 5.488; Audiovisual/High Coherence: P($\theta_{YA} < \theta_{OA}$) = 0.775, log-odds = 1.235; Audiovisual/Low Coherence: P($\theta_{YA} < \theta_{OA}$) = 0.952, log-odds = 2.997). Given we observed lower RTs and higher choice accuracy respectively for YAs compared to OAs, coupled with significant behavioural benefits between AV versus A trial types for YAs, this implicates an age-related difference in the pre-set quantities of sensory evidence required to be accumulated to facilitate choices for AV and A trial types of decreased stimulus salience (i.e., increased categorisation difficulty).

Non-decision time (τ) estimates were: (1) lower for YAs compared to OAs only for V trial types, and (2) higher for YAs compared to OAs only for A trial types, within LC trial types (Visual/High Coherence: P($\tau_{YA} < \tau_{OA}$) = 0.352, log-odds = -0.609; Visual/Low Coherence: P($\tau_{YA} < \tau_{OA}$) = 0.979, log-odds = 3.846; Auditory/High Coherence: P($\tau_{YA} < \tau_{OA}$) = 0.076, log-odds = -2.495; Auditory/Low Coherence: P($\tau_{YA} < \tau_{OA}$) = 0.006, log-odds = -5.158; Audiovisual/High Coherence = 0.374, log-odds = -0.516; Audiovisual/Low Coherence = 0.156, log-odds = -1.686). 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. These results implicates that YAs have a reduced duration for non-decisional processing of V trial types of low evidence salience (i.e., increased categorisation difficulty), but an increased duration for non-decisional processing of A trial types of low evidence salience, relative to OAs. Interestingly, no prevalent differences were found between OAs and YAs for AV trial types were found. Coupled with the different trajectories of non-decisional processing between OAs and YAs, this result underlies the comparative multisensory benefits exhibited between multisensory (i.e., AV) and unisensory (i.e., V and A) trial types on RTs, in which OAs benefit more from non-decisional processing speed from complementary A evidence, and YAs benefit more from non-decisional processing speed from complementary V evidence.

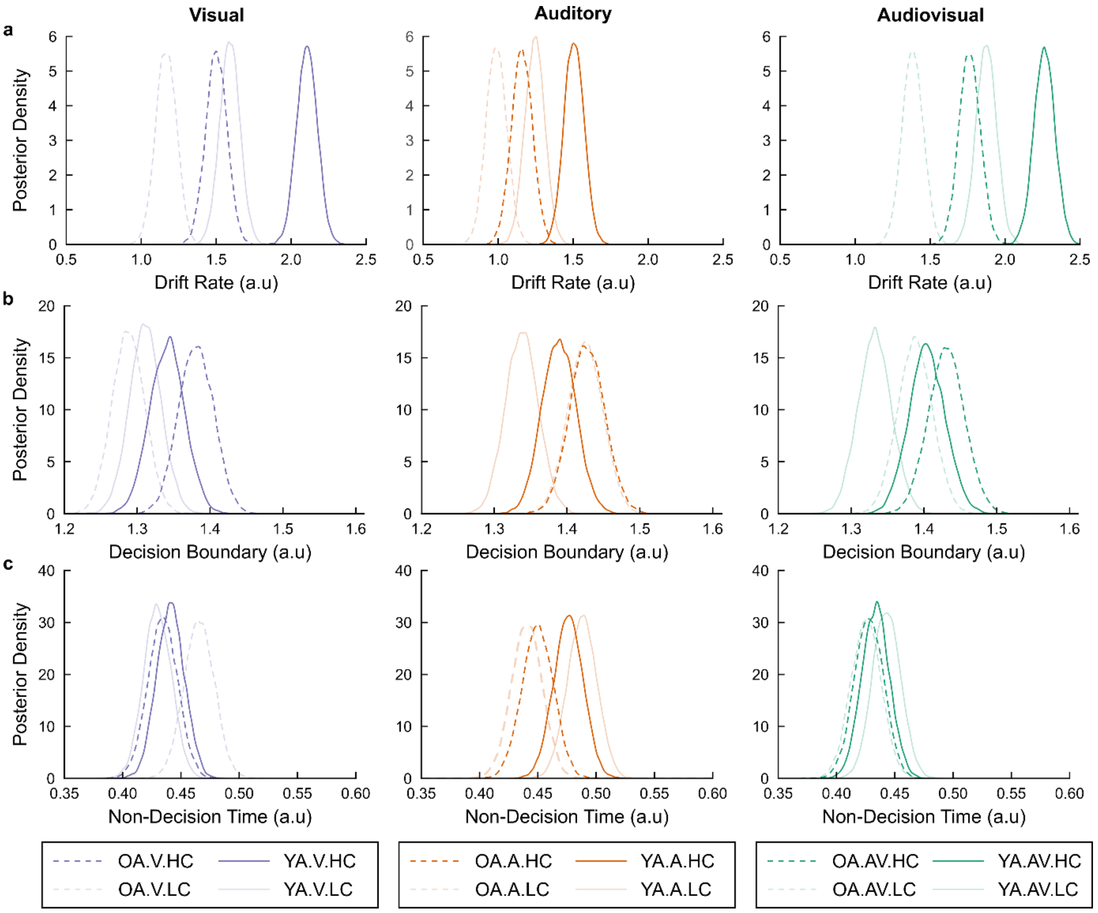


Figure 3.10 HDDM – Age Range Results. Group posterior density distributions illustrating ageing conditional dependencies for **a**, drift rate (δ), **b**, decision boundary (θ), and **c**, non-decision time (τ) parameters across sensory modality conditional dependencies (Visual, V; Auditory; A; Audiovisual; A). All posterior density distributions are derived from the behavioural HDDM, including N = 212 participants and 41765 trials.

3.4.3.2 Sensory Trial Type

Drift rate (δ) estimations were higher for AV versus A trial types (Older Adults/High Coherence: P($\delta_{AV} > \delta_A$) = > 0.999, log-odds = > 6.660; Older Adults/Low Coherence: P($\delta_{AV} > \delta_A$) = > 0.999; Younger Adults/High Coherence: P($\delta_{AV} > \delta_A$) = > 0.999, log-odds = > 6.660; Younger Adults/Low Coherence: P($\delta_{AV} > \delta_A$) = > 0.999, log-odds = > 6.660) and V versus A trial types (Older Adults/High Coherence: P($\delta_V > \delta_A$) = > 0.999, log-odds = >6.660; Older Adults/Low Coherence: P($\delta_V > \delta_A$) = 0.966, log-odds = 3.343; Younger Adults/High Coherence: P($\delta_V > \delta_A$) = > 0.999, log-odds = >6.660; Younger Adults/Low Coherence: P($\delta_V > \delta_A$) = > 0.999, log-odds = > 6.660) across all conditional dependencies. Between AV versus A trial types, drift rate estimations were higher for all conditional dependencies except for YAs within HC trial types (Older Adults/High Coherence: P($\delta_{AV} > \delta_V$) = 0.996, log-odds = 5.538; Older Adults/Low Coherence: P($\delta_{AV} > \delta_V$) = 0.986; Younger Adults/High Coherence: P($\delta_{AV} > \delta_V$) = 0.948, log-odds = 2.902; Younger Adults/Low Coherence: P($\delta_{AV} > \delta_V$) = 0.997, log-odds = 6.570). These results are consistent with the significantly shorter RTs and higher proportions of correct responses observed for AV trial types compared to V trial types, as well as the significant interactions between reduced sensory modality (i.e., AV versus V trial types) and stimulus coherence and age range conditions respectively. These results implicate that the rate of visual sensory evidence accumulation is enhanced when presented simultaneously with complementary auditory information, except in YAs with high stimulus coherence (i.e., lower categorisation difficulty). Older adults, however, still display enhanced multisensory integration, and thus benefits towards perceptual decision formation between AV and V RTs, which is highlighted by the significant interaction, and greater RT difference exhibited.

For decision boundary (θ) estimations, we found (1) higher estimations for AV versus V trial types for OAs for LC trial types, and (2) higher estimations for AV versus V trial types for YAs in HC trial types (Older Adults/High Coherence: P($\theta_{AV} > \theta_V$) = 0.925, log-odds = 2.515; Older Adults/Low Coherence: P($\theta_{AV} > \theta_V$) = 0.999, log-odds = 6.660; Younger Adults/High Coherence: P($\theta_{AV} > \theta_V$) = 0.966, log-odds = 3.355; Younger Adults/Low Coherence: P($\theta_{AV} > \theta_V$) = 0.751, log-odds = 1.106). In comparison, no prevalent differences in decision boundary estimates were found between audiovisual (AV) stimuli, compared to auditory (A) stimuli across any conditional dependencies (Older Adults/High Coherence: P($\theta_{AV} > \theta_A$) = 0.537, log-odds = 0.150; Older Adults/Low Coherence: P($\theta_{AV} > \theta_A$) = 0.135, log-odds = -1.855; Younger Adults/High Coherence: P($\theta_{AV} > \theta_A$) = 0.673, log-odds = 0.721; Younger Adults/Low Coherence: P($\theta_{AV} > \theta_A$) = 0.431, log-odds = -0.277). Given we further observed higher decision boundary estimates for V versus A trial types only for OAs for LC trial types ((Older Adults/High Coherence: P($\theta_V > \theta_A$) = 0.083, log-odds = -2.401; Older Adults/Low Coherence: P($\theta_V > \theta_A$) = > 0.999, log-odds = > 6.660; Younger Adults/High Coherence: P($\theta_V > \theta_A$) = 0.071, log-odds = -2.567; Younger Adults/Low Coherence: P($\theta_V > \theta_A$) = 0.191, log-odds = -1.440), these results implicate that OAs display increased caution in decision policy for AV versus V trial types when complementary A evidence is of decreased salience (i.e., higher categorisation difficulty), yet YAs display a similar increased response caution when complementary A evidence is of increased salience (i.e., lower categorisation difficulty). We observed in our behavioural results that OAs displayed a greater multisensory benefit towards RTs between AV versus V trial types, coupled with no significant differences in choice accuracy between AV versus V trial types, and a significant reduced three-way interaction suggesting such multisensory benefits are impacted by LC trial types. Overall, this implicates that OAs preserve caution in choice responses when complementary sensory evidence is more difficult to categorise, yet YAs preserve caution in choice responses when complementary sensory evidence interferes with less difficult stimuli categorisations.

For non-decision time (τ) estimations, we found (1) lower estimations for AV versus V trial types for OAs in LC trial types (Older Adults/High Coherence: P($\tau_{AV} < \tau_V$) = 0.643, log-odds = 0.587; Older Adults/Low Coherence: P($\tau_{AV} < \tau_V$) = 0.986, log-odds = 4.258; Younger Adults/High Coherence: P($\tau_{AV} < \tau_V$) = 0.674, log-odds = 0.725; Younger Adults/Low Coherence: P($\tau_{AV} < \tau_V$) = 0.224, log-odds = -1.242), (2) lower estimations for AV versus A trial types for YAs across HC/LC trial types (Older Adults/High Coherence: P($\tau_{AV} < \tau_A$) = 0.882, log-odds = 2.010; Older Adults/Low Coherence: P($\tau_{AV} < \tau_A$) = 0.779, log-odds = 1.379 ; Younger Adults/High Coherence: P($\tau_{AV} < \tau_A$) = 0.992, log-odds = 4.877; Younger Adults/Low Coherence: P($\tau_{AV} < \tau_A$) = 0.995, log-odds = 5.246), and (3 lower estimations for V versus A trial types for YAs across HC/LC trial types (Older Adults/High Coherence: P($\tau_V < \tau_A$) = 0.796, log-odds = 1.362; Older Adults/Low Coherence: P($\tau_V < \tau_A$) = 0.087, log-odds = 2.346; Younger Adults/High Coherence: P($\tau_V < \tau_A$) = 0.978, log-odds = 3.782; Younger Adults/Low Coherence: P($\tau_V < \tau_A$) = 0.999, log-odds = 6.660). Interestingly, these results implicate that despite increased response caution in decision policy, OAs had a lower duration for encoding AV versus V information of decreased stimulus salience, whereas YAs saw a further benefit encoding AV versus A information regardless of stimulus salience, enhancing benefits towards RTs and choice accuracy overall. This trend can be reaffirmed given YAs had a lower duration for encoding V versus A trial types (regardless of stimulus salience).

3.4.3.3 Stimulus Coherence

We found higher drift rate (δ) estimations for HC versus LC trial types across all sensory trial type and age range conditions (Older Adults/Visual: P($\delta_{HC} >$

δ_{LC}) = 0.999, log-odds = 6.660; Younger Adults/Visual: P($\delta_{HC} > \delta_{LC}$) = > 0.999, log-odds = > 6.660; Older Adults/Auditory: P($\delta_{HC} > \delta_{LC}$) = 0.957, log-odds = 3.091; Younger Adults/Auditory: P(($\delta_{HC} > \delta_{LC}$) = 0.998, log-odds = 6.064; Older Adults/Audiovisual: P($\delta_{HC} > \delta_{LC}$) = > 0.999, log-odds = > 6.660; Younger Adults/Audiovisual: P($\delta_{HC} > \delta_{LC}$) = > 0.999, log-odds = > 6.660). These findings are consistent with behavioural results illustrating significantly lower RTs for HC compared to LC trials, implicating decreases in the rate of sensory evidence accumulation with decreasing stimulus salience (i.e., increased task difficulty), and vice versa.

For decision boundary (θ) estimations, we found higher estimations for HC versus LC trial types for (a) OAs in V trial types, and (b) YAs in AV trial types (Older Adults/Visual: P($\theta_{HC} > \theta_{LC}$) = 0.998, log-odds = 6.213; Younger Adults/Visual: P($\delta_{HC} > \delta_{LC}$) = 0.826, log-odds = 1.556; Older Adults/Auditory: P($\theta_{HC} > \theta_{LC}$) = 0.529, log-odds = 0.114; Younger Adults/ Auditory : P($\theta_{HC} > \theta_{LC}$) = 0.944, log-odds = 2.832; Older Adults/Audiovisual: P($\theta_{HC} > \theta_{LC}$) = 0.892, log-odds = 2.113; Younger Adults/Audiovisual: P($\theta_{HC} > \theta_{LC}$) = 0.984, log-odds = 4.137). These findings indicate increased response caution was not a factor within AV or A trials for older adults, yet was a factor in AV trials for younger adults, and is consistent with general increases in choice accuracy observed for younger adults, but overall lack of significant findings for choice accuracy between OAs versus YAs within different stimulus coherence trial types.

For non-decision time (τ) estimations, we found lower estimations for HC versus LC trial types for OAs and V trial types only (Older Adults/Visual: P($\tau_{HC} < \tau_{LC}$) = 0.958, log-odds = 3.126; Younger Adults/Visual: P($\tau_{HC} < \tau_{LC}$) = 0.253, log-odds = -1.053; Older Adults/Auditory: P($\tau_{HC} < \tau_{LC}$) = 0.313, log-odds = -0.788; Younger Adults/Auditory: P($\tau_{HC} < \tau_{LC}$) = 0.746, log-odds = 1.078; Older Adults/Audiovisual: P($\tau_{HC} < \tau_{LC}$) = 0.445, log-odds = -0.223; Younger Adults/Audiovisual: P($\tau_{HC} < \tau_{LC}$) = 0.708, log-odds = 0.885). These findings indicate that stimulus salience (i.e., task difficulty) was modality-specific in decreasing the duration of non-decisional processes for older adults with visual information only.

3.4.3.4 Interim Summary of HDDM Results

We observed general age-related decreases in drift rate between OAs and YAs across all sensory trial types, demonstrating OAs accumulated sensory evidence more slowly than YAs. A comparison of drift rates for multisensory (i.e., AV) versus unisensory (i.e., V and A) trial types within OA and YA samples, however, demonstrated that OAs consistently exhibited increases in drift rate for multisensory versus unisensory trial types. Interestingly, YAs did not exhibit pronounced increases in drift rate for AV versus V trial types of increased stimulus salience (i.e., HC trial types), implicating the consolidation of complementary evidence from A trial types did not exhibit multisensory benefits towards perceptual decision formation.

Coupled with this were higher decision boundaries for AV versus A trial types of decreased stimulus salience (i.e., LC trial types) for OAs versus YAs, demonstrating that for general multisensory processing and specific unisensory processing, OAs had a higher response caution in decision policy, in which more sensory evidence needed to be accumulated to facilitate decisions. Finally, only OAs displayed lower non-decision times for AV versus V trial types of decreased stimulus coherence (i.e., LC trial types), implicating that they display an increased reliance on multisensory integration for encoding complementary A information. On the other hand, YAs displayed lower non-decision times for AV versus V trial types regardless of stimulus salience, implicating that benefitted greater from encoding complementary V information relative to OAs.

Conditional Dependencies			Bayesian Hypotheses Testing			
Sensory Modality	Stimulus Coherence	Age Range	Parameter	Hypothesis	Non-Overlap of Posterior Proportions	Logit-Odds Proportion
V	HC	OA versus YA	Drift Rate (δ)	$\delta_{YA} > \delta_{OA}$	**> 0.999**	**> 6.660**
	LC				**> 0.999**	**> 6.660**
	HC	OA versus YA	Decision Boundary (θ)	$\theta_{YA} < \theta_{OA}$ (*$\theta_{YA} > \theta_{OA}$)	0.879 (*0.121)	1.980 (*-1.980)
	LC				0.229 (*0.771)	-1.212 (*1.212)
	HC	OA versus YA	Non-Decision Time (τ)	$\tau_{YA} < \tau_{OA}$	0.352	-0.609
	LC				**0.979**	**3.846**
A	HC	OA versus YA	Drift Rate (δ)	$\delta_{YA} > \delta_{OA}$	**> 0.999**	**> 0.660**
	LC				**0.996**	**5.488**
	HC	OA versus YA	Decision Boundary (θ)	$\theta_{YA} < \theta_{OA}$ (*$\theta_{YA} > \theta_{OA}$)	0.875 (*0.125)	1.949 (*-1.949)
	LC				**0.996 (*0.004)**	**5.488 (*-5.488)**
	HC	OA versus YA	Non-Decision Time (τ)	$\tau_{YA} < \tau_{OA}$	0.076	-2.495 (*2.495)
	LC				**0.006 (*0.994)**	**-5.158 (*5.158)**
AV	HC	OA versus YA	Drift Rate (δ)	$\delta_{YA} > \delta_{OA}$	**> 0.999**	**> 6.660**
	LC				**> 0.999**	**> 6.660**
	HC	OA versus YA	Decision Boundary (θ)	$\theta_{YA} < \theta_{OA}$ (*$\theta_{YA} > \theta_{OA}$)	0.775 (*0.225)	1.235 (*-1.235)
	LC				**0.952 (*0.048)**	**2.997 (*-2.997)**
	HC	OA versus YA	Non-Decision Time (τ)	$\tau_{YA} < \tau_{OA}$	0.374	-0.516
	LC				0.156	-1.686

IV Conditional Dependencies			Bayesian Hypotheses Testing			
Sensory Modality	Stimulus Coherence	Age Range	Parameter	Hypothesis	Non-Overlap of Posterior Proportions	Logit-Odds Proportion
AV versus V	HC	OA	Drift Rate (δ)	$\delta_{AV} > \delta_V$	**0.996**	**5.538**
	LC				**0.986**	**4.229**
	HC	YA			0.948	2.902
	LC				**0.997**	**6.570**
	HC	OA	Decision Boundary (θ)	$\theta_{AV} < \theta_V$ (*$\theta_{AV} > \theta_V$)	0.075 (*0.925)	-2.515 (*2.515)
	LC				0.001 (***0.999**)	-6.660 (***6.660**)
	HC	YA			0.034 (***0.966**)	-3.355 (***3.355**)
	LC				0.249 (*0.751)	-1.106 (*1.106)
	HC	OA	Non-Decision Time (τ)	$\tau_{AV} < \tau_V$	0.643	0.587
	LC				**0.986**	**4.258**
	HC	YA			0.674	0.725
	LC				0.224	-1.242
AV versus A	HC	OA	Drift Rate (δ)	$\delta_{AV} > \delta_A$	**> 0.999**	**> 6.660**
	LC				**> 0.999**	**> 6.660**
	HC	YA			**> 0.999**	**> 6.660**
	LC				**> 0.999**	**> 6.660**
	HC	OA	Decision Boundary (θ)	$\theta_{AV} < \theta_A$ (*$\theta_{AV} > \theta_A$)	0.463 (*0.537)	-0.150 (*0.150)
	LC				0.865 (*0.135)	1.855 (*-1.855)
	HC	YA			0.327 (*0.673)	-0.721 (*0.721)
	LC				0.569 (*0.431)	0.277 (*-0.277)
	HC	OA	Non-Decision Time (τ)	$\tau_{AV} < \tau_A$	0.882	2.010
	LC				0.799	1.379
	HC	YA			**0.992**	**4.877**
	LC				**0.995**	**5.246**
V versus A	HC	OA	Drift Rate (δ)	$\delta_V > \delta_A$	**> 0.999**	**> 6.660**
	LC				**0.966**	**3.343**
	HC	YA			**> 0.999**	**> 6.660**
	LC				**> 0.999**	**> 6.660**
	HC	OA	Decision Boundary (θ)	$\theta_V < \theta_A$ (*$\theta_V > \theta_A$)	0.917 (*0.083)	2.401 (*-2.401)
	LC				**< 0.001 (* > 0.999)**	**< -6.660 (*> 6.660)**
	HC	YA			0.929 (*0.071)	2.567 (*-2.567)
	LC				0.809 (0.191)	1.440 (-1.440)
	HC	OA	Non-Decision Time (τ)	$\tau_V < \tau_A$	0.796	1.362
	LC				0.087	2.346
	HC	YA			**0.978**	**3.782**
	LC				**0.999**	**6.660**

IV Conditional Dependencies			Bayesian Hypotheses Testing			
Sensory Modality	Stimulus Coherence	Age Range	Parameter	Hypothesis	Non-Overlap of Posterior Proportions	Logit-Odds Proportion
V	HC versus LC	OA	Drift Rate (δ)	$\delta_{HC} > \delta_{LC}$	**0.999**	**6.660**
		YA			**> 0.999**	**> 6.660**
	HC versus LC	OA	Decision Boundary (θ)	$\theta_{HC} < \theta_{LC}$ (*$\theta_{HC} > \theta_{LC}$)	**0.002 (*0.998)**	**-6.213 (*6.213)**
		YA			0.174 (*0.826)	1.556
	HC versus LC	OA	Non-Decision Time (τ)	$\tau_{HC} < \tau_{LC}$	**0.958**	**3.126**
		YA			0.253	-1.083
A	HC versus LC	OA	Drift Rate (δ)	$\delta_{HC} > \delta_{LC}$	**0.957**	**3.091**
		YA			**0.998**	**6.064**
	HC versus LC	OA	Decision Boundary (θ)	$\theta_{HC} < \theta_{LC}$ (*$\theta_{HC} > \theta_{LC}$)	0.471 (*0.529)	-0.114 (*0.114)
		YA			0.056 (*0.944)	-2.832 (*2.832)
	HC versus LC	OA	Non-Decision Time (τ)	$\tau_{HC} < \tau_{LC}$	0.313	-0.788
		YA			0.746	1.078
AV	HC versus LC	OA	Drift Rate (δ)	$\delta_{HC} > \delta_{LC}$	**> 0.999**	**> 6.660**
		YA			**> 0.999**	**> 6.660**
	HC versus LC	OA	Decision Boundary (θ)	$\theta_{HC} < \theta_{LC}$ (*$\theta_{HC} > \theta_{LC}$)	0.108 (*0.892)	-2.113 (2.113)
		YA			**0.016 (*0.984)**	**-4.137 (*4.137)**
	HC versus LC	OA	Non-Decision Time (τ)	$\tau_{HC} < \tau_{LC}$	0.445	-0.223
		YA			0.708	0.885

Tables 3.1, 3.2, and 3.3 HDDM Hypothesis Testing Results. HDDM Bayesian hypothesis testing results for age range, sensory trial type, and stimulus coherence conditions respectively. Shaded Hypothesis testing results indicate predictive effects (P() > 0.950) of corresponding hypotheses.

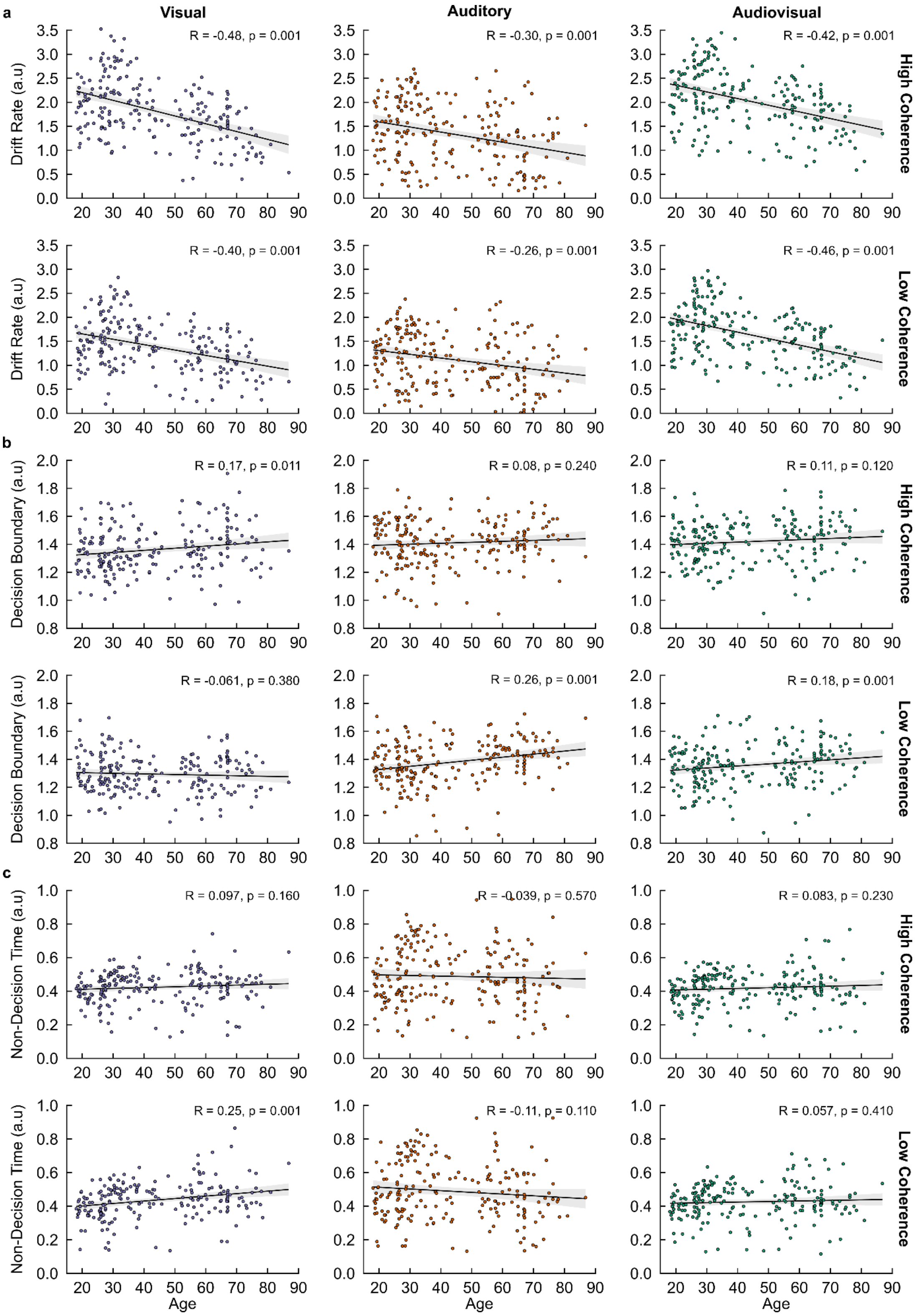


Figure 3.11 HDDM – Posterior Parameter Correlations with Age Range. Individual Pearson correlations of age with **a**, drift rate (δ), **b**, decision boundary (θ), and **c**, non-decision time (τ) parameters across all conditional dependencies (sensory modality: Visual, V; Auditory, A; Audiovisual, AV; stimulus coherence: High Coherence, HC; Low Coherence, LC; age range: Older Adults, OA; Younger Adults, YA). Pearson Correlation Coefficients (R) and p-values are shown for each correlation. Shaded regions indicate 95% Confidence Intervals (CIs).

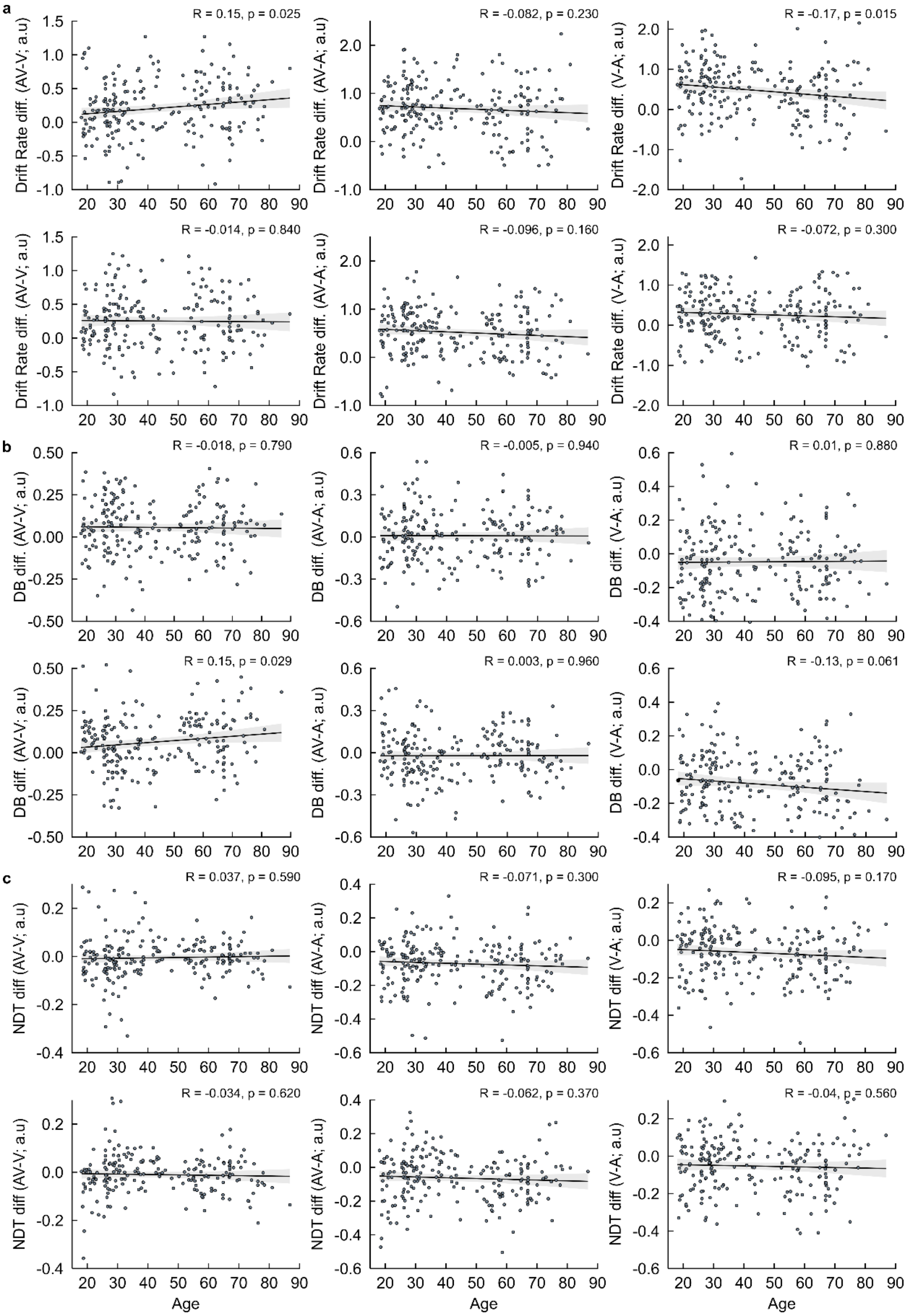


Figure 3.12 HDDM – Posterior Parameter Correlational Differences with Age Range. Individual Pearson's correlations of age with **a,** drift rate (δ), **b,** decision boundary (DB, θ), and **c,** non-decision time (NDT, τ) parameter estimate differences between AV and V trials (*left*), between AV and A trials (*middle*), and between V and A trials (*right*) across levels of stimulus coherence (high/low coherence). Pearson Correlation Coefficients (R) and p-values are shown for each correlation. Shaded regions indicate 95% Confidence Intervals (CIs).

3.5 Discussion

This study coupled an online variant of an audiovisual object categorisation paradigm (Franzen et al., 2020) with a behaviourally-informed HDDM to investigate age-related impacts on the latent psychological processes underlying multisensory perceptual decision-making. This methodology provided a principled account for characterising age-related modulations within the decision formation process. In particular, it permitted a recovery of parameters for modelling observed behavioural performance outcomes (i.e., RTs and binary responses), and mechanistically accounted for multiple alterations in behavioural outcomes suggesting impacts on the utilisation of multisensory information for forming perceptual decisions as we age. In particular, we observed significant alterations in perceptual decision-making between multisensory (i.e., AV) and unisensory (i.e., V and A) trial types, implying that older adults were able to preserve benefits from multisensory enhancements between AV versus V trial types towards choice formation, since we did not observe significant slowing of RT differences (i.e., AV – V Median RT differences) in OAs, as observed between AV versus A (i.e., AV – A Median RT differences) trial types, when compared to YAs. This preservation was identified despite observable age-related declines in decisional speed (i.e., RTs) and choice accuracy across all sensory trial types. In addition, we found a significant interaction between sensory trial types and age range, whereby OAs appeared to benefitted more from complementary audiovisual information compared to auditory, but not from visual information, alone. As mentioned, significant age-related decreases in choice accuracy were found, but not between sensory trial types. Arguably, this demonstrates declines in the accuracy of decisions, formed under time pressure, is not modality-specific, nor does it degrade multisensory integration within perceptual decision formation, implying a preservation of choice accuracy underlying multisensory integrative enhancements.

HDDM fittings towards participants' single-trial RTs and binary responses demonstrated parsimonious fits for reconciling the discrepancies in RTs and choice accuracy observed between younger and older adults for the rate of sensory evidence accumulation (i.e., drift rate), changes in decision caution policy (i.e., decision boundary), and duration of non-decisional processes (i.e., non-decision time). Thus, we could characterise age-related impacts determining the preservation of multisensory integrative benefits between multisensory and unisensory objects as we age. In particular, we observed four key trends: (1) slower rates of uptake in accumulating sensory evidence for OAs (i.e., lower drift rates. See Table 3.1; Figure 3.9; Figure 3.10) across all sensory trial types, irrespective of stimulus difficulty (i.e., stimulus coherence), (2) faster rates of uptake in accumulating sensory evidence between multisensory (i.e., AV) versus unisensory (i.e., V and A) trial types for both OAs and YAs, irrespective of stimulus coherence, (3) increases in response caution (i.e., higher decision boundaries. See Table 3.1, 3.2, Figure 3.9; Figure 3.10) for multisensory (i.e., AV) trial types of reduced evidence salience (i.e., LC stimulus coherence trial types); for both age ranges (i.e., OAs and YAs) and for OAs in comparison with YAs, and (4) a reduced duration of non-decisional processing (i.e., lower non-decision times) between AV versus V trial types for OAs with decreased evidence salience (i.e., LC stimulus coherence trial types), but not for YAs, who only exhibited a reduced duration of non-decisional processing between AV versus A trial types irrespective of stimulus difficulty (i.e., stimulus coherence).

In general, the key findings related to drift rate and decision boundary reaffirm previous observations of perceptual decision-making observed in older adults. First, we observed a decline in drift rate compared to younger adults across all trial types, reaffirming slower perceptual decision formation overall. This can be attributed to inherent age-related declines in sensory acuity and multiple domains of psychological functioning that perceptual decision-making underlies, observed as a function of age (Salthouse, 1996; 2009; see Section 3.2 *Introduction* for further citations) and underlie higher RTs observed for OAs versus YAs across all sensory trial types (see section 3.4.1 *Behavioural Results*). Second, OAs appeared to display decreased response caution when facilitating perceptual decisions with multisensory information of increased categorisation difficulty compared to YAs (Table 3.1; Figure 3.9; Figure 3.10). This finding reaffirms those from previous HDDM studies investigating age-related impacts on the behavioural indices of unisensory perceptual decision-making. Specifically, HDDM fittings to unisensory perceptual decision-making yielded an age-dependent widening of pre-set boundaries quantifying the extent of sensory information needed to facilitate perceptual decision formation (Ratcliff et al., 2001; 2006a; McKoon & Ratcliff, 2013; Scheib, Stoll, & Randerath, 2023; Servant & Evans, 2020; Starns & Ratcliff, 2010; Thapar et al., 2003). Interpretations between such complementary results imply an increase in response caution is increasingly likely as we age in order to compensate for reduced acuity in processing incoming sensory information, particularly when it is increasingly ambiguous to consolidate to inform our choice behaviours.

However, we did not observe age-dependent modulations in decision boundary for visual trial types. Instead, we observed age-dependent modulations in decision boundary for auditory and audiovisual trial types (see Table 3.1). Interestingly, this general widening can be interpreted as modality-specific, and thus can be impacted not only by ageing, but the modality activated to process incoming sensory information. Reaffirming this interpretation concerns the observation from our HDDM results that OAs had larger decision boundaries for AV versus V trial types, but not AV versus A trial types, of decreased evidence salience, whereby complementary information that is difficult to interpret (i.e., A trial types) is treated with a cost in perceptual decision formation to preserve underlying choice accuracy (see Table 3.1; Figure 3.9; Figure 3.10). As such, the general observation of an age-related increased in response caution in decision policies should be treated cautiously, as it was not consistent across both multisensory (i.e., AV) and unisensory (i.e., V and A) trial types, and perhaps underlies the inconsistencies between age-related slowing of information processing and hypotheses demonstrating preserved benefits of multisensory integration within perceptual decision formation (Figure 3.12). A further notion to consider concerns the acuity of the visual and auditory modalities. Given all included participants self-reported normal vision and/or hearing, a comparison with OAs who display deficits in visual and/or auditory acuity could reaffirm our assumption that the widening of response boundaries across multisensory and modality-specific unisensory conditions could further our understanding of whether increased response caution in decision policy is sensory-driven or modulated by post-sensory decision dynamics, as evidenced by our significant interactions for AV versus A, but not AV versus V sensory trial types.

Alternatively, the notion that older adults typically display general age-dependent increases in response caution as a compensatory mechanism can be challenged. Given the prevalent difficulties in ascertaining why bidirectional age-related modulations in drift rate and decision boundary are observed for unisensory paradigms, it implicates that differences in the readout of multisensory information differs depending on the context of the decision facilitation process not otherwise seen with unisensory information (Kelly & O’Connell. 2015). In consideration of the hypotheses previously outlined (see section 3.2 *Introduction*), preserved multisensory integration for objects that are more difficult to process may arise from the principle of inverse effectiveness (Holmes, 2009; Meredith & Stein, 1986a; 1986b; Stein & Stafford, 2008), and/or deficits in attentional control; either in higher-order selective attention, or lower-order “baseline” unisensory filtering levels (Hugenschmidt et al., 2009a; 200b). Overall, there seems to be a modulation of multisensory integrative levels compensating where unisensory processing performance is poorest. This interpretation is further reinforced by increases in the duration of sensory encoding and/or motor response production latency (i.e., increased non-decision times) for older adults, compared to younger adults, across unisensory conditions (i.e., V and A; see Table 3.1). This suggests declines in sensory systems, resulting in degraded function, and longer processing times. Similarly, we further observe that younger adults did not exhibit multisensory benefits through increased drift rate differences between AV versus V object types of increased evidence salience (i.e., HC trial types), whereas older adults did. It is possible that younger adults are less likely to exhibit multisensory benefits when complementary unisensory information does not weight the reliability of information already available, and decision-making performance is already near-optimal. Whereas a ceiling effect may underline why younger adults did not exhibit such benefits in this finding (i.e., as observed in chronological age correlations in which younger adults were more likely to ascertain full choice accuracy across unisensory and multisensory trial types; see Figure 3.3), it is possible that younger adults are less susceptible to manipulations in stimulus salience (i.e., task difficulty), thus reducing the likelihood of the principal of inverse effectiveness arising to account for deficits in attentional control in filtering complementary unisensory information.

Contradictory to these arguments, however, are the declines in non-duration time for AV versus V object types, implying that whereas increased noisy at unisensory baseline levels remains inherent; irrespective of stimulus object properties or attentional state (Fisher et al., 2023; Hernández et al., 2019; Rowe et al., 2006), the integration of information preserves enhancements towards multisensory perceptual decision-making, in which less cognitive demand is needed to process information from independent unisensory modalities. In consideration of such inferences, we contend that preserved benefits from multisensory integration are specific to the modality whereby information is more reliable, possibly inferring a preservation of reliability-weighting mechanisms (Ernst & Banks, 2002; Ernst & Bülthoff, 2004; Fetsch et al., 2012; Rohe & Noppeney, 2016) in older adults. Research supporting this interpretation comes from findings that applied alternative computational modelling methodologies (see Jones & Noppeney, 2021, for review). Park & Kayser (2020), for example, investigated whether age-related impacts on audiovisual spatial perception arose from reduced sensory precision with age or alterations to the process of causal inference when inferring sensorimotor cues. Using a single-trial audiovisual localisation paradigm, coupled with models of Bayesian causal inference, they analysed participants' abilities to localise auditory and visual stimuli, and their susceptibility to the ventriloquist effect (i.e., within multisensory trials) and the ventriloquist aftereffect (i.e., between multisensory trials) biasing their judgements (Alais & Burr, 2004). The findings of their model-based approach predicted that while older adults exhibited reduced precision in representing auditory information, they were not susceptible to follow-up biases in auditory spatial localisation. Their results therefore suggested that older adults benefitted from a recalibration in multisensory representations based on their previous decisional responses, not previous sensory representations. Our results reaffirm this interpretation, since we observed prominent evidence of age-related increases in response caution when complementary auditory information was less salient, and thereby increasingly difficult to precisely consolidate.

To assess the effect of reliability-weighting mechanisms, future replications of this study should consider replacing self-reporting of no known impairments to vision and hearing, as conducted in our study, with metrics for assessing sensory acuity. For example, visual acuity can be assessed using logMAR charts (Hazel & Elliott, 2002), whereas contrast sensitivity can be assessed using Pelli-Robson charts (Elliott, Sanderson, & Conkey, 1990). Recording of such metrics as an exclusion criterion could be considered in order to stabilise the degree of sensory acuity and reliability of participants across the adult lifespan. In addition, however, sensory acuity and/or sensitivity analyses with both categorical and chronological age, and subsequent relationship with HDDM parameters for conceptualising multisensory perceptual decision-making, could enable us to discern why we observe modality-specific differences between AV versus V and AV versus A sensory trial types between younger and older adults. Thus, we can probe if reliability-weighting is an inherent mechanism most prominently impacted in natural ageing and consolidate whether it is the inherent declines in sensory processing, or in post-sensory dynamics impacting the cognitive mechanisms for consolidating multisensory representations towards perceptual decision formation (Hirst et al., 2019; 2022).

A further recommendation for future replications of this study should exploit the effects of learning, specifically, how older adults react to trial-by-trial feedback on decisional speed and accuracy in categorising multisensory and unisensory trial types relative to younger adults. Previous DDM analyses of age-related impacts on unisensory perceptual decision-making, for example, have found that older adults typically begin experiments with slower and less accurate choice performance, but improve at a rate that exceeds those of younger adults. The effects of learning have found that they improve in proportion to the salience of the available sensory information, leading to decreased response caution in decision policy (i.e., lower decision boundaries) and increases in the uptake of sensory information (i.e., higher drift rates) in exploiting signals of increased evidence salience (Ratcliff, Thapar, & McKoon, 2006). However, further DDM findings have highlighted that older adults are more likely to compromise SATOs in order to minimize errors in trial-by-trial decision-making accuracy, and do not reduce their increased response caution in decision policy (i.e., higher decision boundaries) to benefit their Reward Rate Optimal Boundary (RROB), a single value metric for decision boundary parameters that captures the highest proportion of correct responses, when unisensory signals are of decreased evidence salience (Starn & Ratcliff, 2010). Conceptualising if such findings are generalisable to multisensory perceptual decision-making would yield insights into whether the preservation of multisensory benefits observed in our HDDM results are similarly impacted through trial-by-trial feedback, and subsequent associative learning.

In order to consolidate our interpretations for disentangling age-related impacts on multisensory perceptual decision-making, it would be beneficial to follow up on these preliminary findings by investigating age-related impacts on the neural signatures underlying decision formation. In particular, examining age-dependent modulations in EEG response profiles towards multisensory and unisensory stimuli using neurally-informed modelling approaches. Considering unisensory perceptual decision-making research, McGovern et al. (2018) investigated age-related effects on EEG activity of participants who engaged in continuous random dot motion discrimination and contrast-change detection paradigms. They extracted ERPs corresponding to CPP, and performed comparisons of behaviourally-informed and neurally-informed DDM models; modelling RTs/choice accuracy and CPP amplitudes respectively. Interestingly, they found that the fitting of the maximum amplitude, captured by the CPP, as regressors within the DDM predicted task-dependent slowing of accumulation-to-bound processes (i.e., lower drift rates) among older adults, as well as decreased inter-trial drift rate variability, suggesting neural mechanisms for sustained visual attention as a compensatory mechanism for deficits in unisensory processing. Similarly, Forstmann et al. (2011) investigated age-related degradations in the structure of pre-SMA and striatum; a region thought to determine decision policy adjustments, and therefore decision boundaries underlying response caution. By simultaneously incorporating diffusion weighted structural imaging and behavioural performance outcomes into a LBA model, they found a decreased flexibility in adjusting decision policies for SATOs in older adults correlated with a reduction in white matter integrity in the corticostriatal pathways. Finally, Jaworska et al. (2019) used an information theoretic framework to predict age-related modulations in EEG activity corresponding to the processing of facial features. They reported modulations in occipitotemporal regions that represented dynamic evidence accumulation of eyes in face-versus-noise images at approximately 170ms post-stimulus onset. Furthermore, their information theoretic framework reconciled their neural locus with an approximate 40ms increase in RTs, inferring an impact of sensory evidence accumulation not indicative of a slowing of visual active sensing.

Importantly, these studies recommend that constraining modelling fits towards ageing data with neural indices of perceptual decision-making may benefit *when* and *where* in the brain we observe separate decision processes for sensory evidence consolidation and motor response facilitation. In order to develop upon this research domain of application, we contend that utilising a neurally-informed modelling approach to provide parsimonious fits towards simultaneously collected behavioural and neuroimaging data may prove beneficial in linking our preliminary interpretations of preserved multisensory mechanisms with preservations of function in their corresponding neural substrate(s). In consideration of this, a study that recorded EEG signals underlying older and younger adults' behavioural performance in the audiovisual face-versus-car categorisation paradigm, and subsequent analysis utilising multivariate LDA to characterise the underlying neural dynamics, would provide the opportunity to incorporate a neurally-informed HDDM (Wiecki et al., 2013) to dissect the underpinning neural processes capturing the exhibited benefits of multisensory integration towards perceptual decision formation within older adults. In particular, a consolidation of the neural components capturing increases in drift rate between multisensory (i.e., AV) and unisensory (i.e., V and A) trial types, coupled with increased decision boundary and decreased non-decision time between AV and V trial types of low stimulus coherence, could capture the spatiotemporal trajectory of earlier sensory encoding and later changes to post-sensory decision dynamics, and modulatory differences as a function of age, further informing our insights into the resultant cognitive mechanisms preserved or degraded in preserving multisensory integrative benefits towards perceptual decision formation. In line with the hypotheses outlined in Chapter 1 (see subsection 1.2.3 *Where Do the Benefits of Multisensory Integration Arise during Perceptual Decision-Making*), an understanding of how the ageing brain modulates the integration and consolidation of sensory information across modalities, through capturing particular modulations to temporally earlier and later neural correlates, would highlight if the modulations in HDDM parameters arise from natural deficits to sensory systems (i.e., bottom-up processing) and/or decision-making cognition (i.e., top-down processing), or modulated activity attributed to a means of compensating for such deficits to preserve multisensory perceptual decision-making behaviour.

In conclusion, we demonstrated novel insights into the key computations determining preserved multisensory integration, and therefore preserved benefits in perceptual decision formation, in older adults. Notably, we conceptualised that despite an age-related slowing of sensory information processing for both multisensory and unisensory information, compensatory mechanisms for (a) the functions of unisensory modalities, and (b) decreased salience of stimulus properties may be apparent in older adults to ensure perceptual decision-making is not fully degraded. We therefore recommend further exploration of the relationship between sensory representation precision, stimulus effectiveness, and multisensory enhancements towards perceptual decision formation, both at a computational level and at a cortical level, to further disentangle the extent older adults may benefit from multisensory integration in choice formation (Mozolic et al., 2012).

Chapter 4

Investigating the Neuromodulatory Effects of Perceptual Learning on Multisensory Decision-Making

4.1 Abstract

Learning is associated with long-term improvements to an organism's perceptual abilities, including the formation of rapid low-level decisions. Neurobiological accounts of unisensory perceptual learning have implicated multiple timescales and locations for improvements in decisional behaviour, i.e., lower RTs and higher choice accuracy, arguing that alterations to neuronal plasticity may arise in sensory cortices; and are therefore associated with the strengthening of sensory representations, or associative cortices; and are therefore associated with a strengthening of decision dynamics in consolidating post-sensory decision evidence representations. Such unisensory perceptual learning accounts, however, are not compatible with localising where we may observe benefits to perceptual learning towards multisensory decision formation. Undoubtedly, considering the spatiotemporal locus of perceptual learning benefits towards multisensory perceptual decision-making has not been explored in concordance with unisensory perceptual learning accounts. To address this, we coupled an audiovisual object categorisation paradigm with EEG to (a) analyse the neural activity underpinning the ability of participants to categorise faces and cars in embedded visual, auditory, and audiovisual stimuli, and (b) assess the effects of learning and experience on decision-making performance (i.e., RTs and choice accuracy) from completing the paradigm over three consecutive days. While we observed generalisable benefits towards perceptual decision formation from completing the study over three days, through significantly decreased RTs and higher choice accuracy, we did not observe any significant differences between categorisations of multisensory and unisensory stimuli. A methodological issue was identified with the application of multivariate LDA, in which there was a lack of sufficient statistical power in the classifier's ability to identify categorisations of faces and cars in the embedded stimuli. An alternative analysis, however, decoding categorisations between multisensory (i.e., AV) and unisensory (i.e., V and A trial types), did uncover neural components suggesting earlier sensory and later post-sensory neural dynamics underpinning the beneficial effects of multisensory integration towards perceptual decision formation. Interestingly, the spatiotemporal locus of the identified components varied depending on the

unisensory modality providing the complementary evidence to be integrated. We conclude our findings discussing the potential implications of why we did not observe our results corresponding to our hypotheses, and recommend alterations to the paradigm for how to improve the LDA classification performance in order to further our understanding of the neuromodulatory effects of multisensory perceptual learning.

4.2 Introduction

Perceptual learning defines sustained, long-lasting, alterations to an organism's perceptual system which improves its ability to respond to its environment (Gibson, 1963; 1968; Seitz, 2017). Specifically, it defines enhancements in our ability to categorise, detect, and/or discriminate sensory stimuli (Gilbert, Sigman, & Crist, 2001; Law & Gold, 2008; Goldstone, 1998). Importantly, perceptual learning is inextricably related to both unisensory (see Gold & Ding, 2011; Philiastides, Diaz, & Gherman, 2017 for reviews) and multisensory perceptual decision-making (Powers, Hevey, & Wallace, 2012; Proulx, Brown, Pasqualotto, & Meijer, 2014). 2AFC paradigms; considered the *de facto* choice of paradigm in perceptual decision-making literature (see section 1.1.2 *Human Neuroimaging Research*), instruct participants to categorise, detect, and/or discriminate sensory stimulus presentations as quickly and as accurately as possible, thereby facilitating a response from a limited range of choice alternatives. The expected outcome of perceptual learning in 2AFC paradigms is that with training, practice and/or task-dependent experience, long-lasting improvements to decisional speed and accuracy improves during task participation, hence providing a benefit to the formation of perceptual decisions.

Despite the obvious ubiquity of perceptual learning benefitting the formation of decisions, the underlying neural plasticity still requires further research. Previous studies have provided insights into the neurobiological mechanisms of perceptual learning (see Philiastides, Diaz, & Gherman, 2017, for review), but remain conflicted over *when* and *where* in an organism's perceptual system we observe alterations to neural plasticity that underlie induced benefits towards decision formation. The residing notion is that learning-induced neuronal plasticity can occur over multiple timescales and regions in the brain (Fahle, 2005). In consideration of the theoretical framework for (unisensory) perceptual decision-making, traditional viewpoints posited that the influence of repeated exposure modulates earlier neural mechanisms in the sensory cortices (Seitz & Dinse, 2007). Findings demonstrating support for this long-standing hypothesis stem from studies identifying markers of neuronal plasticity in visual (Karni & Sagi, 1991; Sagi & Tanne, 1994), auditory (Bao, 2015; Irvine, 2018), and somatosensory (Bejjanki et al., 2011) cortices. Human neuroimaging research from Chen et al. (2015), for example, coupled visual 2AFC motion discrimination paradigm with fMRI to investigate the consequential effects of perceptual learning. They found improved choice accuracy over eight training days, with underlying enhancements in BOLD responses in area V3(A) of the visual cortex, as well as its effective connectivity to the IPS. They thus implicated that perceptual learning improves the functional interconnectivity of neurons in visual cortex for forced decision-making. Comparably, Jehee et al. (2012) measured BOLD activity of participants who engaged in delayed decision-making in a visual orientation discrimination paradigm. Their results demonstrated neural enhancements among early areas of the visual cortex for repeatedly induced orientations, implicating that a functional reorganisation of neural substrates within early sensory cortices underlies the basis for prolonged benefits to perceptual decision formation.

Alternatively, a more recent viewpoint posits that learning-induced neuronal plasticity does not modulate neural activity in sensory cortices, but rather, modulates neural activity in associative cortices. In other words, it does not alter representations of incoming sensory information, but how they are interpreted to infer the decision dynamics facilitating choice responses. A seminal study by Law & Gold (2008), for example, identified neural correlates in the middle temporal area (MT) and LIP of Macaque monkeys as they engaged in a RDK, which displayed different modulatory patterns as behavioural sensitivity to determining motion coherence increased through training. Specifically, motion-driven responses of LIP, but not MT, neurons correlated with this training-induced increase in behavioural sensitivity, suggesting inputs from sensory cortices did not strengthen the underlying neural pathways, but rather, it was the neural pathways for resultant sensory evidence representations being strengthened. Furthermore, an EEG study by Diaz et al. (2017) tested participants who completed the visual face-versus-car categorisation paradigm over three consecutive days, in order to identify the benefits towards visual perceptual decision-making from forced choice practice. Their results yielded two EEG components that corresponded with the time points of those uncovered from research by Philiastides et al. (Philiastides & Sajda, 2006a; 2006b; Philiastides, Ratcliff, & Sajda, 2006; see section 1.1.2 *Human Neuroimaging Research*). Specifically, they uncovered *Early* and *Late* components of perceptual decision formation at approximately 183ms and 431ms over occipitotemporal and frontroparietal brain regions respectively. Completion of the experiment over three consecutive days elicited the observation that participants displayed comparative improvements in decisional speed (i.e., lower RTs) and accuracy (i.e., higher proportion of correct responses), and was attributed to simultaneous enhancements in classification performance to the *Late*, but not the *Early*, component. Furthermore, the onset of the *Late* component was delayed in time, and attributed to increases in RTs), as the categorisation of faces versus cars became more difficult with decreasing evidence salience. The authors implied that their findings demonstrated a strengthening of the connections between early sensory encoding and downstream associative regions, driving modulatory improvements to the post-sensory decision dynamics as a product of perceptual learning.

Despite the contradictory findings inferring *where* and *when* we observe the neuromodulatory influences of perceptual learning towards decision formation, it can be concluded that perceptual learning is inextricably linked to perceptual decision-making. In order to further our understanding of the consequential effects to neural plasticity, an understanding of how perceptual learning is inextricable linked to multisensory perceptual learning may contribute to narrowing our overall understanding. Comparably, similar parallels can be considered between unisensory hypotheses of the nature and location of perceptual learning in the brain, and hypotheses of the nature and location of multisensory perceptual decision-making in the brain (see section 1.2.3 *Where Do the Benefits of Multisensory Integration Occur Within Perceptual Decision-Making?*). Put concisely, the spatiotemporal locus of multisensory perceptual learning may coincide with the spatiotemporal locus of benefits from perceptual learning.

In relation to early multisensory perceptual learning benefits, for example, Powers et al. (2012) analysed fMRI BOLD activity of participants who were trained to improve decision-making performance in an audiovisual simultaneity judgement training paradigm. They characterised regions of auditory and visual cortices that exhibited decreases in BOLD activity during testing on the second day, with dynamic causal modelling predicting an increase in resting state effective connectivity, which reduced the window for temporal binding (i.e., the *Temporal Binding Window*; TBW), thereby strengthening early bidirectional sensory connections for multisensory integration. Comparably, Ross et al. (2022) coupled fMRI with recordings of continuous narrative speech in order to inquire into the neural regions that showed audiovisual enhancement during articulatory speech perception. BOLD fMRI responses demonstrated functional connectivity of thalamic brain regions along the visual and auditory pathways with bilateral Superior Temporal Sulcus (STS) and amygdalae; regions notably associated with early multisensory integration in speech perception (Calvert & Thesen, 2004). However, further studies have instead inferred neural indices that indicate perceptual learning elicits later, post-sensory, multisensory benefits towards decision formation. For example, Yang et al. (2018) applied EEG to older and younger participants who engaged with an audiovisual spatiotemporal discrimination paradigm instructing them to categorise spatiotemporally congruent and incongruent presentations of auditory and visual stimuli. Their results highlighted greater P300 ERP amplitudes in older adults after spatiotemporal perceptual training; a temporal-parietal potential associated with selective attention (Polich, 2007), but no significant effects on P300 ERP amplitudes in younger adults. This implied that multisensory perceptual learning improves cognitive abilities, with greater modulatory effects on malleable regions of the ageing brain associated with post-sensory decision dynamics. Furthermore, Zilber et al. (2014) trained participants to discriminant visual colour coherence between RDKs that displayed visual only, audiovisual, and visual with auditory noise RDKs while recording brain activity using MEG. Selectivity of ventrolateral prefrontal cortex (vlPFC) responses were demonstrated across all three RDK variants towards learned coherence levels, but neural selectivity in the human Medial Temporal complex (hMT+) was only demonstrated in the audiovisual RDK variant, correlating with post-sensory decision-making performance. Their findings demonstrate empirical support for the *Reverse Hierarchy Theory of Learning* (Ahissar & Hochstein, 2004), which theories that supramodal processing is optimised from late decisional processes capitalizing on sensory invariant representations and feeding back globally towards sensory modalities.

Despite increasing interest, whether multisensory perceptual learning enhances multisensory integrative benefits, and if so how, remains an underexplored and conflicting field of research. Specifically, a focus on *where* we may observe multisensory perceptual learning benefits, in conjunction with the processes theorised in a multisensory perceptual decision-making framework, remains underexplored in the research field, and is limited by the wide variety of experimental paradigms and modelling variants utilised. In particular, this confounds the assumption of context invariance; defined as the assumption which states unisensory signal processing is independent of its embedment in the induced experimental context (i.e., instructions; Liu & Otto, 2020). As such, conflations of enhanced multisensory integration within perceptual decision formation as purely sensory, or post-sensory in respect to modulating decision dynamics (i.e., perceptual inference), are prevalent in limiting our understanding of how learning enhances the formation of rapid multisensory decisions (Noppeney, 2021).

In this study, we sought to explore how perceptual learning modulates the underlying cognitive and neural mechanisms of multisensory perceptual decision formation. In particular, we sought to see if any of the unisensory perceptual learning hypotheses could account for the development of multisensory benefits towards decision formation during audiovisual object categorisation. In addition, we further sought to pinpoint whether perceptual learning would heighten any observed multisensory integrative benefits towards perceptual decision formation (i.e., decreased RTs and/or increased choice accuracy) over the course of three consecutive days. This was achieved through comparison of behavioural performance differences in RTs and choice accuracy (i.e., binary responses) between multisensory (i.e., audiovisual) and unisensory (i.e., visual or auditory) trial types over three consecutive days. We used another modified variant of Franzen et al. (2020)'s audiovisual face-versus-car categorisation paradigm, coupled with EEG, to further discern the modulatory effects of learning on the underlying neural recordings. To characterise the neural dynamics underlying any observable behavioural benefits, we analysed single-trial EEG activity using multivariate LDA (Parra et al., 2002; 2005; Sajda et al., 2009; Philiastides & Sajda, 2006a; 2006b; Philiastides, Ratcliff, & Sajda, 2006; Philiastides et al., 2014), to identify patterns of activity discriminating (1) face-versus-car stimulus categorisations (i.e., *within*), and (2) *between* multisensory and unisensory trial types respectively.

4.3 Materials and Methodology

4.3.1 Participants

22 participants (male = 14, female = 8; mean age = 23.70; standard deviation = 2.79; age range = 18-29) were opportunistically recruited from the University of Leeds, and received £45 (UK sterling) for successful completion of the full experiment (i.e., over three consecutive days; see section 4.3.3 *Experimental Paradigm and Procedure*). All participants were right-handed, reported normal hearing, normal or corrected-to-normal vision, and had no history of neurological conditions. The study was approved by the ethics committee of the Faculty of Biological Sciences at the University of Leeds (BIOSCI 19-021) and conducted in accordance with the Declaration of Helsinki (World Medical Association, 2013).

4.3.2 Stimuli

We used the same stimuli, outlined in Chapter 3 (see section 3.3.2 *Stimuli*), consisting of 36 grayscale images (18 images of faces and cars respectively) and 36 sounds (18 sounds of human speech and car/street-related sounds respectively) adapted from previous experiments (Diaz et al., 2017; Franzen et al., 2020; Philiastides & Sajda, 2006a; 2006b; 2007; Philiastides, Ratcliff, & Sajda, 2006). However, all images were instead displayed on a purple background (RGB [102 0 128]) instead of a white background (RGB [255 255 255]), and feedback was replaced by two images of a cartoon scientist illustrating two possible outcomes: (a) smiling and holding his thumb up for correct responses, and (b) frowning and holding his thumb down for incorrect responses (see sections 3.3.2 *Stimul*i and 3.3.3 *Experimental Paradigm and Procedure*). The visual stimulus display was controlled by a Stone 64-bit based machine (CPU: i7-9700; RAM: 500GB SSD) running Windows Professional 7 (Linux-x86_64-bit) and PsychoPy presentation software (version 1.92.01; Pierce et al., 2019). All images were presented on an Iiyama ProLite B2484HSU 24-inch monitor (resolution: 1920 x 1080 pixels; refresh rate: 75 Hz). Participants were seated 60cm from the visual stimulus display at all times. Auditory stimuli were presented using Sennheiser HD 280 headphones, with the sound intensity of all tones matched to 72 DbA SPL (across left and right ears).

4.3.3 Experimental Paradigm and Procedure

We employed a similar modified variant of the audiovisual face-versus-car categorisation paradigm adapted from Franzen et al. (2020). As was previously outlined in Chapter 3 (see subsection 3.2.*3 Experimental Paradigm and Procedure*), this is a 2AFC paradigm that requires participants to categorise, on a single-trial basis, whether a face or a car is embedded in a presented stimulus. The presented stimuli consisted of the following: (a) images of faces and cars (visual stimuli; V), (b) sounds of human speech or car/street-related sounds (auditory stimuli; A), or (c) simultaneously presented and matching (i.e., congruent) images and sounds of faces/human speech and cars/car or street-related sounds (audiovisual stimuli; AV). All stimuli were presented for a period of 300ms and in a pseudorandomised sequence. Two levels of stimulus phase coherence were used to vary the amount of sensory evidence in presented images and sounds (i.e., “noise”; High Coherence; HC and Low Coherence; LC, see sections 3.3.2 *Stimuli*, 3.4.3 *Experimental Paradigm*, and 4.3.2 *Stimuli* respectively), therefore manipulating the difficulty of categorising embedded stimuli. Participants were instructed to indicate their categorised decision via a standard keyboard button press as quickly and accurately as possible, with RTs and binary responses (i.e., correct or incorrect) collected as dependent variable measurements quantifying behavioural performance (and therefore perceptual decision formation). Each participant performed this paradigm on three consecutive days, with the experiment taking place at the same time on each day. Successful completion of the experiment involved a commitment to completing all trials on each day. On the first day, participants completed a maximum of 15 trials in order to familiarize themselves with the structure and pace of the paradigm.

Figure 4.1 illustrates the procedure on a single-trial basis. Prior to completing the experimental paradigm, participants were presented with an ethical consent form, study information, and instructions for preparation prior to completing the paradigm. The instructions explicitly specified the instructions of the paradigm itself, whereby they would be shown a series of quick and distorted stimuli consisting of images only (V), sounds only (A), and images and sounds, presented together, of faces/human speech or cars/street-related noises (AV), and asked to categorise the presented stimulus as either a face or a car, using the left and right arrow standard keyboard buttons, as quickly and as accurately as possible, with further instructions to position their left index and middle fingers over the *left* and *right* arrow keys respectively.

Notably, they were informed that audiovisual stimuli would always be matching (i.e., congruent) and that images and sounds would neither be mismatching (i.e., face images-traffic-related sounds, car images-human speech sounds), and instructed to refrain from categorising images and sounds individually in these trials. They were further informed that if they were unsure in their decision to guess to the best of their capabilities since they had a maximum time limit to indicate a response (1250ms). Finally, participants were instructed that they would receive visual feedback following each response.

Participants completed the experiment in a darkened and sound-attenuated room on each day. Each trial started with a fixation cross presented centrally on-screen for 1000ms. Then, one of three stimuli (i.e., images only, sounds only, simultaneously presented images and sounds) were presented for 300ms. Auditory stimuli were accompanied by an image of a speakerphone on screen to immediately indicate to the participant that the presented stimulus is a sound only. Participants were then instructed to categorise, as quickly and as accurately as possible, the presented stimuli using the left and right arrow response keys for face and car stimuli respectively. The response deadline was set at 1250ms. Feedback was presented centrally for 500ms for two possible outcomes: (a) a cartoon scientist smiling and holding his thumb up for correct responses, or (b) a cartoon scientist frowning and holding his thumb down for incorrect responses (see section 4.3.2 *Stimuli*). On each day, we presented 216 trials (divided equally between the two response categories, i.e. 108 face trials and 108 car trials, and the two levels of stimulus coherence, i.e. 108 HC trials and 108 LC trials), presented in three blocks of 72 trials each (divided equally between the stimuli, i.e. 24 V, 24 A, and 24 AV trials, totalling 72 trials for each stimuli respectively), with a 60 second rest period between blocks. In total, we presented 648 trials over the three days. The entire experiment lasted approximately 20-25 minutes per day.

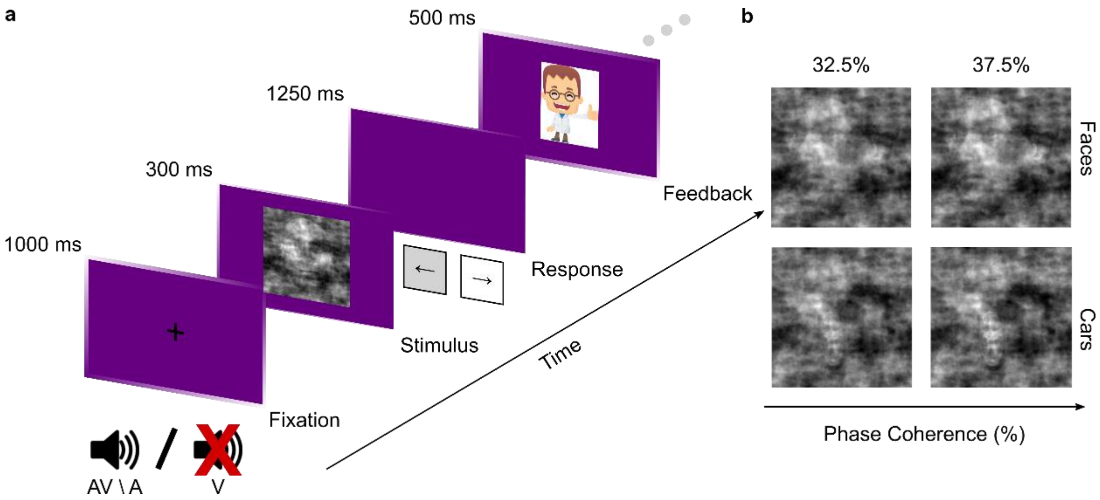


Figure 4.1 Experimental Paradigm. a, Schematic representation of the experimental paradigm on a single-trial basis. Participants were instructed to categorise noisy representations of faces and cars. First, a fixation cross was presented centrally on-screen for 1000ms. Then, a brief stimulus, which was either an image (V), a sound (A), or a simultaneously presented image and sound (AV), was presented for 300ms, and followed by a delay period of a maximum of 1250ms during which participants were instructed to indicate their response with a keyboard button press (*left*/*right* keyboard button presses for *face*/*car* categorisations respectively). Following their response, feedback was then presented for 500ms (a scientist holding a thumbs up or a thumbs down, for a correct or incorrect response, respectively), which preceded an inter-stimulus interval (ISI) of 500ms (illustrated by the ellipsis). **b**, Sample face (*top*) and car (*bottom*) images at the two levels of stimulus phase coherence used in the experimental paradigm (Low Coherence: 32.5%, High Coherence: 37.5%).

4.3.4 Statistical Analysis of Behavioural Data

For each participant, RTs (calculated in milliseconds) and choice accuracy (calculated as a binary variable of correct and incorrect responses) were collected as single-trial dependent variable measurements quantifying behavioural performance (and hence perceptual decision formation) for three categorical independent variables: (1) sensory trial type (three levels: V | A | AV), (2) stimulus coherence (two levels: HC | LC), and (3) testing day (three levels: Day One | Day Two | Day Three). Trials in which participants did not facilitate a choice response within the 1250ms delay period were excluded from further analyses. Overall, 410 trials were excluded from an initial 14256 trials, leaving a remaining total of 13846 trials. Next, an Anderson-Darling test was used to assess if RTs were normally distributed (Anderson & Darling, 1954; Nelson, 1998; Stephens, 1974). As we have previously outlined (see sections 2.3.5 and 3.3.4 *Statistical Analysis of Behavioural Data* respectively), this is because it is commonly observed in research that RT distributions tend to be not normally distributed, with histograms prominently illustrating a positive right skew (Marmolejo-Ramos et al., 2015; Whelan, 2008). As expected, the assumption of normality was found to be violated for RTs over the full pre-processed behavioural dataset ($A^2 = 325.45$, $p < 0.001$).

Therefore, we quantified participants’ behavioural performance using GLMMs using the *lme4* package in RStudio (Bates et al., 2014; R Core Team, 2022), specifying *gamma* and *binomial logit* models for log-transformed RTs and choice accuracy respectively[4]. Our GLMM analyses included main effects and two-way interactions of the three predictors: sensory modality (three levels: Visual, V; Auditory, A; Audiovisual, AV), stimulus coherence (two levels: High Coherence, HC; Low Coherence, LC), and testing day (three levels: Day One; Day Two; Day Three), along with by-participant random slopes and random intercepts. Random correlations were excluded for all GLMMs. All GLMMs were fit using a *bobyqa* optimizer to ensure model convergence. *Post Hoc* likelihood-ratio (χ^2) model comparisons were used to quantify the predictive power and significance of all main effects and interactions in our GLMM analyses. These *Post Hoc* likelihood-ratio (χ^2) model comparisons compared full models (i.e., models including main effects, their two-way interactions, and random effects) to reduced models that excluded the main predictor or two-way interaction in question.

4.3.5 EEG Recording and Preprocessing

Continuous EEG data was recorded simultaneously with experiment participation in a darkened and sound-attenuated room using a 64-channel EEG amplifier system (actiCHamp Plus, Brain Products, GmbH, Germany). Ag/AgCl (sinter) scalp electrodes were placed in accordance with the international 10-20 positioning system on an actiCAP cap (Brain Products GmbH, Germany). A separate ground electrode was used, with all channels initially referenced to electrode FCz (placed on the midline sagittal plane of the scalp) during EEG recording. The input impedance of all channels was adjusted to <15 kΩ through application of a high viscosity electrolyte gel (SuperVisc, Brain Products GmbH, Germany) between each electrode and the scalp to minimize the resistance to electrode current flow. Brain Vision

---

[4] See section 3.3.4 *Statistical Analysis of Behavioural Data* for an outline of the rationale of using GLMMs for statistical analysis of behavioural data.

Recorder (version 1.22.01, Brain Products GmbH, Germany) was used to record EEG data and trial-specific information, including event triggers for displayed stimuli, response key presses, and feedback presentations. All EEG recordings were sampled at a rate of 1000 Hz and underwent online filtering using a 0.0016-250Hz analog band-pass filter. Upon completion of data collection, recordings were stored for offline preprocessing and analysis in MATLAB (Mathworks).

Datasets from individual sessions (i.e., testing days) for each participant were preprocessed offline separately in MATLAB using *EEGLAB* (Delorme & Makeig, 2004). All datasets were re-referenced to obtain a common average reference, which computes the signal average at all EEG channel locations, and subtracts it from the EEG signal at each electrode location at the sampled time point. Each dataset was then filtered using a Butterworth band-pass filter between 0.5-40Hz cut-off frequencies in order to minimise linear trends from phase-related frequency distortions. This minimises the introduction of artefact filtering at shorter time intervals (i.e., trials). Independent Component Analysis (ICA) was used to identify potential noise artefacts, both internal (e.g., eye blinks and saccades) and external (e.g. line and/or channel noise) for subsequent removal (Delorme et al., 2004; 2007; 2012; Makeig et al., 1995; 2004; Onton, Delorme, & Makeig, 2005). This was applied twice to ensure sufficient convergence during feature decomposition. The built-in function *ICLabel* (Pion-Tonachini, Kreutz-Delgado, & Makeig, 2019) was used to distinguish independent components that were characteristic of sources originating from the brain from those that were characteristic of non-task related eye and bodily movements, heartbeats, and/or noisy electrode channels. Finally, epochs from -100ms to 1000ms relative to stimulus-onset were extracted on a single-trial basis. For two participants (SO2 and SO5), single noisy channels (TP9 and FP2 respectively) were interpolated to minimise noise artefacts from unexpected non-neural resources, and to reconstruct EEG signals from neighbouring electrodes.

4.3.6 EEG Signal Analysis – Linear Discriminant Analysis

We applied single-trial multivariate LDA (Parra et al., 2002; 2005; Philiastides & Sajda, 2006a; 2006b; Philiastides, Ratcliff, & Sajda, 2006; Philiastides et al., 2014; Sajda et al., 2009) to extract stimulus-locked EEG components that discriminated (1) face-versus-car trial types *within* sensory trial types (Visual;

V, Auditory; A, and Audiovisual; AV), and (2) AV-versus-V and AV-versus-A trial types *between* multisensory and unisensory trial types, across testing days (Day 1, Day 2, and Day 3). As we have previously outlined (see section 2.3.7 *EEG Signal Analysis – Linear Discriminant Analysis*), 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, and reduces effects common between two conditions of interest. Compared to univariate trial-averaging approaches, notably 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 SNR (Sajda et al., 2011) for preserved task-relevant information.

We used a sliding window approach (Parra et al., 2005; Sajda et al., 2009) to identify projections of the multichannel EEG signal, $x_i(t)$, where $i = [1 \dots N\ trials]$ and N is the total number of trials, within 100ms time windows that maximally discriminated (1) face-versus-car trial types within sensory trial types (Visual; V, Auditory; A, and Audiovisual; AV) and (2) AV-versus-V and AV-versus-A trial types between sensory trial types, across testing day (Day 1, Day 2, and Day 3). All time windows had a width of 100ms, with the window centre t shifted from -100ms to 1000ms, relative to the onset of stimulus presentations (i.e., 0ms), in 10ms increments. Specifically, we used logistic regression (Parra et al., 2002; 2005) to learn a 64-channel spatial weighting vector $w(t)$ that achieved maximal discrimination within each time window. This yielded 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 (1) categorizing face-versus-car trial types within sensory trial types, and (2) categorizing sensory trial types, thus separating activity maximizing differences and minimizing similarities of neural processes common to two conditions of interest. In discriminating individual conditions of interest, the classifier maps negative and positive discriminant component amplitudes to (1) face-versus-car trial types respectively within each sensory trial type, and (2) V-versus-AV and A-versus-AV trial types between sensory trial types respectively. Thus, larger negative values indicate a higher likelihood of categorizing (1) face trial types, and (2) V (versus AV) and A (versus AV) trial types respectively, whereas larger positive values indicate a higher likelihood of categorizing (1) car trial types, and (2) AV (versus V) and AV (versus A trials respectively, with values near zero reflecting less discriminative component amplitudes.

We quantified classification performance for each time window using the area under a receiver operating characteristic (ROC) curve (Green & Swets, 1966), referred to as an A_z value, using a leave-one-out cross-validation procedure (Gherman & Philiastides, 2015; Philiastides & Sajda, 2006a; 2006b). To determine group significance thresholds for discriminator performance, we implemented a permutation test, whereby (1) face-versus-car and (2) V-versus-AV and A-versus-AV trial labels were randomized and submitted to the leave-one-out procedure. This randomization procedure was repeated 100 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 for V-versus-AV and A-versus-AV trial types, resulting from equation (1), by estimating a forward model as:

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

In which 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 10ms 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).

4.4 Results

4.4.1 Behavioural Results

Generalised Linear Mixed-Effects Models (GLMMs) and *post hoc* likelihood-ratio (χ^2) model comparisons were used to analyse log-transformed RTs and choice accuracy (using *gamma* and *binomial logit* models for log-transformed RTs and choice accuracy respectively) as a function of sensory trial type (Visual; V, Auditory; A, and Audiovisual; AV), stimulus coherence (High Coherence; HC, and Low Coherence; LC) trial types, and testing day (Day One, Day Two, and Day Three), as well as their subsequent two-way interactions and three way interactions.

4.4.1.1 Reaction Time GLMM Analyses Results

Our log-transformed RT GLMM analyses demonstrated significant predictive power of the main effects of sensory trial type (Median_{V}: 5.99 log-RT post-stimulus offset, Median_{A} = 6.15 log-RT post-stimulus offset, Median_{AV}: 5.95 log-RT post-stimulus offset; χ^2 = 1108.62, df = 2, p < 0.001), stimulus coherence trial type (Median_{HC} = 5.99 log-RT post-stimulus offset, Median_{LC} = 6.07 log-RT post-stimulus offset; χ^2 = 107.74, df = 1, p < 0.001), and testing day ($\text{Median}_{\text{Day One}}$ = 6.15 log-RT post-stimulus offset, $\text{Median}_{\text{Day Two}}$ = 6.03 post-stimulus offset, $\text{Median}_{\text{Day Three}}$ = 5.95 log-RT post-stimulus offset; χ^2 = 651.41, df = 2, p < 0.001).

Furthermore, our RT GLMMs demonstrated significant predictive power of the main two-way interactions between sensory trial type and stimulus coherence ($\text{Median}_{V \mid HC}$ = 5.95 log-RT post-stimulus offset, $\text{Median}_{A \mid HC}$ = 6.15 log-RT post-stimulus offset, $\text{Median}_{AV \mid HC}$ = 5.91 log-RT post-stimulus offset, $\text{Median}_{V \mid LC}$ = 6.03 log-RT post-stimulus offset, $\text{Median}_{A \mid LC}$ = 6.18 log-RT post-stimulus offset, Median$_{\text{AV | LC}}$ = 5.99 log-RT post-stimulus offset; χ^2 = 107.40, df = 2, p < 0.001), but not of the main interactions between sensory trial type and testing day (Median$_{\text{V | Day One}}$ = 6.11 log-RT post-stimulus offset, Median$_{\text{V | Day Two}}$ = 5.99 log-RT post-stimulus offset, Median$_{\text{V | Day Three}}$ = 5.91 log-RT post-stimulus offset, Median$_{\text{A | Day One}}$ = 6.25 log-RT post-stimulus offset, Median$_{\text{A | Day Two}}$ = 6.15 log-RT post-stimulus offset, Median$_{\text{A | Day Three}}$ = 6.07 log-RT post-stimulus offset, Median$_{\text{AV | Day One}}$ = 6.03 log-RT post-stimulus offset, Median$_{\text{AV | Day Two}}$ = 5.95 log-RT post-stimulus offset, Median$_{\text{AV | Day Three}}$ = 5.86 log-RT post-stimulus offset; χ^2 = 0.27, df = 2, p = 0.875) or between stimulus coherence trial type and testing day (Median$_{\text{HC | Day One}}$ = 6.11 log-RT post-stimulus offset, Median$_{\text{HC | Day Two}}$ = 5.99 log-RT post-stimulus offset, Median$_{\text{HC | Day Three}}$ = 5.91 log-RT post-stimulus offset, Median$_{\text{LC | Day One}}$ = 6.18 log-RT post-stimulus offset, Median$_{\text{LC | Day Two}}$ = 6.07 log-RT post-stimulus offset, Median$_{\text{LC | Day Three}}$ = 5.99 log-RT post-stimulus offset; χ^2 = 3.09, df = 2, p = 0.079).

Finally, our RT GLMMs demonstrated no significant predictive power of the main three-way interaction between sensory trial type, stimulus coherence trial type, and testing day conditions (Figure 4.2a; Median$_{\text{V | HC | Day One}}$ = 6.07 log-RT post-stimulus offset, Median$_{\text{V | HC | Day Two}}$ = 5.95 log-RT post-stimulus offset, Median$_{\text{V | HC | Day Three}}$ = 5.86 log-RT post-stimulus offset, Median$_{\text{V | LC | Day One}}$ = 6.15 log-RT post-stimulus offset, Median$_{\text{V | LC | Day Two}}$ = 6.07 log-RT post-stimulus offset, Median$_{\text{V | LC | Day Three}}$ = 5.95 log-RT post-stimulus offset, Median$_{\text{A | HC | Day One}}$ = 6.21 log-RT post-stimulus offset, Median$_{\text{A | HC | Day Two}}$ = 6.13 log-RT post-stimulus offset, Median$_{\text{A | HC | Day Three}}$ = 6.03 log-RT post-stimulus offset, Median$_{\text{A | LC | Day One}}$ = 6.28 log-RT post-stimulus offset, Median$_{\text{A | LC | Day Two}}$ = 6.18 log-RT post-stimulus offset, Median$_{\text{A | LC | Day Three}}$ = 6.09 log-RT post-stimulus offset, Median$_{\text{AV | HC | Day One}}$ = 5.99 log-RT post-stimulus offset, Median$_{\text{AV | HC | Day Two}}$ = 5.91 log-RT post-stimulus offset, Median$_{\text{AV | HC | Day Three}}$ = 5.81 log-RT post-stimulus offset, Median$_{\text{AV | LC | Day One}}$ = 6.07 log-RT post-stimulus offset, Median$_{\text{AV | LC | Day Two}}$ = 5.99 log-RT post-stimulus offset, Median$_{\text{AV | LC | Day Three}}$ = 5.86 log-RT post-stimulus offset, χ^2 = 1.42, df = 2, p = 0.491)

4.4.1.2 Choice Accuracy GLMM Analyses Results

Our choice accuracy GLMMs demonstrated significant predictive power of the main effects of sensory trial type (Proportion Correct$_{\text{V}}$ = 0.913, Proportion

Correct$_{\text{A}}$ = 0.908, Proportion Correct$_{\text{AV}}$ = 0.955; χ^2 = 140.64, df = 2, p < 0.001), stimulus coherence trial type (Proportion Correct$_{\text{HC}}$ = 0.949, Proportion Correct$_{\text{LC}}$ = 0.903; χ^2 = 90.22, df = 1, p < 0.001), and testing day (Proportion Correct$_{\text{Day One}}$ = 0.901, Proportion Correct$_{\text{Day Two}}$ = 0.935, Proportion Correct$_{\text{Day Three}}$ = 0.942; χ^2 = 90.70, df = 2, p < 0.001).

Furthermore, our choice accuracy GLMMs demonstrated significant predictive power of the main two-way interactions between sensory trial type and stimulus coherence trial type (Proportion Correct$_{\text{V | HC}}$ = 0.944, Proportion Correct$_{\text{A | HC}}$ = 0.929, Proportion Correct$_{\text{AV | HC}}$ = 0.973, Proportion Correct$_{\text{V | LC}}$ = 0.884, Proportion Correct$_{\text{A | LC}}$ = 0.888, Proportion Correct$_{\text{AV | LC}}$ = 0.938; χ^2 = 6.60, df = 2, p = 0.037), but not of the main interactions between sensory trial type and testing day (Proportion Correct$_{\text{V | Day One}}$ = 0.878, Proportion Correct$_{\text{V | Day Two}}$ = 0.924, Proportion Correct$_{\text{V | Day Three}}$ = 0.940, Proportion Correct$_{\text{A | Day One}}$ = 0.892, Proportion Correct$_{\text{A | Day Two}}$ = 0.912, Proportion Correct$_{\text{A | Day Three}}$ = 0.921, Proportion Correct$_{\text{AV | Day One}}$ = 0.933, Proportion Correct$_{\text{AV | Day Two}}$ = 0.968, Proportion Correct$_{\text{AV | Day Three}}$ = 0.965; χ^2 = 3.75, df = 2, p = 0.154), or between stimulus coherence trial type and testing day (Proportion Correct$_{\text{HC | Day One}}$ = 0.930, Proportion Correct$_{\text{HC | Day Two}}$ = 0.51, Proportion Correct$_{\text{HC | Day Three}}$ = 0.964, Proportion Correct$_{\text{LC | Day One}}$ = 0.872, Proportion Correct$_{\text{LC | Day Two}}$ = 0.918, Proportion Correct$_{\text{LC | Day Three}}$ = 0.920; χ^2 = 0.60, df = 2, p = 0.439).

Finally, our choice accuracy GLMMs demonstrated no significant predictive power of the main three-way interaction between sensory trial type, stimulus coherence trial type, and testing day conditions (Figure 4.2b; Proportion Correct$_{\text{V | HC | Day One}}$ = 0.920, Proportion Correct$_{\text{V | HC | Day Two}}$ = 0.946, Proportion Correct$_{\text{V | HC | Day Three}}$ = 0.964, Proportion Correct$_{\text{V | LC | Day One}}$ = 0.836, Proportion Correct$_{\text{V | LC | Day Two}}$ = 0.901, Proportion Correct$_{\text{V | LC | Day Three}}$ = 0.915, Proportion Correct$_{\text{A | HC | Day One}}$ = 0.916, Proportion Correct$_{\text{A | HC | Day Two}}$ = 0.927, Proportion Correct$_{\text{A | HC | Day Three}}$ = 0.944, Proportion Correct$_{\text{A | LC | Day One}}$ = 0.868, Proportion Correct$_{\text{A | LC | Day Two}}$ = 0.897, Proportion Correct$_{\text{A | LC | Day Three}}$ = 0.898, Proportion Correct$_{\text{AV | HC | Day One}}$ = 0.953, Proportion Correct$_{\text{AV | HC | Day Two}}$ = 0.980, Proportion Correct$_{\text{AV | HC | Day Three}}$ = 0.985, Median$_{\text{AV | LC | Day One}}$ = 0.912, Proportion Correct$_{\text{AV | LC | Day Two}}$ = 0.955, Median$_{\text{AV | LC | Day Three}}$ = 0.945, χ^2 = 0.42, df = 2, p = 0.812).

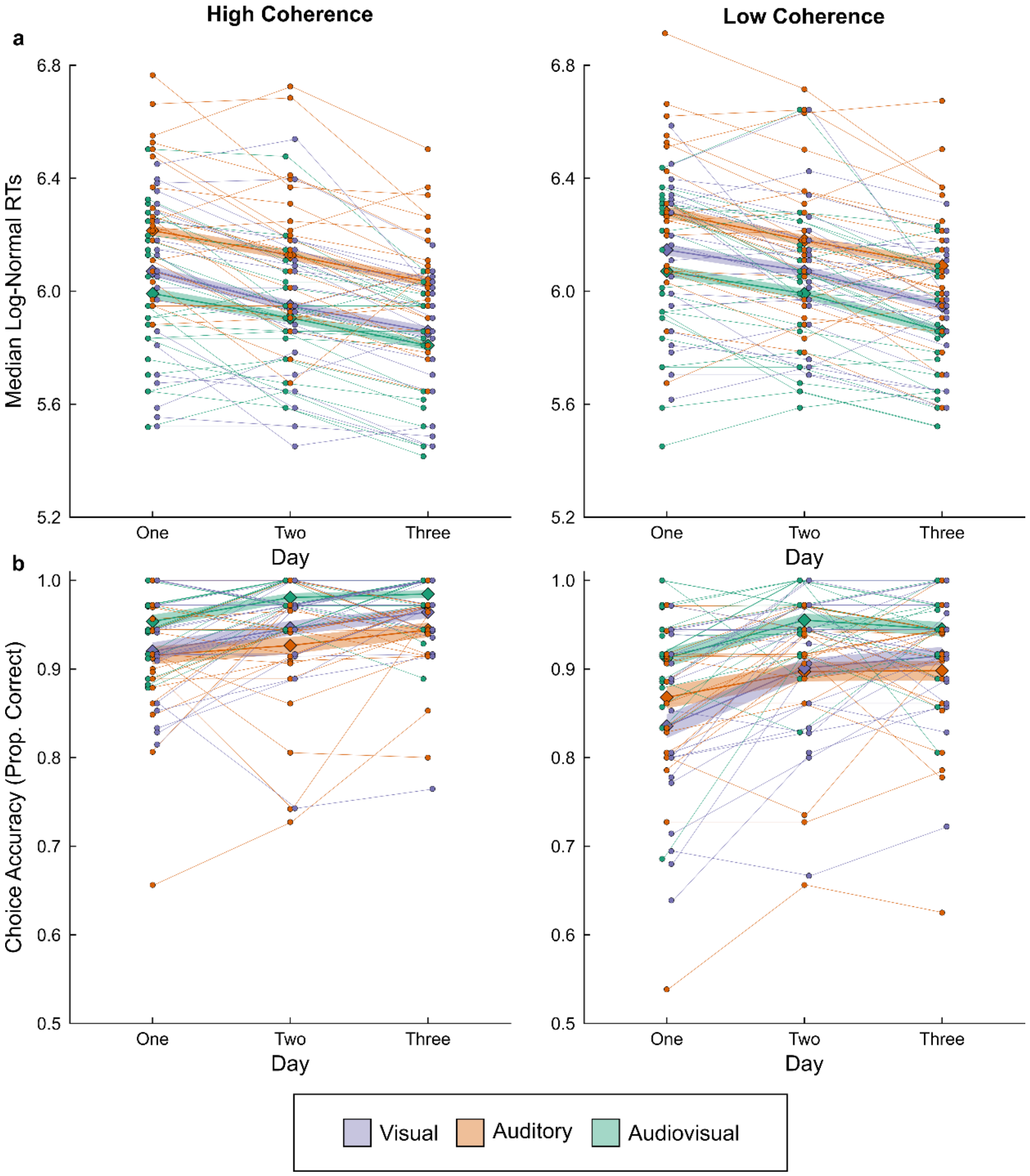


Figure 4.2 Behavioural Results. a, Median log-transformed RTs and **b,** choice accuracy (proportion of correct responses) for condition (diamonds) and participants (scatter points) across two levels of stimulus coherence (*Left*. High Coherence trials; *Right*. Low Coherence trials) as a function across three levels of sensory trial type (Visual: *purple*, Auditory: *orange*, Audiovisual: *green*) and testing days (Day One, Day Two, and Day Three). For all graphs, standard errors of the median and mean were computed for median log-transformed RTs and choice accuracy respectively, as represented by the shaded regions across groups.

4.4.2 EEG Signal Analysis Results

First, we analysed the EEG data to identify neural components that discriminated between face-versus-car trial types embedded within visual (V), auditory (A), and audiovisual (AV) sensory trial types across testing days. For each participant, we performed single-trial multivariate LDA to identify linear spatial weightings (i.e., spatial filters) of the EEG sensors that discriminated face-versus-car trial types. The identified weightings yielded a projection of the 64-dimensional EEG space that maximally discriminated stimulus-locked face-versus-car trial types within pre-defined time windows of 100ms. Second, we analysed the EEG data to identify neural components that discriminated between multisensory (i.e., AV) and unisensory (i.e., A or V) trial types across testing days. For each participant, we performed single-trial multivariate LDA to identify linear spatial weightings (i.e., spatial filters) of the EEG sensors that discriminated multisensory (i.e., AV) versus unisensory (i.e., A or V) trial types across testing days. The identified weightings produced a projection of the 64-dimensional EEG space that maximally discriminated stimulus-locked AV-versus-V, and AV-versus-A trial types, within pre-defined time windows of 100ms.

Application of the linear spatial filters to the single-trial EEG dataset produced measurements quantifying the resulting component amplitudes discriminating between the aforementioned conditions of interest (γ, see sections 2.3.7 *EEG Signal Analysis – Linear Discriminant Analysis* and 4.3.6 *EEG Signal Analysis – Linear Discriminant Analysis*). These components amplitudes can be used as an index of the quality of categorising: (1) face-versus-car trial types embedded within V, A, and AV sensory trial types, and (2) AV-versus-V and AV-versus-A trial types across testing days respectively. Namely, higher amplitudes, negative or positive, indicate higher neural evidence for two conditions of interest at a time, while amplitudes closer to zero indicate less neural evidence. For all multivariate analyses, we quantified the discriminator’s classification performance using area under a receiver operating characteristic curves (i.e., AUC-ROC; A_z value), coupled with leave-one-trial-out cross validation approaches, in order to control for overfitting. Compared to traditional approaches, which assume an A_z value of 0.5 as chance performance, we performed permutation analyses using leave-one-trial-out procedures that produced A_z randomization distributions, computing group-average A_z values, that lead to a conventional significance level of $p = 0.05$.

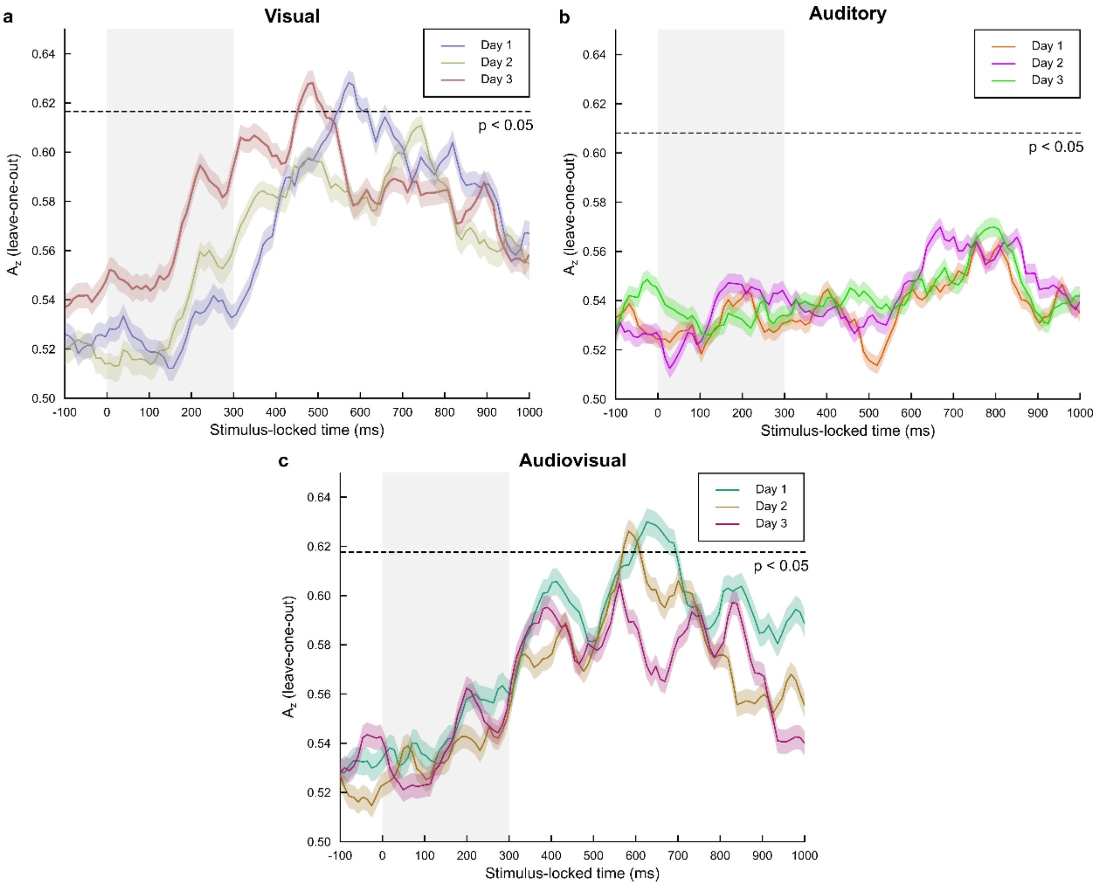


Figure 4.3 Face-Versus-Car Multivariate Linear Discriminant Analysis Results. Mean multivariate discriminator performance (A_Z), quantified by leave-one-out cross-validation procedure, during face-versus-car trial type outcome discrimination of stimulus-locked EEG responses, as a function of testing day (Day 1, Day 2, and Day 3) within **a**, Visual (V), **b**, Auditory (A), and **c**, Audiovisual (AV) sensory trial types. Dashed black lines represent the group average permutation thresholds at $p < 0.05$ for face-versus-car trial type discriminator performance for all trial types across testing days. Shaded error bars across testing days denote the standard error of the mean across participants. Shaded area denotes the presentation of stimuli, from 0ms (post-stimulus onset) to 300ms (post-stimulus offset).

The performance of our stimulus-locked discriminators for decoding face-versus-car trial types (within sensory trial types) across testing days did not appear to reveal broad time windows in which face-versus-car trial type decoding could be considered statistically reliable (-100-1000ms post-stimulus onset). As observed in Figure 4.3, this is due to the performance of all discriminators predominantly remaining below the significance levels estimated from permutation testing. We did observe exceptions within V and AV trial types, in which face-versus-car decoding did exceed A_Z values in which classification performance could be considered statistically reliable. These were localised at approximately 400-600ms post-stimulus onset within

V trial types (Figure 4.3a) and at approximately 600-700ms post-stimulus onset within AV trial types (Figure 4.3c). However, we did not observe any systematic differences in classification performance over testing day. Overall, these results indicate that multivariate LDA classification could not exhibit significant neural information for classification of face-versus-car trial types, nor could it exhibit pronounced components to correlate peak discriminant output to identify temporally-specific time points to infer the underlying neural dynamics of multisensory perceptual learning. Consequently, we chose not to examine the resultant forward models, since any inferences for where in the brain would also not be statistically reliable.

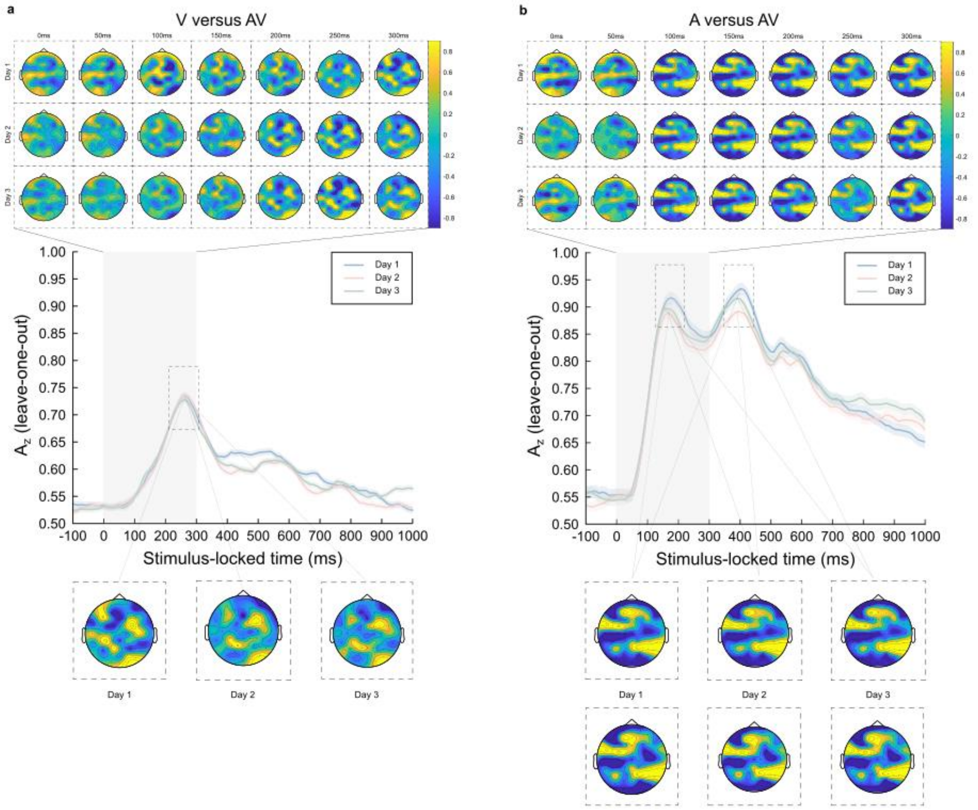


Figure 4.4 AV-versus-AV and AV-versus-A Trial Type Multivariate Linear Discriminant Analysis Results. Mean multivariate linear discriminator performance (A_Z), quantified by leave-one-out trial cross-validation performance, during outcome discrimination of stimulus-locked EEG responses, as a function of **a**, V versus AV trial types, and **b**, A versus AV trial types over testing days (Day 1; blue, Day 2; red, Day 3; green). Shaded error regions denote the standard error of the mean across participants. Shaded areas denote the stimulus presentation duration (300ms). *Top*. Scalp topographies at representative time windows corresponding to stimulus presentation duration in 50ms increments from 0ms (post-stimulus onset) to 300ms (post-stimulus offset) over for **a**, V versus AV and **b**, A versus AV conditions, over testing days. *Bottom*. Scalp topographies at representative

time windows corresponding to EEG components, defined by peaks of maximum discrimination for **a**, V versus AV trial types (at approximately 270ms), and **b**, A versus AV trial types (at approximately 170ms and 400ms respectively).

However, the performance of our stimulus-locked discriminators for decoding AV-versus-V and AV-versus-A sensory trial types did reveal broad time windows for which decoding of sensory trial types could be considered statistically reliable. For AV-versus-V decoding, we observed increased discriminator performance from 0-300ms post-stimulus onset. Specifically, discriminator performance within this range was characterised by three components, all of which were localised at approximately 270ms post-stimulus onset (Figure 4.4a; Day 1: $A_z = 0.736$; Day 2: $A_z = 0.735$; Day 3: $A_z = 0.727$). The corresponding scalp topographies, obtained from using the forward model, correlated peak discriminant output and EEG data (averaged over a 100ms time window centered on the classification performance peak) consistently over right-lateralized occipital, centro-parietal, and left-lateralized fronto-central electrodes. For AV-versus-A decoding, we observed increased discriminator performance from approximately 0-400ms post-stimulus onset, characterized by two temporally-specific components arising at 170ms (Figure 4.4b; Day 1: $A_z = 0.917$; Day 2: $A_z = 0.890$; Day 3: $A_z = 0.898$) and 400ms (Figure 4.4b; Day 1: $A_z = 0.934$; Day 2: $A_z = 0.892$; Day 3: $A_z = 0.916$) post-stimulus onset. The corresponding scalp topographies, obtained from using the forward model, correlated peak discriminant output and EEG data (averaged over a 100ms time window centered on the classification performance peaks) consistently over right-lateralized parieto-occipital and left and central-lateralized fronto-central electrodes. Across participants, however, ANOVA analyses did not find any significant effects of testing day between AV-versus-A trial type peak components (170ms post-stimulus onset: $F(2, 63) = 0.460$, $p = 0.636$; 400ms post-stimulus onset: $F(2, 63) = 1.530$, $p = 0.224$) nor AV-versus-V trial type peak components (270ms post-stimulus onset: $F(2, 63) = 0.040$, $p = 0.957$).

4.5 Discussion

In our investigation of the effects of learning on multisensory perceptual decision-making, we did not find any significant evidence to demonstrate enhancements to decisional behaviour, nor modulations to the underlying neural mechanisms, between multisensory and unisensory trial types. Whereas our behavioural findings implied a general benefit of perceptual learning on behavioural measurements of decision-making performance, in which we observed significantly lower RTs and higher choice accuracy across testing days, we did not observe any significant interactions between sensory trial type and testing day. In addition, no significant neural components were found for classifying face-versus-car trial types either within sensory trial types (i.e., visual, auditory, and audiovisual) across testing days. Since temporally-specific *Early* and *Late* components were replicated across previous studies utilising similar variants of the face-versus-car object categorisation paradigm (Diaz et al., 2017; Franzen et al., 2020; Philiastides & Sajda, 2006a; 2006b; Philiastides, Ratcliff, & Sajda, 2006; 2009), we contend that our initial multivariate analysis lacked sufficient statistical power to not only capture similar developmental trajectories of discriminated EEG activity, but further capture potential effects of testing day as an indicator of the benefits of multisensory perceptual learning. Given previous research has discerned decision modulators, notably stimulus strength (Palmer, Huk, & Shadlen, 2005) and pre-stimulus anticipation (Summerfield & De Lange, 2014) exert an influence on choice benefits (i.e., decreased speed and increased accuracy of facilitated responses), we contend the overall scope of improvement we initially aimed to elicit during our investigations of multisensory perceptual decision-making was confounded by further decisional modulators, as well as paradigm and stimulus dependencies, not considered. Thus, we outline recommendations for future replications of our study below that evaluate potential decision (neuro)modulators that may have confounding influences on our initial observations.

One potential modulator to consider concerns the duration of stimulus presentations. We fixed the duration to 300ms across our implementations of the audiovisual face-versus-car categorisation paradigm (see sections 3.3.2 *Stimuli*, 3.3.3 *Experimental Paradigm and Procedure*, 4.3.2 *Stimuli*, and 4.3.3 *Experimental Paradigm and Procedure* in Chapters 3 and 4 respectively). In Chapter 3, our findings showed significant behavioural improvements in multisensory trial types (i.e., AV trial types) compared to unisensory trial types (i.e., V and/or A trial types; see section 3.4.1 *Behavioural Results*). In particular, we found faster RTs and higher proportions of correct responses, implying multisensory integrative benefits towards perceptual decision formation. Interestingly, this finding is contradictory to Franzen et al.’s findings

(Franzen et al., 2020), who did not find any significant effects of sensory trial type (i.e., AV versus V trial types) on RTs, nor a significant interaction with stimulus coherence trial type. Instead, they observed that RTs were, on average, slower for AV versus V trial types. Instead, they identified a multisensory integrative benefit for the accuracy of choice responses. Taking into consideration that the authors implemented a duration of 50ms for presenting sensory trial types, it is possible that the lack of replication in our RT analyses may be attributed to the timeframe for the encoding of stimulus presentations alters multisensory perceptual decision-making, and influences whether or not multisensory integrative benefits for decisional speed and/or accuracy may be discerned. Previous literature has outlined that paradigm and stimulus dependencies can govern the likelihood of multisensory temporal processes benefitting perceptual decision formation (Otto, Dassy, & Mamassion, 2013; Stevenson & Wallace, 2013). In consideration of the principle of temporal concordance outlined in Chapter 1 (see sections 1.2.1 *Multisensory Integration* and 1.2.2 *The Benefits of Multisensory Integration for Perceptual Decision-Making* respectively), it may be important to consider that the variability in behavioural results observed across our studies should be considered in accordance to research outlining that the length of the Temporal Binding Window may influence the recorded behavioural indices of multisensory perceptual decision-making and multisensory perceptual learning.

In particular, it is possible that our 300ms duration for stimulus presentations was not effable in our aims to elucidate the benefits of complementary multisensory information on perceptual decision formation over testing days. Instead, it appeared too long a duration in order to adequately operationalise conditions for multisensory perceptual learning, as well as for improvement in decision-making, as highlighted by no significant interaction between sensory trial type and testing day (see section 4.4.1 *Behavioural Results*). Whereas we observed behavioural performance was typically improved in multisensory versus unisensory trial types within each day, we did not observe an increase in multisensory perceptual decision-making performance that exceeded unisensory perceptual decision-making performance, and further identified participants who displayed maximum choice accuracy over the first testing day, therefore not leaving a scope for improvement of decisional accuracy over the second and third testing days (see Figure 4.2). Our findings contradict those reported by Diaz et al. (2017), who found significant effects of testing day on both RTs and binary responses for visual perceptual learning. Whereas our 300ms stimulus presentation duration enabled us to discern age-related impacts on the behavioural indices of multisensory perceptual decision-making (see Chapter 3), and thus overcome the hypothesis that our results would be confounded by possible wider temporal windows of integration in older adults, it did not enable us to discern learning-related impacts. Consequently, future replications of this study using the audiovisual face-versus-car categorisation paradigm should consider an adequate stimulus presentation duration that can replicate previous findings, thereby permitting for an increased possibility of perceptual learning improving multisensory decision-making performance.

A further limitation concerns the lack of sufficient statistical power for our multivariate LDA analyses (see section 4.4.2 *EEG Signal Analysis Results*). As previously outlined, we did not find any significant neural evidence for classifying face-versus-car trial types; either within sensory modality conditions (i.e., visual, auditory, and audiovisual), nor across testing days. Our findings are not compatible with the findings of EEG studies who applied similar variants of face-versus-car categorisation paradigm and multivariate LDA analyses (Diaz et al., 2017; Franzen et al., 2020; Philiastides & Sajda, 2006a; 2006b; Philiastides, Ratcliff, & Sajda, 2006; 2009). Undoubtedly, these studies demonstrated major discriminating EEG components characterising *Early* and *Late* temporal shifts in neural components underlying facilitated choices, and further characterised a smooth developmental trajectory from which they could infer the underlying constituent processes which modulate perceptual decision formation. In contrast, our LDA analyses for visual trial types in particular did not even remotely capture similar patterns of discriminated neural activity, nor salient components to utilise in a cognitively-informed modelling approach for the effects of multisensory perceptual learning. One possible contention is the number of trials across sensory trial types and testing days we implemented was insufficient for our supervised classifiers to discern EEG activity patterns for discriminating face-versus-car trial types. Reviewing the methodology of previous studies, and identifying trial quantities, reaffirms this notion. For example, studies from Philiastides et al., (2006a; 2006b) and Diaz et al. (2017) presented 288 trials, split over the two trial types (i.e., 144 face/car trial types respectively), with Diaz et al. presenting the same amount per testing day, whereas Franzen et al. (2020) presented 720 trials, split over all stimulus trial types (i.e., 360 face/car trial types; 360

visual/audiovisual trial types, 180 27.5%/30%/32.5% and 35% phase coherence conditions respectively). Given we presented 216 trials, split over the three sensory trial types (i.e., 36 face and car trial types for visual/auditory/audiovisual trial types respectively), it is possible that we did not have participants complete a sufficient quantity of trials for which to quantify underlying EEG signals. Thus, we recommend that future researchers ensure enough trials for which supervised dimensionality reduction analyses; such as single-trial multivariate LDA, can discern robust patterns of EEG signal classification. The aforementioned recommendation is supported by our alternate findings identifying neural components discriminating between sensory trial types, in which we achieved robust, stable, temporal trajectories of our identified neural components (see Figure 4.4).

Chapter 5
General Summary

5.1 Overview

Over the past two decades, substantial advances in our understanding of the process of perceptual decision-making across multiple domains of cognitive functioning have been developed. A collaborative set of experimental scenarios (in particular, 2AFC paradigms), neuroimaging methodologies, and computational models have yielded cognitive modelling approaches (either behaviourally informed or neurally-informed) for discerning the underlying neural and cognitive mechanisms for rapidly translating sensory information into overt behavioural responses. The application of cognitively-informed approaches for characterising similar mechanisms for perceptual decision-making with multisensory information remains a somewhat emerging field of research and many fundamental questions remain to address how we observe multisensory integrative benefits towards perceptual decision formation. To contribute to the existing literature, we sought to observe the exerted influence of three decision modulators (i.e., unisensory-formulated cross-modal associations, ageing, perceptual learning) on multisensory perceptual decision-making utilising our own cognitive modelling approaches.

In Chapter 2, 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.

In Chapter 3, using a behaviourally-informed cognitive modelling approach, we highlighted novel findings into age-related impacts on cognitive processes underlying multisensory perceptual decision-making. We disentangled HDDM findings predicting that, despite general decision-making deficits (i.e., increased RTs and reduced choice accuracy), ageing is associated with an increased functional reliance on multimodal information integration, whereby reductions in sensory representations, through decreased salience or increased task difficulty, are compensated for by the principle of inverse effectiveness compensating for general unisensory processing deficits. Overall, our findings demonstrate how computational modelling can reconcile contrasting hypotheses of age-related changes in processes underlying multisensory decision-making behaviour and develop upon computational modelling research of unisensory decision-making behavioural changes.

Chapter 4 was not as successful in being able to identify a temporal locus for the neural benefits of multisensory perceptual learning. We attribute this to (i) a lack of sufficient statistical power due to the number of trials across sensory conditions and testing days, and (ii) our stimulus presentation duration of 300ms, which confounded task difficulty as participants had too long to categorise embedded stimuli. Nonetheless, an analysis of EEG activity between multisensory (i.e., audiovisual) versus unisensory (i.e., visual or auditory) trial types uncovered neural components demonstrating greater discrimination for categorisations of stimuli. In particular, we observe prominent earlier and later components for audiovisual versus auditory stimuli, but only prominent early components for audiovisual versus visual stimuli. Our findings demonstrate preliminary evidence of different timescales for multisensory integration benefiting perceptual decision-making, depending on the sensory modality providing the complementary information.

5.2 Limitations of EEG

An immediate limitation of the findings from this current thesis is its reliance on utilising EEG to record neural activity underlying multisensory perceptual decision-making. One of the most predominant advantages of using EEG is its excellent temporal resolution. Changes in electrical activity can be recorded on the scale of milliseconds (ms; 1/1000th of a second). This benefits how we observe when changes in mesoscale brain dynamics unfold in real time for particular cognitive processes underlying multisensory perceptual decision-making. However, EEG has a predominant disadvantage of poor spatial resolution. The recorded signals are limited to generated electrical activity across the scalp, meaning that each electrode records activity not exclusively from the neurons below the electrode, but also from surrounding electrodes. This further includes specific structures in the deep laminae of cerebral cortex within the brain. A solution to rectify this is concerns simultaneously recording generative surface potentials over more electrodes. For example, the number of electrodes could be increased to 256. This would comparatively improve spatial resolution and the signal-to-noise ratio. Despite this, it is important to consider the electrical propagation, or movement direction in signal transmission, between neurons. Signal transmission between neurons is bi-directional, so signals can be propagated both towards and away from multiple neurons, depending on factors such as neuron type and location. Due to this, it remains difficult where to identify a particular process occurs in the brain. The limitations of EEG should therefore be considered concerning the neuroimaging findings presented in this thesis (see Chapters 2 and 4), since we cannot specify brain structures that determine the cognitive mechanisms facilitating multisensory perceptual decision-making. However, this does not fully limit the scope of our findings. Using our results, we were able to infer the spatiotemporal locus of multisensory integrative and decision formation processes, and identify where we may observe choice facilitative benefits to overt behaviours. Therefore, our findings can be compared and contrasted with findings that used methodologies such as fMRI and/or MEG, which achieve greater spatial resolution. Future research can even consider dual neuroimaging methodologies, in order to achieve both greater spatial and temporal resolution. This has been by Philiastides, Tu, & Sajda, 2021, who combined EEG and fMRI to investigate unisensory perceptual decision-making. Consequently, findings from an extensive variety of neurally-informed modelling approaches should be considered to further our understanding of the underlying neural and cognitive mechanisms of multisensory perceptual decision-making.

5.3 Conclusion

The current thesis presented research that characterised the underlying cognitive and neural mechanisms of multisensory perceptual decision-making in humans. Using both behaviourally-informed (Chapter 3) and neurally-informed (Chapters 2 and 4) modelling approaches, it details accounts of how decision (neuro)modulators can exert their influence on the various processes of multisensory perceptual decision-making. In particular, the use of the widespread accumulation-to-bound HDDM posited findings that demonstrated when, where, and/or how our nominated modulators: unisensory-formulated cross-modal associations, ageing, and perceptual learning, can influence the integration of multisensory integration to benefit perceptual decision formation. It enabled us to functionally characterise the latent processes underlying such multisensory enhancements. Chapter 2 showed clear enhancements of associative congruency when formulating cross-modal associations between auditory pitch-frequency and visual object-size from unisensory information alone. In dissecting two neurocomputational mechanisms that characterised different internal benefits (i.e., lower RTs), we demonstrate neurally-informed evidence in support of a dual-integration hypothesis for localising benefits towards perceptual decision formation. These mechanisms characterised benefits even in the absence of multisensory information, implicating that cross-modal binding, and not multisensory integration, may be a unique mechanism to optimise adaptive behaviours (Bizley, Maddox, & Lee, 2016b). Chapter 3 highlighted a behaviourally-informed modelling account of age-related changes in the use of multisensory information for perceptual decision-making. We particularly highlight prominent evidence of a generalised age-dependent slowing of sensory evidence accumulation for both multisensory and unisensory signals, but further evidence for compensatory mechanisms linked to the principle of inverse effectiveness and increased baseline filtering in unisensory modalities for older adults. This highlights further opportunities to exploit neurally-informed modelling approaches to obtain the corresponding neural underpinnings. Finally, whilst we did not observe any prominent multisensory perceptual learning benefits in Chapter 4, we observed neural evidence of multisensory benefits that were temporally-dependent on the quality of signals provided by the complementary sensory modality.

Bibliography

Aarts, E., Verhage, M., Veenvliet, J. V., Dolan, C. V., & Van Der Sluis, S. (2014). A solution to dependency: using multilevel analysis to accommodate nested data. *Nature Neuroscience*, *17*(4), 491-496.

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 Computational Biology*, *14*(7), e1006110.

Adam, R., & Noppeney, U. (2014). A phonologically congruent sound boosts a visual target into perceptual awareness. *Frontiers in Integrative Neuroscience*, *8*(70), 1-13.

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*(10), 457-464.

Alain, C., Chow, R., Lu, J., Rabi, R., Sharma, V. V., Shen, D., Anderson, N. D., Binns, M., Hasher, L., Yao, D., & Freedman, M. (2022). Aging enhances neural activity in auditory, visual, and somatosensory cortices: the common cause revisited. *Journal of Neuroscience*, *42*(2), 264-275.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257-262.

Alais, D., Newell, F., & Mamassian, P. (2010). Multisensory processing in review: from physiology to behaviour. *Seeing and Perceiving*, *23*(1), 3-38.

Albertazzi, L., Canal, L., Micciolo, R., & Hachen, I. (2020). Cross-modal perceptual organization in works of art. *i-Perception*, *11*(4), 2041669520950750.

Aller, M., Giani, A., Conrad, V., Watanabe, M., & Noppeney, U. (2015). A spatially collocated sound thrusts a flash into awareness. *Frontiers in Integrative Neuroscience*, *9*(16), 1-8.

Anderson, D., & Burnham, K. (2004). Model selection and multi-model inference (Second Edition). *NY: Springer-Verlag*, *63*(2020), 10.

Anderson, T. W., & Darling, D. A. (1954). A test of goodness of fit. *Journal of the American Statistical Association*, *49*(268), 765-769.

Angelaki, D. E., Gu, Y., & DeAngelis, G. C. (2009). Multisensory integration: psychophysics, neurophysiology, and computation. *Current Opinion in Neurobiology*, *19*(4), 452-458.

Anwyl-Irvin, A.L., Dalmaijer, E.S., Hodges, N., Evershed, J.K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavioural Research Methods*, *53*, 1407-1425.

Anwyl-Irvine, A.L., Massonié, J., Flitton, A., Kirkham, N.Z., & Evershed, J.K. (2020). Gorilla in our midst: an online behavioural experiment builder. *Behavioural Research Methods*, *52*, 388-407.

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. *Journal of Neuroscience*, *26*(47), 12260-12265.

Bang, J. W., & Rahnev, D. (2017). Stimulus expectation alters decision criterion but not sensory signal in perceptual decision making. *Scientific Reports*, *7*(1), 1-12.

Bao, S. (2015). Perceptual learning in the developing auditory cortex. *European Journal of Neuroscience*, *41*(5), 718-724.

Basharat, A., Mahoney, J. R., & Barnett-Cowan, M. (2019). Temporal metrics of multisensory processing change in the elderly. *Multisensory Research*, *32*(8), 715-744.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv Preprint arXiv:1406.5823*, 1-51.

Bates, S. L., & Wolbers, T. (2014). How cognitive aging affects multisensory integration of navigational cues. *Neurobiology of Aging*, *35*(12), 2761-2769.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412.

Beauchamp, M. S., Pasalar, S., & Ro, T. (2010). Neural substrates of reliability-weighted visual-tactile multisensory integration. *Frontiers in Systems Neuroscience*, *4*(25), 1-11.

Bedard, G., & Barnett-Cowan, M. (2016). Impaired timing of audiovisual events in the elderly. *Experimental Brain Research*, *234*, 331-340.

Bejjanki, V. R., Beck, J. M., Lu, Z. L., & Pouget, A. (2011). Perceptual learning as improved probabilistic inference in early sensory areas. *Nature Neuroscience*, *14*(5), 642-648.

Bennett, P. J., Sekuler, R., & Sekuler, A. B. (2007). The effects of aging on motion detection and direction identification. *Vision Research*, *47*(6), 799-809.

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.

Billino, J., & Pilz, K. S. (2019). Motion perception as a model for perceptual aging. *Journal of Vision*, *19*(4), 3-3.

Birren, J. E., & Fisher, L. M. (1995). Aging and speed of behavior: Possible consequences for psychological functioning. *Annual Review of Psychology*, *46*, 329-353.

Bizley, J. K., Jones, G. P., & Town, S. M. (2016a). Where are multisensory signals combined for perceptual decision-making?. *Current Opinion in Neurobiology*, *40*, 31-37.

Bizley, J. K., Maddox, R. K., & Lee, A. K. (2016b). Defining auditory-visual objects: behavioral tests and physiological mechanisms. *Trends in Neurosciences*, *39*(2), 74-85.

Bizley, J. K., Walker, K. M., Nodal, F. R., King, A. J., & Schnupp, J. W. (2013). Auditory cortex represents both pitch judgments and the corresponding acoustic cues. *Current Biology*, *23*(7), 620-625.

Blanz, V., & Vetter, T. (1999, July). A morphable model for the synthesis of 3D faces. In *Proceedings of the 26th Annual Conference on Computer Graphics and iInteractive Techniques* (pp. 187-194).

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. *Journal of Mathematical Psychology*, *87*, 46-75.

Boldt, A., Schiffer, A. M., Waszak, F., & Yeung, N. (2019). Confidence predictions affect performance confidence and neural preparation in perceptual decision making. *Scientific Reports*, *9*(1), 4031.

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 Journal of Neuroscience*, *46*(10), 2565-2577.

Brant, L. J., & Fozard, J. L. (1990). Age changes in pure-tone hearing thresholds in a longitudinal study of normal human aging. *The Journal of the Acoustical Society of America*, *88*(2), 813-820.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *Journal of Neuroscience*, *12*(12), 4745-4765.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1993). Responses of neurons in macaque MT to stochastic motion signals. *Visual Neuroscience*, *10*(6), 1157-1169.

Brunel, L., Carvalho, P. F., & Goldstone, R. L. (2015). It does belong together: cross-modal correspondences influence cross-modal integration during perceptual learning. *Frontiers in Psychology*, *6*, 358-368.

Brunetti, R., Indraccolo, A., Del Gatto, C., Spence, C., & Santangelo, V. (2018). Are crossmodal correspondences relative or absolute? Sequential effects on speeded classification. *Attention, Perception, & Psychophysics*, *80*(2), 527-534.

Bucur, B., Allen, P. A., Sanders, R. E., Ruthruff, E., & Murphy, M. D. (2005). Redundancy gain and coactivation in bimodal detection: Evidence for the preservation of coactive processing in older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *60*(5), P279-P282.

Burnham, K. P., & Anderson, D. R. (1998). Practical use of the information-theoretic approach. In *Model Selection and Inference* (pp. 75-117). Springer, New York, NY.

Calvert, G. A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal of Physiology-Paris*, *98*(1-3), 191-205.

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., Rouiller, E. M., & Barone, P. (2009). Multisensory anatomical pathways. *Hearing Research*, *258*(1-2), 28-36.

Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2010). Auditory–visual multisensory interactions in humans: timing, topography, directionality, and sources. *Journal of Neuroscience*, *30*(38), 12572-12580.

Cerella, J. (1985). Information processing rates in the elderly. *Psychological Bulletin*, *98*(1), 67-83.

Chandrasekaran, C. (2017). Computational principles and models of multisensory integration. *Current Opinion in Neurobiology*, *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), 1-22.

Chen, N., Bi, T., Zhou, T., Li, S., Liu, Z., & Fang, F. (2015). Sharpened cortical tuning and enhanced cortico-cortical communication contribute to the long-term neural mechanisms of visual motion perceptual learning. *NeuroImage*, *115*, 17-29.

Chiou, R., & Rich, A. N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception*, *41*(3), 339-353.

Choi, I., Lee, J. Y., & Lee, S. H. (2018). Bottom-up and top-down modulation of multisensory integration. *Current Opinion in Neurobiology*, *52*(1), 115-122.

Cienkowski, K. M., & Carney, A. E. (2002). Auditory-visual speech perception and aging. *Ear and Hearing*, *23*(5), 439-449.

Cerella, J. (1985). Information processing rates in the elderly. *Psychological Bulletin*, *98*(1), 67-83.

Cerella, J. (1990). Aging and information-processing rate. *Handbook of the Psychology of Aging*, *3*, 201-221.

Clemo, H. R., Keniston, L. P., & Meredith, M. A. (2012). Structural Basis of Multisensory Processing. *The Neural Bases of Multisensory Processes*.

Cohen, J. (1992a). Statistical power analysis. *Current Directions in Psychological Science*, *1*(3), 98-101.

Cohen, J. (1992b). A power primer. *Psychological Bulletin, 112*(1), 155–159.

Courchesne, E., Hillyard, S. A., & Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography and Clinical Neurophysiology*, *39*(2), 131-143.

Cullum, S., Huppert, F. A., McGee, M., Dening, T., Ahmed, A., Paykel, E. S., & Brayne, C. (2000). Decline across different domains of cognitive function in normal ageing: results of a longitudinal population- based study using CAMCOG. *International Journal of Geriatric Psychiatry*, *15*(9), 853-862.

Dahl, C. D., Logothetis, N. K., & Kayser, C. (2009). Spatial organization of multisensory responses in temporal association cortex. *Journal of Neuroscience*, *29*(38), 11924-11932.

Dakin, S. C., Hess, R. F., Ledgeway, T., & Achtman, R. L. (2002). What causes non-monotonic tuning of fMRI response to noisy images?. *Current Biology*, *12*(14), 476-477.

Davis, R. (1961). The fitness of names to drawings: A cross-cultural study in Tanganyika. *British Journal of Psychology*, 52, 259–268.

Deary, I. J., Corley, J., Gow, A. J., Harris, S. E., Houlihan, L. M., Marioni, R. E., Penke, L., Rafnsson, S. B., & Starr, J. M. (2009). Age-associated cognitive decline. *British Medical Bulletin*, *92*(1), 135-152.

De Dieuleveult, A. L., Siemonsma, P. C., Van Erp, J. B., & Brouwer, A. M. (2017). Effects of aging in multisensory integration: a systematic review. *Frontiers in Aging Neuroscience*, 9(80), 1-14).

Demattè, M. L., Sanabria, D., & Spence, C. (2006). Cross-modal associations between odors and colors. *Chemical Senses*, *31*(6), 531-538.

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.

Delis, I., Ince, R. A., Sajda, P., & Wang, Q. (2022). Neural encoding of active multi-sensing enhances perceptual decision-making via a synergistic cross-modal interaction. *Journal of Neuroscience*, *42*(11), 2344-2355.

Delong, P., & Noppeney, U. (2021). Semantic and spatial congruency mould audiovisual integration depending on perceptual awareness. *Scientific Reports*, *11*(1), 10832.

DeLoss, D. J., Pierce, R. S., & Andersen, G. J. (2013). Multisensory integration, aging, and the sound-induced flash illusion. *Psychology and Aging*, *28*(3), 802-812.

De Meo, R., Murray, M. M., Clarke, S., & Matusz, P. J. (2015). Top-down control and early multisensory processes: chicken vs. egg. *Frontiers in Integrative Neuroscience*, *9*, 1-17.

Diaconescu, A. O., Alain, C., & McIntosh, A. R. (2011). The co-occurrence of multisensory facilitation and cross-modal conflict in the human brain. *Journal of Neurophysiology*, *106*, 2896-2909.

Diaconescu, A. O., Hasher, L., & McIntosh, A. R. (2013). Visual dominance and multisensory integration changes with age. *NeuroImage*, *65*, 152-166.

Diaz, J. A., Delis, I., Philaistides, M. G. The neural signatures of the effects of prior probability on perceptual decision-making. Unpublished manuscript. The University of Glasgow.

Diaz, J. A., Queirazza, F., & Philiastides, M. G. (2017). Perceptual learning alters post-sensory processing in human decision-making. *Nature Human Behaviour*, *1*(2), 1-9.

Diederich, A. (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology*, *39*(2), 197-215.

Diederich, A. (2008). A further test of sequential-sampling models that account for payoff effects on response bias in perceptual decision tasks. *Perception & Psychophysics*, *70*(2), 229-256.

Diederich, A., Colonius, H., & Schomburg, A. (2008). Assessing age-related multisensory enhancement with the time-window-of-integration model. *Neuropsychologia*, *46*(10), 2556-2562.

Dobreva, M. S., O'Neill, W. E., & Paige, G. D. (2012). Influence of age, spatial memory, and ocular fixation on localization of auditory, visual, and bimodal targets by human subjects. *Experimental Brain Research*, *223*(4), 441-455.

Domenech, P., & Dreher, J. C. (2010). Decision threshold modulation in the human brain. *Journal of Neuroscience*, *30*(43), 14305-14317.

Doya, K. (2008). Modulators of decision making. *Nature Neuroscience*, *11*(4), 410-416.

Drugowitsch, J., DeAngelis, G. C., Angelaki, D. E., & Pouget, A. (2015). Tuning the speed-accuracy trade-off to maximize reward rate in multisensory decision-making. *Elife*, *4*, 1-11.

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.

Dully, J., McGovern, D. P., & O'Connell, R. G. (2018). The impact of natural aging on computational and neural indices of perceptual decision making: A review. *Behavioural Brain Research*, *355*, 48-55.

Dutilh, G., & Rieskamp, J. (2016). Comparing perceptual and preferential decision making. *Psychonomic Bulletin & Review*, *23*(3), 723-737.

Eimer, M. (2011). The face-sensitivity of the n170 component. *Frontiers in Human Neuroscience*, *5*, 119.

Elliott, D. B. (1987). Contrast sensitivity decline with ageing: a neural or optical phenomenon?. *Ophthalmic and Physiological Optics*, *7*(4), 415-419.

Elliott, D. B., Sanderson, K., & Conkey, A. (1990). The reliability of the Pelli-Robson contrast sensitivity chart. *Ophthalmic and Physiological Optics*, *10*(1), 21-24.

Elliott, D., Whitaker, D., & MacVeigh, D. (1990). Neural contribution to spatiotemporal contrast sensitivity decline in healthy ageing eyes. *Vision Research*, *30*(4), 541-547.

Engel, A. K., Senkowski, D., & Schneider, T. R. (2012). Multisensory integration through neural coherence. In *The neural bases of multisensory processes*. CRC Press/Taylor & Francis.

Erlich, J. C., Brunton, B. W., Duan, C. A., Hanks, T. D., & Brody, C. D. (2015). Distinct effects of prefrontal and parietal cortex inactivations on an accumulation of evidence task in the rat. *Elife*, *4*, e05457.

Ernst, M .O., & Banks, M .S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429-433.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162-169.

Eusop, E., Sebban, C., & Piette, F. (2001). Aging and cognitive slowing: example of attentional processes--evaluation procedures and related questions. *L'encephale*, *27*(1), 39-44.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*(1), 6-6.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience*, *22*(13), 5749-5759.

Fahle, M. (2005). Perceptual learning: specificity versus generalization. *Current Opinion in Neurobiology*, *15*(2), 154-160.

Falkenstein, M., Yordanova, J., & Kolev, V. (2006). Effects of aging on slowing of motor-response generation. *International Journal of Psychophysiology*, *59*(1), 22-29.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research 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. *Behavior Research Methods*, *39*(2), 175-191.

Ferrari, A., & Noppeney, U. (2021). Attention controls multisensory perception via two distinct mechanisms at different levels of the cortical hierarchy. *PLoS Biology*, *19*(11), e3001465.

Fetsch, C. R., Pouget, A., DeAngelis, G. C., & Angelaki, D. E. (2012). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience*, *15*(1), 146-154.

Filimon, F., Philiastides, M. G., Nelson, J. D., Kloosterman, N. A., & Heekeren, H. R. (2013). How embodied is perceptual decision making? Evidence for separate processing of perceptual and motor decisions. *Journal of Neuroscience*, *33*(5), 2121-2136.

Fisher, V. L., Dean, C. L., Nave, C. S., Parkins, E. V., Kerkhoff, W. G., & Kwakye, L. D. (2023). Increases in sensory noise predict attentional disruptions to audiovisual speech perception. *Frontiers in Human Neuroscience*, *16*(1), 1-14

Forstmann, B. U., Anwander, A., Schäfer, A., Neumann, J., Brown, S., Wagenmakers, E. J., Bogacz, R., & Turner, R. (2010). Cortico-striatal connections predict control over speed and accuracy in perceptual decision making. *Proceedings of the National Academy of Sciences*, *107*(36), 15916-15920.

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., Von Cramon, D. Y., Ridderinkhof, K. R., & Wagenmakers, E. J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences*, *105*(45), 17538-17542.

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E. J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Psychology*, *67*(1), 641-666.

Forstmann, B. U., Tittgemeyer, M., Wagenmakers, E. J., Derrfuss, J., Imperati, D., & Brown, S. (2011). The speed-accuracy tradeoff in the elderly brain: a structural model-based approach. *Journal of Neuroscience*, *31*(47), 17242-17249.

Fouragnan, E., Queirazza, F., Retzler, C., Mullinger, K. J., & Philiastides, M. G. (2017). Spatiotemporal neural characterization of prediction error valence and surprise during reward learning in humans. *Scientific Reports*, *7*(1), 1-18.

Fouragnan, E., Retzler, C., Mullinger, K., & Philiastides, M. G. (2015). Two spatiotemporally distinct value systems shape reward-based learning in the human brain. *Nature Communications*, *6*(1), 8107.

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. *Cognitive Brain Research*, *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., Ritter, W., & Murray, M. M. (2002). Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. *Journal of Neurophysiology*, *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. *Journal of Neuroscience*, *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. *Nature Communications*, *11*(1), 1-14.

Freigang, C., Schmidt, L., Wagner, J., Eckardt, R., Steinhagen-Thiessen, E., Ernst, A., & Rübsamen, R. (2011). Evaluation of central auditory discrimination abilities in older adults. *Frontiers in Aging Neuroscience*, *3*(6), 1-11.

Freiherr, J., Lundström, J. N., Habel, U., & Reetz, K. (2013). Multisensory integration mechanisms during aging. *Frontiers in human neuroscience*, *7*, 863.

Fusi, S., Miller, E. K., & Rigotti, M. (2016). Why neurons mix: high dimensionality for higher cognition. *Current Opinion in Neurobiology*, *37*(1), 66-74.

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics*, *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. *International Statistical Review*, *71*(2), 369-382.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457-472.

Getz, L. M., & Kubovy, M. (2018). Questioning the automaticity of audiovisual correspondences. *Cognition*, *175*, 101-108.

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory?. *Trends in Cognitive Sciences*, *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.

Gherman, S., & Philiastides, M. G. (2018). Human VMPFC encodes early signatures of confidence in perceptual decisions. *Elife*, *7*, e38293.

Gibson, E. J. (1963). Perceptual learning. *Annual Review of Psychology*, *14*(1), 29-56.

Gibson, E. J. (1969). Principles of perceptual learning and development. New York: Appleton-Century-Crofts.

Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron*, *31*(5), 681-697.

Glicksohn, A., & Cohen, A. (2013). The role of cross-modal associations in statistical learning. *Psychonomic Bulletin & Review*, *20*(6), 1161-1169.

Gold, J. I., & Ding, L. (2013). How mechanisms of perceptual decision-making affect the psychometric function. *Progress in Neurobiology*, *103*, 98-114.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*(1), 535-574.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*(1), 585-612.

Gondan, M., & Minakata, K. (2016). A tutorial on testing the race model inequality. *Attention, Perception, & Psychophysics*, *78*(3), 723-735.

Grech, R., Cassar, T., Muscat, J., Camilleri, K. P., Fabri, S. G., Zervakis, M., Xanthopoulosm ,P., Sakkalis V., & Vanrumste, B. (2008). Review on solving the inverse problem in EEG source analysis. *Journal of Neuroengineering and Rehabilitation*, *5*(1), 1-33.

Greenweld, A.G., McGhee, D.E., & Schwartz, J.L.K. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, *47*(6), 1464-1480.

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics* (Vol. 1). New York: Wiley.

Guerreiro, M. J., Anguera, J. A., Mishra, J., Van Gerven, P. W., & Gazzaley, A. (2014). Age-equivalent top–down modulation during cross-modal selective attention. *Journal of Cognitive Neuroscience*, *26*(12), 2827-2839.

Guerreiro, M. J., Murphy, D. R., & Van Gerven, P. W. (2013). Making sense of age-related distractibility: The critical role of sensory modality. *Acta Psychologica*, *142*(2), 184-194.

Guerreiro, M. J., Eck, J., Moerel, M., Evers, E. A., & Van Gerven, P. W. (2015). Top-down modulation of visual and auditory cortical processing in aging. *Behavioural Brain Research*, *278*, 226-234.

Habets, B., Bruns, P., & Röder, B. (2017). Experience with crossmodal statistics reduces the sensitivity for audio-visual temporal asynchrony. *Scientific Reports*, *7*(1), 1-7.

Hairston, W. D., Laurienti, P. J., Mishra, G., Burdette, J. H., & Wallace, M. T. (2003). Multisensory enhancement of localization under conditions of induced myopia. *Experimental Brain Research*, *152*, 404-408.

Hauser, C. K. and Salinas, E. (2014) ‘Perceptual Decision Making’, in Encyclopedia of Computational Neuroscience. New York, NY: Springer New York, pp. 1–21. https://doi.org/10.1007/978-1-4614-7320-6_317-1

Hazel, C. A., & Elliott, D. B. (2002). The dependency of logMAR visual acuity measurements on chart design and scoring rule. *Optometry and Vision Science*, *79*(12), 788-792.

Hawkins, G. E., Mittner, M., Forstmann, B. U., & Heathcote, A. (2017). On the efficiency of neurally-informed cognitive models to identify latent cognitive states. *Journal of Mathematical Psychology*, *76*, 142-155.

Hedden, T., & Gabrieli, J.D. (2004). Insights into the ageing mind: a view from cognitive neuroscience. *Nature Reviews Neuroscience*, *5*(2), 87-96.

Heekeren, H. R., Marrett, S., Bandettini, P. A., & Ungerleider, L. G. (2004). A general mechanism for perceptual decision-making in the human brain. *Nature*, *431*(7010), 859-862.

Heekeren, H. R., Marrett, S., & Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nature Reviews Neuroscience*, *9*(6), 467-479.

Heekeren, H. R., Marrett, S., Ruff, D. A., Bandettini, P. A., & Ungerleider, L. G. (2006). Involvement of human left dorsolateral prefrontal cortex in perceptual decision making is independent of response modality. *Proceedings of the National Academy of Sciences*, *103*(26), 10023-10028.

Heikkilä, J., Alho, K., Hyvönen, H., & Tiippana, K. (2015). Audiovisual semantic congruency during encoding enhances memory performance. *Experimental Psychology*, *62*(2), 123-130.

Henschke, J. U., Noesselt, T., Scheich, H., & Budinger, E. (2015). Possible anatomical pathways for short-latency multisensory integration processes in primary sensory cortices. *Brain Structure and Function*, *220*, 955-977.

Herding, J., Ludwig, S., von Lautz, A., Spitzer, B., & Blankenburg, F. (2019). Centro-parietal EEG potentials index subjective evidence and confidence during perceptual decision making. *NeuroImage*, *201*, 116011.

Hernández, B., Setti, A., Kenny, R. A., & Newell, F. N. (2019). Individual differences in ageing, cognitive status, and sex on susceptibility to the sound-induced flash illusion: A large-scale study. *Psychology and Aging*, *34*(7), 978.

Heron, J., Roach, N. W., Hanson, J. V., McGraw, P. V., & Whitaker, D. (2012). Audiovisual time perception is spatially specific. *Experimental Brain Research*, *218*(3), 477-485.

Hipp, J. F., & Siegel, M. (2013). Dissociating neuronal gamma-band activity from cranial and ocular muscle activity in EEG. *Frontiers in Human Neuroscience*, *7*, 338.

Hirst, R. J., Cassarino, M., Kenny, R. A., Newell, F. N., & Setti, A. (2022). Urban and rural environments differentially shape multisensory perception in ageing. *Aging, Neuropsychology, and Cognition*, *29*(2), 197-212.

Hirst, R. J., Setti, A., Kenny, R. A., & Newell, F. N. (2019). Age-related sensory decline mediates the Sound-Induced Flash Illusion: Evidence for reliability weighting models of multisensory perception. *Scientific Reports*, *9*(1), 19347.

Holmes, N. P. (2009). The principle of inverse effectiveness in multisensory integration: some statistical considerations. *Brain Topography*, *21*(3), 168-176.

Holmes, N. P., & Spence, C. (2005). Multisensory integration: space, time and superadditivity. *Current Biology*, *15*(18), R762-R764.

Hou, H., Zheng, Q., Zhao, Y., Pouget, A., & Gu, Y. (2018). Neural correlates of optimal multisensory decision making. *BioRxiv*, 480178.

Huang, Y., & Rao, R. P. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), 580-593.

Hugenschmidt, C. E., Peiffer, A. M., McCoy, T. P., Hayasaka, S., & Laurienti, P. J. (2009a). Preservation of crossmodal selective attention in healthy aging. *Experimental Brain Research*, *198*, 273-285.

Hugenschmidt, C. E., Mozolic, J. L., Tan, H., Kraft, R. A., & Laurienti, P. J. (2009b). Age-related increase in cross-sensory noise in resting and steady-state cerebral perfusion. *Brain Topography*, *21*, 241-251.

Huk, A. C., & Shadlen, M. N. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *Journal of Neuroscience*, *25*(45), 10420-10436.

Ince, R. A., Paton, A. T., Kay, J. W., & Schyns, P. G. (2021). Bayesian inference of population prevalence. *Elife*, *10*, e62461, 1-22.

Irvine, D. R. (2018). Auditory perceptual learning and changes in the conceptualization of auditory cortex. *Hearing Research*, *366*, 3-16.

Jamal, Y., Lacey, S., Nygaard, L., & Sathian, K. (2017). Interactions between auditory elevation, auditory pitch and visual elevation during multisensory perception. *Multisensory Research*, *30*(3-5), 287-306.

Jaworska, K., Yi, F., Ince, R. A., Van Rijsbergen, N. J., Schyns, P. G., & Rousselet, G. A. (2020). Healthy aging delays the neural processing of face features relevant for behavior by 40 ms. *Human Brain Mapping*, *41*(5), 1212-1225.

Jones, A. (2015). Independent effects of bottom-up temporal expectancy and top-down spatial attention. An audiovisual study using rhythmic cueing. *Frontiers in Integrative Neuroscience*, *8*(96), 1-9.

Jones, S. A., Beierholm, U., Meijer, D., & Noppeney, U. (2019). Older adults sacrifice response speed to preserve multisensory integration performance. *Neurobiology of Aging*, *84*, 148-157.

Jones, S. A., & Noppeney, U. (2021). Ageing and multisensory integration: A review of the evidence, and a computational perspective. *Cortex*, *138*, 1-23.

Kaiser, J., Lennert, T., & Lutzenberger, W. (2007). Dynamics of oscillatory activity during auditory decision making. *Cerebral Cortex*, *17*(10), 2258-2267.

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences*, *88*(11), 4966-4970.

Kayser, S. J., & Kayser, C. (2018). Trial by trial dependencies in multisensory perception and their correlates in dynamic brain activity. *Scientific Reports*, *8*(1), 3742.

Kayser, C., Petkov, C. I., Augath, M., & Logothetis, N. K. (2005). Integration of touch and sound in auditory cortex. *Neuron*, *48*(2), 373-384.

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.

Kelly, S. P., Corbett, E. A., & O'Connell, R. G. (2021). Neurocomputational mechanisms of prior-informed perceptual decision-making in humans. *Nature Human Behaviour*, *5*(4), 467-481.

Kelly, S. P., & O'Connell, R. G. (2015). The neural processes underlying perceptual decision making in humans: recent progress and future directions. *Journal of Physiology-Paris*, *109*(1-3), 27-37.

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.

Klapetek, A., Ngo, M. K., & Spence, C. (2012). Does crossmodal correspondence modulate the facilitatory effect of auditory cues on visual search?. *Attention, Perception, & Psychophysics*, *74*(6), 1154-1167.

Kim, J., Chung, Y. G., Park, J. Y., Chung, S. C., Wallraven, C., Bülthoff, H. H., & Kim, S. P. (2015). Decoding accuracy in supplementary motor cortex correlates with perceptual sensitivity to tactile roughness. *PloS one*, *10*(6), e0129777.

Kim, J. N., & Shadlen, M. N. (1999). Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nature Neuroscience*, *2*(2), 176-185.

Köhler, W. (1929). Gestalt Psychology. New York, USA: Liveright.

Köhler, W. (1947). Gestalt psychology (2nd Edition). New York, USA: Liveright.

Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, *114*(1), 19-28.

Kim, R. S., Seitz, A. R., & Shams, L. (2008). Benefits of stimulus congruency for multisensory facilitation of visual learning. *PLoS One*, *3*(1), e1532.

King, A. J., & Palmer, A. R. (1985). Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. *Experimental Brain Research*, *60*(3), 492-500.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS one*, *2*(9), 943-953.

Laurienti, P. J., Burdette, J. H., Maldjian, J. A., & Wallace, M. T. (2006). Enhanced multisensory integration in older adults. *Neurobiology of Aging*, *27*(8), 1155-1163.

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. *Experimental Brain Research*, *158*(4), 405-414.

Lauzon, S. Á., Abraham, A. E., Curcin, K., Butler, B. E., & Stevenson, R. A. (2022). The relationship between multisensory associative learning and multisensory integration. *Neuropsychologia*, *174*, 108336.

Law, C. T., & Gold, J. I. (2008). Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nature Neuroscience*, *11*(4), 505-513.

Lerche, V., & Voss, A. (2016). Model complexity in diffusion modeling: Benefits of making the model more parsimonious. *Frontiers in Psychology*, *7*(1324), 1-14.

Lee, A. K., Maddox, R. K., & Bizley, J. K. (2019). An object-based interpretation of audiovisual processing. *Multisensory Processes: The Auditory Perspective*, 59-83.

Lee, F. S., Matthews, L. J., Dubno, J. R., & Mills, J. H. (2005). Longitudinal study of pure-tone thresholds in older persons. *Ear and Hearing*, *26*(1), 1-11.

Lee, A., Ryu, H., Kim, J. K., & Jeong, E. (2018). Multisensory integration strategy for modality-specific loss of inhibition control in older adults. *International Journal of Environmental Research and Public Health*, *15*(4), 718.

Lewis, R., & Noppeney, U. (2010). Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. *Journal of Neuroscience*, *30*(37), 12329-12339.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764-766.

Lippert, M., Logothetis, N. K., & Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound. *Brain Research*, *1173*, 102-109.

Liu, Y., & Otto, T. U. (2020). The role of context in experiments and models of multisensory decision making. *Journal of Mathematical Psychology*, *96*, 102352.

Mahani, M. A. N., Sheybani, S., Bausenhart, K. M., Ulrich, R., & Ahmadabadi, M. N. (2017). Multisensory perception of contradictory information in an environment of varying reliability: evidence for conscious perception and optimal causal inference. *Scientific Reports*, *7*(1), 3167.

Mahoney, J. R., Cotton, K., & Verghese, J. (2019). Multisensory integration predicts balance and falls in older adults. *The Journals of Gerontology: Series A*, *74*(9), 1429-1435.

Mahoney, J. R., Li, P. C. C., Oh-Park, M., Verghese, J., & Holtzer, R. (2011). Multisensory integration across the senses in young and old adults. *Brain Research*, *1426*, 43-53.

Mahoney, J. R., & Verghese, J. (2020). Does cognitive impairment influence visual-somatosensory integration and mobility in older adults?. *The Journals of Gerontology: Series A*, *75*(3), 581-588.

Majka, P., Rosa, M. G., Bai, S., Chan, J. M., Huo, B. X., Jermakow, N., Meng, K. L., Yeonsook, S. T., Wolkowicz, I. H., Worthy, K. H., Rajan, R., Reser, D. H., Wójcik, D. K., Okano, H., & Mitra, P. P. (2019). Unidirectional monosynaptic connections from auditory areas to the primary visual cortex in the marmoset monkey. *Brain Structure and Function*, *224*, 111-131.

Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(3), 384-394.

Marks, L. E. (2004). Cross-modal interactions in speeded classification. In G. Calvert., C. Spence., & B.E. Stein (Eds.), *The Handbook of Multisensory Processes* (pp. 85-105). Cambridge, MA: The MIT Press.

Marks, L. E., Ben-Artzi, E., & Lakatos, S. (2003). Cross-modal interactions in auditory and visual discrimination. *International Journal of Psychophysiology*, *50*(1-2), 125-145.

Marmolejo-Ramos, F., Cousineau, D., Benites, L., & Maehara, R. (2015). On the efficacy of procedures to normalize Ex-Gaussian distributions. *Frontiers in Psychology*, *5*, 1548-1561.

Matsuhashi, M., Ikeda, A., Ohara, S., Matsumoto, R., Yamamoto, J., Takayama, M., ... & Shibasaki, H. (2004). Multisensory convergence at human temporo-parietal junction–epicortical recording of evoked responses. *Clinical Neurophysiology*, *115*(5), 1145-1160.

Mayr, S., Erdfelder, E., Buchner, A., & Faul, F. (2007). A short tutorial of GPower. *Tutorials in Quantitative Methods for Psychology*, *3*(2), 51-59.

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience*, *9*(5), 605-610.

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.

McDowd, J., Vercruyssen, M., & Birren, J. E. (2020). Aging, divided attention, and dual-task performance. In *Multiple-Task Performance* (pp. 387-414). CRC Press.

McGovern, D. P., Roudaia, E., Stapleton, J., McGinnity, T. M., & Newell, F. N. (2014). The sound-induced flash illusion reveals dissociable age-related effects in multisensory integration. *Frontiers in Aging Neuroscience*, *6*(250), 1-9.

McGovern, D. P., Roudaia, E., Newell, F. N., & Roach, N. W. (2016). Perceptual learning shapes multisensory causal inference via two distinct mechanisms. *Scientific Reports*, *6*(24673), 1-11.

McKoon, G., & Ratcliff, R. (2013). Aging and predicting inferences: A diffusion model analysis. *Journal of Memory and Language*, *68*(3), 240-254.

Mercier, M. R., & Cappe, C. (2020). The interplay between multisensory integration and perceptual decision making. *NeuroImage*, 116970, 1-16.

Mercier, M. R., Foxe, J. J., Fiebelkorn, I. C., Butler, J. S., Schwartz, T. H., & Molholm, S. (2013). Auditory-driven phase reset in visual cortex: human electrocorticography reveals mechanisms of early multisensory integration. *NeuroImage*, *79*, 19-29.

Meredith, M. A., & Stein, B. E. (1986a). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research*, *365*(2), 350-354.

Meredith, M. A., & Stein, B. E. (1986b). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, *56*(3), 640-662.

Mishra, J., & Gazzaley, A. (2014). Attentional control of multisensory integration is preserved in aging. *Cognitive Electrophysiology of Attention. Signals of the Mind*, 190-203.

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. *Cognitive Brain Research*, *14*(1), 115-128.

Molholm, S., Sehatpour, P., Mehta, A. D., Shpaner, M., Gomez-Ramirez, M., Ortigue, S., Dyke, J. P., Schwartz, T. H., & Foxe, J. J. (2006). Audio-visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. *Journal of Neurophysiology*, *96*(2), 721-729.

Moors, A., & De Houwer, J. (2006). Automaticity: a theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297-326.

Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(2), 133-136.

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Cognitive Brain Research*, *17*(1), 154-163.

Mozolic, J. L., Hugenschmidt, C. E., Peiffer, A. M., & Laurienti, P. J. (2008a). Modality-specific selective attention attenuates multisensory integration. *Experimental Brain Research*, *184*, 39-52.

Mozolic, J. L., Joyner, D., Hugenschmidt, C. E., Peiffer, A. M., Kraft, R. A., Maldjian, J. A., & Laurienti, P. J. (2008b). Cross-modal deactivations during modality-specific selective attention. *BMC Neurology*, *8*(1), 1-11.

Mozolic, J. L., Hugenschmidt, C. E., Peiffer, A. M., & Laurienti, P. J. (2012). Multisensory integration and aging. In: The Neural Bases of Multisensory Processes. CRC Press/Taylor & Francis, Boca Raton (FL); 2012. PMID: 22593868.

Murray, C. A., de Larrea-Mancera, E. S. L., Glicksohn, A., Shams, L., & Seitz, A. R. (2020). Revealing multisensory benefit with diffusion modeling. *Journal of Mathematical Psychology*, *99*, 102449-102549.

Murray, M. M., Eardley, A. F., Edginton, T., Oyekan, R., Smyth, E., & Matusz, P. J. (2018). Sensory dominance and multisensory integration as screening tools in aging. *Scientific Reports*, *8*(1), 1-11.

Murray, M. M., Thelen, A., Thut, G., Romei, V., Martuzzi, R., & Matusz, P. J. (2016). The multisensory function of the human primary visual cortex. *Neuropsychologia*, *83*, 161-169.

Navarro, D. J., & Fuss, I. G. (2009). Fast and accurate calculations for first-passage times in Wiener diffusion models. *Journal of Mathematical Psychology*, *53*(4), 222-230.

Naveh-Benjamin, M., Craik, F. I., Guez, J., & Kreuger, S. (2005). Divided attention in younger and older adults: effects of strategy and relatedness on memory performance and secondary task costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 520.

Nelson, L. S. (1998). The Anderson-Darling test for normality. *Journal of Quality Technology*, *30*(3), 298-299.

Newell, F., & Shams, L. (2007). New insights into multisensory perception. *Perception*, *36*(10), 1415-1417.

Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature*, *341*(6237), 52-54.

Nickl-Jockschat, T., Kleiman, A., Schulz, J. B., Schneider, F., Laird, A. R., Fox, P. T., Eickhoff, S.B., & Reetz, K. (2012). Neuroanatomic changes and their association with cognitive decline in mild cognitive impairment: a meta-analysis. *Brain Structure and Function*, *217*(1), 115-125.

Nidiffer, A. R., Diederich, A., Ramachandran, R., & Wallace, M. T. (2018). Multisensory perception reflects individual differences in processing temporal correlations. *Scientific Reports*, *8*(1), 1-15.

Noppeney, U. (2021). Perceptual inference, learning, and attention in a multisensory world. *Annual Review of Neuroscience*, *44*, 449-473.

Norman, J. F., Cheeseman, J. R., Pyles, J., Baxter, M. W., Thomason, K. E., & Calloway, A. B. (2013). The effect of age upon the perception of 3-D shape from motion. *Vision Research*, *93*, 54-61.

Norman, J. F., Crabtree, C. E., Bartholomew, A. N., & Ferrell, E. L. (2009). Aging and the perception of slant from optical texture, motion parallax, and binocular disparity. *Perception & Psychophysics*, *71*(1), 116-130.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*(12), 1729-1735.

O’Connell, R. G., Shadlen, M. N., Wong-Lin, K., & Kelly, S. P. (2018). Bridging neural and computational viewpoints on perceptual decision-making. *Trends in Neurosciences*, *41*(11), 838-852.

Odegaard, B., Wozny, D. R., & Shams, L. (2015). Biases in visual, auditory, and audiovisual perception of space. *PLoS computational biology*, *11*(12), e1004649.

Odegaard, B., Wozny, D. R., & Shams, L. (2016). The effects of selective and divided attention on sensory precision and integration. *Neuroscience Letters*, *614*, 24-28.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*(156869), 1-9.

Orchard-Mills, E., Alais, D., & Van der Burg, E. (2013). Cross-modal associations between vision, touch, and audition influence visual search through top-down attention, not bottom-up capture. *Attention, Perception, & Psychophysics*, *75*(8), 1892-1905.

Osgood, C. E. (1960). The cross-cultural generality of visual-verbal synesthetic tendencies. *Behavioral Science*, *5*(2), 146-169.

Otto, T. U., Dassy, B., & Mamassian, P. (2013). Principles of multisensory behavior. *Journal of Neuroscience*, *33*(17), 7463-7474.

Otto, T. U., & Mamassian, P. (2017). Multisensory decisions: the test of a race model, its logic, and power. *Multisensory Research*, *30*(1), 1-24.

Oyama T, Yamada H, Iwasawa H (1998) Synesthetic tendencies as the basis of sensory symbolism: a review of a series of experiments by means of semantic differential. Psychologia 41(1), 203–215.

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, *5*(5), 1-1.

Paramei, G. V. (2012). Color discrimination across four life decades assessed by the Cambridge Colour Test. *JOSA A*, *29*(2), A290-A297.

Paraskevoudi, N., Balcı, F., & Vatakis, A. (2018). “Walking” through the sensory, cognitive, and temporal degradations of healthy aging. *Annals of the New York Academy of Sciences*, *1426*(1), 72-92.

Pareek, V., Rallabandi, V. S., & Roy, P. K. (2018). A Correlational study between microstructural white matter properties and macrostructural gray matter volume across normal ageing: conjoint DTI and VBM analysis. *Magnetic Resonance Insights*, *11*, https://doi.org/10.1177/1178623X18799926.

Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences*, *111*(16), 6104-6108.

Parise, C. V., & Pavani, F. (2011). Evidence of sound symbolism in simple vocalizations. *Experimental Brain Research*, *214*(3), 373-380.

Parise, C., & Spence, C. (2008). Synesthetic congruency modulates the temporal ventriloquism effect. *Neuroscience Letters*, *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. *Experimental Brain Research*, *220*(3-4), 319-333.

Parise, C.V., & Spence, C. (2013). Audiovisual cross-modal correspondences in the general population. In Julia Simner & Edward M. Hubbard (Eds). *Oxford Handbook of Synesthesia* (1st Ed., pp. 790-815). Oxford. Oxfordshire. Oxford University Press.

Park, H., & Kayser, C. (2019). Shared neural underpinnings of multisensory integration and trial-by-trial perceptual recalibration in humans. *Elife*, *8*, e47001.

Park, H., & Kayser, C. (2020). Diversification of perceptual mechanisms underlying preserved multisensory behaviour in healthy aging. *bioRxiv*. https://doi.org/10.1101/2020.02.12.945949

Park, D. C., & Reuter-Lorenz, P. (2009). The adaptive brain: aging and neurocognitive scaffolding. *Annual Review of Psychology*, *60*, 173-196.

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., Christoforou, C., Gerson, A. C., Dyrholm, M., Luo, A., Wagner, M., Philiastides, M. G., & Sajda, P. (2007). Spatiotemporal linear decoding of brain state. *IEEE Signal Processing Magazine*, *25*(1), 107-115.

Parra, L. C., Spence, C. D., Gerson, A. D., & Sajda, P. (2005). Recipes for the linear analysis of EEG. *Neuroimage*, *28*(2), 326-341.

Pasalar, S., Ro, T., & Beauchamp, M. S. (2010). TMS of posterior parietal cortex disrupts visual tactile multisensory integration. *European Journal of Neuroscience*, *31*(10), 1783-1790.

Peiffer, A. M., Mozolic, J. L., Hugenschmidt, C. E., & Laurienti, P. J. (2007). Age-related multisensory enhancement in a simple audiovisual detection task. *Neuroreport*, *18*(10), 1077-1081.

Peiffer, A. M., Hugenschmidt, C. E., Maldjian, J. A., Casanova, R., Srikanth, R., Hayasaka, S., Burdette, J. H., Kraft, R. A., & Laurienti, P. J. (2009). Aging and the interaction of sensory cortical function and structure. *Human brain mapping*, *30*(1), 228-240.,

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195-203.

Philiastides, M. G., Diaz, J. A., & Gherman, S. (2017). Spatiotemporal characteristics and modulators of perceptual decision-making in the human brain. In *Decision Neuroscience* (pp. 137-147). Academic Press.

Philiastides, M. G., & Heekeren, H. R. (2009). Spatiotemporal characteristics of perceptual decision making in the human brain. In *Handbook of Reward and Decision Making* (pp. 185-212). Academic Press.

Philiastides, M. G., Heekeren, H. R., & Sajda, P. (2014). Human scalp potentials reflect a mixture of decision-related signals during perceptual choices. *Journal of Neuroscience*, *34*(50), 16877-16889.

Philiastides, M. G., Auksztulewicz, R., Heekeren, H. R., & Blankenburg, F. (2011). Causal role of dorsolateral prefrontal cortex in human perceptual decision making. *Current Biology*, *21*(11), 980-983.

Philiastides, M. G., Ratcliff, R., & Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. *Journal of Neuroscience*, *26*(35), 8965-8975.

Philiastides, M. G., & Sajda, P. (2006a). 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. (2006b). Causal influences in the human brain during face discrimination: a short-window directed transfer function approach. *IEEE Transactions on Biomedical Engineering*, *53*(12), 2602-2605.

Philiastides, M. G., & Sajda, P. (2007). EEG-informed fMRI reveals spatiotemporal characteristics of perceptual decision making. *Journal of Neuroscience*, *27*(48), 13082-13091.

Philiastides, M., Tu, T., & Sajda, P. (2021). Inferring macroscale network dynamics via EEG-fMRI fusion. *Annual Review of Neuroscience*, *44*, 315-334.

Pisauro, M. A., Fouragnan, E., Retzler, C., & Philiastides, M. G. (2017). Neural correlates of evidence accumulation during value-based decisions revealed via simultaneous EEG-fMRI. *Nature Communications*, *8*(1), 15808.

Pleger, B., Ruff, C. C., Blankenburg, F., Bestmann, S., Wiech, K., Stephan, K. E., Capilla, A., Friston, K. J., & Dolan, R. J. (2006). Neural coding of tactile decisions in the human prefrontal cortex. *Journal of Neuroscience*, *26*(48), 12596-12601.

Ploran, E. J., Nelson, S. M., Velanova, K., Donaldson, D. I., Petersen, S. E., & Wheeler, M. E. (2007). Evidence accumulation and the moment of recognition: dissociating perceptual recognition processes using fMRI. *Journal of Neuroscience*, *27*(44), 11912-11924.

Poliakoff, E., Ashworth, S., Lowe, C., & Spence, C. (2006). Vision and touch in ageing: crossmodal selective attention and visuotactile spatial interactions. *Neuropsychologia*, *44*(4), 507-517.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128-2148.

Powers, A. R., Hevey, M. A., & Wallace, M. T. (2012). Neural correlates of multisensory perceptual learning. *Journal of Neuroscience*, *32*(18), 6263-6274.

Proulx, M. J., Brown, D. J., Pasqualotto, A., & Meijer, P. (2014). Multisensory perceptual learning and sensory substitution. *Neuroscience & Biobehavioral Reviews*, *41*, 16-25.

Rach, S., Diederich, A., & Colonius, H. (2011). On quantifying multisensory interaction effects in reaction time and detection rate. *Psychological Research*, *75*(2), 77-94.

Raij, T., Ahveninen, J., Lin, F. H., Witzel, T., Jääskeläinen, I. P., Letham, B., Israeli, E., Sahyoun, C., Vasios, C., Stufflebeam, S., Hämäläinen, M., & Belliveau, J. W. (2010). Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. *European Journal of Neuroscience*, *31*(10), 1772-1782.

Ramachandran, V. S., & Hubbard, E. M. (2001a). Synaesthesia--a window into perception, thought and language. *Journal of Consciousness Studies*, *8*(12), 3-34.

Ramachandran, V. S., & Hubbard, E. M. (2001b). Psychophysical investigations into the neural basis of synaesthesia. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1470), 979-983.

Raposo, D., Kaufman, M. T., & Churchland, A. K. (2014). A category-free neural population supports evolving demands during decision-making. *Nature Neuroscience*, *17*(12), 1784.

Raposo, D., Sheppard, J. P., Schrater, P. R., & Churchland, A. K. (2012). Multisensory decision-making in rats and humans. *Journal of Neuroscience*, *32*(11), 3726-3735.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *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 Computation*, *20*(4), 873-922.

Ratcliff, R., Philiastides, M. G., & Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proceedings of the National Academy of Sciences*, *106*(16), 6539-6544.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*(2), 333-367.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260-281.

Ratcliff, R., Smith, P. L., & McKoon, G. (2015). Modeling regularities in response time and accuracy data with the diffusion model. *Current directions in Psychological Science*, *24*(6), 458-470.

Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging*, *16*(2), 323-341.

Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004a). A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychology and Aging*, *19*(2), 278-289.

Ratcliff, R., Thapar, A., & McKoon, G. (2004b). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language*, *50*(4), 408-424.

Ratcliff, R., Thapar, A., & McKoon, G. (2006a). Aging and individual differences in rapid two-choice decisions. *Psychonomic Bulletin & Review*, *13*(4), 626-635.

Ratcliff, R., Thapar, A., & McKoon, G. (2006b). Aging, practice, and perceptual tasks: a diffusion model analysis. *Psychology and Aging*, *21*(2), 353.

R Core Team (2022). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL https://www.R-project.org/

Redfern, M. S., Jennings, J. R., Martin, C., & Furman, J. M. (2001). Attention influences sensory integration for postural control in older adults. *Gait & Posture*, *14*(3), 211-216.

Redfern, M. S., Jennings, J. R., Mendelson, D., & Nebes, R. D. (2009). Perceptual inhibition is associated with sensory integration in standing postural control among older adults. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *64*(5), 569-576.

Regenbogen, C., Seubert, J., Johansson, E., Finkelmeyer, A., Andersson, P., & Lundström, J. N. (2018). The intraparietal sulcus governs multisensory integration of audiovisual information based on task difficulty. *Human Brain Mapping*, *39*(3), 1313-1326.

Rienäcker, F., Van Gerven, P. W., Jacobs, H. I., Eck, J., Van Heugten, C. M., & Guerreiro, M. J. (2020). The neural correlates of visual and auditory cross-modal selective attention in aging. *Frontiers in Aging Neuroscience*, *12*, 498978.

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 and Language*, *128*(1), 18-24.

Rhodes, M. G. (2004). Age-related differences in performance on the Wisconsin card sorting test: a meta-analytic review. *Psychology and Aging*, *19*(3), 482-494.

Rieck, J. R., Baracchini, G., & Grady, C. L. (2021). Contributions of brain function and structure to three different domains of cognitive control in normal aging. *Journal of Cognitive Neuroscience*, *33*(9), 1811-1832.

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. *Nature Communications*, *10*(1), 1-17.

Rohe, T., & Noppeney, U. (2015a). Sensory reliability shapes perceptual inference via two mechanisms. *Journal of Vision*, *15*(5), 22-22.

Rohe, T., & Noppeney, U. (2015b). Cortical hierarchies perform Bayesian causal inference in multisensory perception. *PLOS Biology*, *13*(2), e1002073.

Rohe, T., & Noppeney, U. (2016). Distinct computational principles govern multisensory integration in primary sensory and association cortices. *Current Biology*, *26*(4), 509-514.

Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience*, *22*(21), 9475-9489.

Romo, R., & de Lafuente, V. (2013). Conversion of sensory signals into perceptual decisions. *Progress in Neurobiology*, *103*, 41-75.

Roque, J., Lafraire, J., & Auvray, M. (2020). Audiovisual crossmodal correspondence between bubbles' size and pouring sounds' pitch in carbonated beverages. *Foods*, *9*(8), 966-985.

Romo, R., Lemus, L., & de Lafuente, V. (2012). Sense, memory, and decision-making in the somatosensory cortical network. *Current Opinion in Neurobiology*, *22*(6), 914-919.

Rosenthal, R., Cooper, H., & Hedges, L. (1994). Parametric measures of effect size. *The Handbook of Research Synthesis*, *621*(*2*), 231-244.

Ross, L. A., Molholm, S., Butler, J. S., Del Bene, V. A., & Foxe, J. J. (2022). Neural correlates of multisensory enhancement in audiovisual narrative speech perception: a fMRI investigation. *NeuroImage*, *263*, 119598.

Sadaghiani, S., Maier, J. X., & Noppeney, U. (2009). Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing. *Journal of Neuroscience*, *29*(20), 6490-6499.

Sagi, D., & Tanne, D. (1994). Perceptual learning: learning to see. *Current Opinion in Neurobiology*, *4*(2), 195-199.

Sajda, P., Philiastides, M.G., Heekeren, H.H., & Ratcliff, R. (2011) Linking neuronal variability to perceptual decision making via neuroimaging. In *Neuronal Variability and Its Functional Significance* (pp.214-232), Eds: Ding, M., Glanzman, D. Oxford, Oxford University Press.

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 Reviews in Biomedical Engineering*, *2*, 97-109.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*(3), 403.

Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological Psychology*, *54*(1-3), 35-54.

Salthouse, T. A. (2009). When does age-related cognitive decline begin?. *Neurobiology of Aging*, *30*(4), 507-514.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, *12*(3), 225-239.

Schaffhauser, T., Boubenec, Y., & Mamassian, P. (2021). EEG signature of evidence integration suggests distinct visual and auditory representation in multisensory context. *Journal of Vision*, *21*(9), 2627-2627.

Scheib, J. P., Stoll, S., & Randerath, J. (2023). Does aging amplify the rule-based efficiency effect in action selection?. *Frontiers in Psychology*, 14(1), 1-10.

Schroeder, C. E., & Foxe, J. (2005). Multisensory contributions to low-level, ‘unisensory’ processing. *Current Opinion in Neurobiology*, *15*(4), 454-458.

Schroeder, C. E., Smiley, J., Fu, K. G., McGinnis, T., O'Connell, M. N., & Hackett, T. A. (2003). Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing. *international Journal of Psychophysiology*, *50*(1-2), 5-17.

Scurry, A. N., Lovelady, Z., Lemus, D. M., & Jiang, F. (2021). Impoverished inhibitory control exacerbates multisensory impairments in older fallers. *Frontiers in aging neuroscience*, *13*, 700787.

Seilheimer, R. L., Rosenberg, A., & Angelaki, D. E. (2014). Models and processes of multisensory cue combination. *Current Opinion in Neurobiology*, *25*, 38-46.

Seitz, A. R. (2017). Perceptual learning. *Current Biology*, *27*(13), R631-R636.

Seitz, A. R., & Dinse, H. R. (2007). A common framework for perceptual learning. *Current Opinion in Neurobiology*, *17*(2), 148-153.

Sereno, M. I., & Huang, R. S. (2014). Multisensory maps in parietal cortex. *Current Opinion in Neurobiology*, *24*, 39-46.

Servant, M., & Evans, N. J. (2020). A diffusion model analysis of the effects of aging in the Flanker Task. *Psychology and Aging*, *35*(6), 831.

Shadlen, M. N., & Newsome, W. T. (1996). Motion perception: seeing and deciding. *Proceedings of the National Academy of Sciences*, *93*(2), 628-633.

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, *86*(4), 1916-1936.

Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends in Cognitive Sciences*, *14*(9), 425-432.

Shang, N., & Styles, S. J. (2023). Implicit Association Test (IAT) Studies Investigating Pitch-Shape Audiovisual Cross-modal Associations Across Language Groups. *Cognitive Science*, *47*(1), e13221.

Sheppard, J. P., Raposo, D., & Churchland, A. K. (2013). Dynamic weighting of multisensory stimuli shapes decision-making in rats and humans. *Journal of Vision*, *13*(6), 4-4.

Shi, Z., & Burr, D. (2016). Predictive coding of multisensory timing. *Current Opinion in Behavioral Sciences*, *8*, 200-206.

Shima, K., & Tanji, J. (1998). Both supplementary and presupplementary motor areas are crucial for the temporal organization of multiple movements. *Journal of Neurophysiology*, *80*(6), 3247-3260.

Siegel, M., Engel, A. K., & Donner, T. H. (2011). Cortical network dynamics of perceptual decision-making in the human brain. *Frontiers in Human Neuroscience*, *5*, 21.

Silva, A. E., Barakat, B. K., Jimenez, L. O., & Shams, L. (2017). Multisensory congruency enhances explicit awareness in a sequence learning task. *Multisensory Research*, *30*(7-8), 681-689.

Silva, D. M. R., & Bellini-Leite, S. C. (2020). Cross-modal correspondences in sine wave: Speech versus non-speech modes. *Attention, Perception, & Psychophysics*, *82*(3), 944-953.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*(3), 161-168.

Spear, P. D. (1993). Neural bases of visual deficits during aging. *Vision Research*, *33*(18), 2589-2609.

Spence, C., & Deroy, O. (2013). How automatic are crossmodal correspondences?. *Consciousness and Cognition*, *22*(1), 245-260.

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology*, *28*(2), 61-70.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971-995.

Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. *Annals of the New York Academy of Sciences*, *1296*(1), 31-49.

Spence, C., & Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony. *Current Biology*, *13*(1), 519-521.

Sperdin, H. F., Cappe, C., Foxe, J. J., & Murray, M. M. (2009). Early, low-level auditory-somatosensory multisensory interactions impact reaction time speed. *Frontiers in Integrative Neuroscience*, *3*(2), 1-10.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583-639.

Spitzer, B., & Blankenburg, F. (2011). Stimulus-dependent EEG activity reflects internal updating of tactile working memory in humans. *Proceedings of the National Academy of Sciences*, *108*(20), 8444-8449.

Starns, J. J., & Ratcliff, R. (2010). The effects of aging on the speed–accuracy compromise: Boundary optimality in the diffusion model. *Psychology and Aging*, *25*(2), 377.

Starrfelt, R., & Gerlach, C. (2007). The visual what for area: words and pictures in the left fusiform gyrus. *NeuroImage*, *35*(1), 334-342.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. The MIT Press, Cambridge, Massachusetts.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*(4), 255-266.

Stephens, M. A. (1974). EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, *69*(347), 730-737.

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society*, *8*(3), 448-460.

Stevens, J. C., & Marks, L. E. (1965). Cross-modality matching of brightness and loudness. *Proceedings of the National Academy of Sciences*, 54(2), 407–411.

Stevenson, R. A., Fister, J. K., Barnett, Z. P., Nidiffer, A. R., & Wallace, M. T. (2012). Interactions between the spatial and temporal stimulus factors that influence multisensory integration in human performance. *Experimental Brain Research*, *219*(1), 121-137.

Stevenson, R. A., Ghose, D., Fister, J. K., Sarko, D. K., Altieri, N. A., Nidiffer, A. R., Kurela, L. R., Siemann, J. K., James, T. W., & Wallace, M. T. (2014).

Identifying and quantifying multisensory integration: a tutorial review. *Brain Topography*, *27*, 707-730.

Stevenson, R. A., Nelms, C. E., Baum, S. H., Zurkovsky, L., Barense, M. D., Newhouse, P. A., & Wallace, M. T. (2015). Deficits in audiovisual speech perception in normal aging emerge at the level of whole-word recognition. *Neurobiology of Aging*, *36*(1), 283-291.

Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: task and stimulus dependencies. *Experimental Brain Research*, *227*, 249-261.

Stevenson, R. A., Wallace, M. T., & Altieri, N. (2014). The interaction between stimulus factors and cognitive factors during multisensory integration of audiovisual speech. *Frontiers in Psychology*, *5*, 352.

Stine, E. A. L., Wingfield, A., & Myers, S. D. (1990). Age differences in processing information from television news: The effects of bisensory augmentation. *Journal of Gerontology*, *45*(1), P1-P8.

St-Laurent, M., Abdi, H., Burianová, H., & Grady, C. L. (2011). Influence of aging on the neural correlates of autobiographical, episodic, and semantic memory retrieval. *Journal of Cognitive Neuroscience*, *23*(12), 4150-4163.

Summerfield, C., & De Lange, F. P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), 745-756.

Swenor, B. K., Wang, J., Varadaraj, V., Rosano, C., Yaffe, K., Albert, M., & Simonsick, E. M. (2019). Vision impairment and cognitive outcomes in older adults: the Health ABC Study. *The Journals of Gerontology: Series A*, *74*(9), 1454-1460.

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. *Scientific Reports*, *9*(1), 1-12.

Taitelbaum-Swead, R., & Fostick, L. (2016). The effect of age and type of noise on speech perception under conditions of changing context and noise levels. *Folia Phoniatrica et Logopaedica*, *68*(1), 16-21.

Talsma, D. (2015). Predictive coding and multisensory integration: an attentional account of the multisensory mind. *Frontiers in Integrative Neuroscience*, *9*(*19*), 1-13.

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*(9), 400-410.

Tatz, J. R., Undorf, M., & Peynircioğlu, Z. F. (2021). Effect of impoverished information on multisensory integration in judgments of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(3), 481-497.

Ten Oever, S., Schroeder, C. E., Poeppel, D., Van Atteveldt, N., & Zion-Golumbic, E. (2014). Rhythmicity and cross-modal temporal cues facilitate detection. *Neuropsychologia*, *63*, 43-50.

Thapar, A., Ratcliff, R., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychology and Aging*, *18*(3), 415.

Theisen, M., Lerche, V., von Krause, M., & Voss, A. (2021). Age differences in diffusion model parameters: A meta-analysis. *Psychological Research*, *85*, 2012-2021.

Tong, J., Li, L., Bruns, P., & Röder, B. (2020). Crossmodal associations modulate multisensory spatial integration. *Attention, Perception, & Psychophysics*, *82*(7), 3490-3506.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97-136.

Tremel, J. J., & Wheeler, M. E. (2015). Content-specific evidence accumulation in inferior temporal cortex during perceptual decision-making. *Neuroimage*, *109*(1), 35-49.

Troje, N. F., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research*, *36*(12), 1761-1771.

Turkkan, N., & Pham-Gia, T. (1993). Computation of the highest posterior density interval in Bayesian analysis. *Journal of Statistical Computation and Simulation*, *44*(3-4), 243-250.

Turner, B. M., Forstmann, B. U., Love, B. C., Palmeri, T. J., & Van Maanen, L. (2017). Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology*, *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.

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. *Journal of Neuroscience*, *36*(28), 7346-7352.

Tye-Murray, N., Sommers, M., Spehar, B., Myerson, J., & Hale, S. (2010). Aging, audiovisual integration, and the principle of inverse effectiveness. *Ear and Hearing*, *31*(5), 636.

Ursino, M., Cuppini, C., & Magosso, E. (2014). Neurocomputational approaches to modelling multisensory integration in the brain: a review. *Neural Networks*, *60*, 141-165.

Verdonck, S., Loossens, T., & Philiastides, M. G. (2022). Sensorimotor decisions rely on the entanglement of evidence and motor accumulation processes. *bioRxiv*, 2022-05.

Wasylyshyn, C., Verhaeghen, P., & Sliwinski, M. J. (2011). Aging and task switching: a meta-analysis. *Psychology and Aging*, *26*(1), 15.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638-667.

Wenzlaff, H., Bauer, M., Maess, B., & Heekeren, H. R. (2011). Neural characterization of the speed–accuracy tradeoff in a perceptual decision-making task. *Journal of Neuroscience*, *31*(4), 1254-1266.

Werner, S., & Noppeney, U. (2010). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *Journal of Neuroscience*, *30*(7), 2662-2675.

Wheeler, M. E., Woo, S. G., Ansel, T., Tremel, J. J., Collier, A. L., Velanova, K., Ploran, E.J., & Yang, T. (2015). The strength of gradually accruing probabilistic evidence modulates brain activity during a categorical decision. *Journal of Cognitive Neuroscience*, *27*(4), 705-719.

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, *58*(3), 475-482.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the drift-diffusion model in Python. *Frontiers in Neuroinformatics*, *7*, 1-14.

World Medical Association. (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *Jama*, *310*(20), 2191-2194.

Van der Burg, E., Alais, D., & Cass, J. (2018). Rapid recalibration to audiovisual asynchrony follows the physical—not the perceived—temporal order. *Attention, Perception, & Psychophysics*, *80*(8), 2060-2068.

Vandekerckhove, J., Tuerlinckx, F., & Lee, M.D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods*, *16*(1), 44-62.

Van Hooren, S. A. H., Valentijn, A. M., Bosma, H., Ponds, R. W. H. M., Van Boxtel, M. P. J., & Jolles, J. (2007). Cognitive functioning in healthy older adults aged 64–81: A cohort study into the effects of age, sex, and education. *Aging, Neuropsychology, and Cognition*, *14*(1), 40-54.

von Krause, M., Lerche, V., Schubert, A. L., & Voss, A. (2020). Do non-decision times mediate the association between age and intelligence across different content and process domains?. *Journal of Intelligence*, *8*(3), 33.

Vroomen, J., & Gelder, B. D. (2000). Sound enhances visual perception: cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1583-1590.

Yang, W., Guo, A., Li, Y., Qiu, J., Li, S., Yin, S., ... & Ren, Y. (2018). Audio-visual spatiotemporal perceptual training enhances the P300 component in healthy older adults. *Frontiers in Psychology*, *9*, 2537.

Yang, W., Li, S., Xu, J., Li, Z., Yang, X., & Ren, Y. (2020). Selective and divided attention modulates audiovisual integration in adolescents. *Cognitive Development*, *55*, 100922.

Yau, Y., Dadar, M., Taylor, M., Zeighami, Y., Fellows, L. K., Cisek, P., & Dagher, A. (2020). Neural correlates of evidence and urgency during human perceptual decision-making in dynamically changing conditions. *Cerebral Cortex*, *30*(10), 5471-5483.

Yordanova, J., Kolev, V., Hohnsbein, J., & Falkenstein, M. (2004). Sensorimotor slowing with ageing is mediated by a functional dysregulation of motor-generation processes: evidence from high-resolution event-related potentials. *Brain*, *127*(2), 351-362.

Zangaladze, A., Epstein, C. M., Grafton, S. T., & Sathian, K. (1999). Involvement of visual cortex in tactile discrimination of orientation. *Nature*, *401*(6753), 587-590.

Zaninotto, P., Batty, G. D., Allerhand, M., & Deary, I. J. (2018). Cognitive function trajectories and their determinants in older people: 8 years of follow-up in the English Longitudinal Study of Ageing. *Journal of Epidemiological and Community Health*, *72*(8), 685-694.

Zeljko, M., Kritikos, A., & Grove, P. M. (2019). Lightness/pitch and elevation/pitch crossmodal correspondences are low-level sensory effects. *Attention, Perception, & Psychophysics*, *81*(5), 1609-1623.

Zhang, J., & Rowe, J. B. (2014). Dissociable mechanisms of speed-accuracy tradeoff during visual perceptual learning are revealed by a hierarchical drift-diffusion model. *Frontiers in Neuroscience*, *8*(69), 1-13.

Zhao, S., Wang, C., Feng, C., Wang, Y., & Feng, W. (2022). The interplay between audiovisual temporal synchrony and semantic congruency in the cross-modal boost of the visual target discrimination during the attentional blink. *Human Brain Mapping*, *43*(8), 2478-2494.

Zigler, M. J. (1930). Tone shapes: A novel type of synaesthesia. *The Journal of General Psychology*, *3*(2), 277-287.

Zilber, N., Ciuciu, P., Gramfort, A., Azizi, L., & van Wassenhove, V. (2014). Supramodal processing optimizes visual perceptual learning and plasticity. *NeuroImage*, *93*, 32-46.