# Neural signatures of surprise in auditory learning 

Lance Abel BSci (Maths) BCom (Business Strategy/Economic Management)

Supervisor: Dr. Paul Sowman

Associate Supervisor: Jordan Wehrman



# MACQUARIE <br> University 

SYDNEY•AUSTRALIA

Faculty Of Medicine, Health, and Human Sciences
Department of Cognitive Science

A thesis submitted in partial fulfilment of the requirements for the degree of:
BPhil/Master of Research

Submitted on 15/7/2022

## Table Of Contents

Abstract .....
Statement Of Originality ..... vi
Acknowledgements ..... viii
Abbreviations ..... ix
Introduction ..... 1
Linking Homeostasis To Learning ..... 1
Predictive Processing ..... 2
Mismatch Responses (MMRs) ..... 3
Auditory MMRs ..... 3
Age ..... 4
Sensory Memory. ..... 5
Quantifying Expected Surprise ..... 6
The Hierarchical Gaussian Filter (HGF) ..... 6
Applications. ..... 8
Trajectories and transitions ..... 8
Modifying The HGF. ..... 10
The Auditory Roving Oddball Paradigm ..... 12
Eliminating Behaviour, Minimising Attention ..... 13
Magnetoencephalography (MEG) ..... 13
Statistical Analysis ..... 13
Summary ..... 15
Aims ..... 16
Method ..... 17
Participants ..... 17
Equipment ..... 18
Head Shape Measurement ..... 18
MEG ..... 18
Setup. ..... 18
Movement. ..... 18
The Stimuli ..... 19
Visual Stimuli ..... 19
Tones ..... 19
Past participants. ..... 19
New recruits. ..... 20
Number of tones ..... 20
Post-processing ..... 21
Generation Of Predictors ..... 22
Data Analysis Statistics ..... 23
Condition Split And Regression ..... 23
Age Splits ..... 24
Condition-by-age Interactions ..... 24
GFPs ..... 24
Regression On Brain Activity ..... 25
Permutation Cluster Analyses ..... 25
Times Of Interest (TOIs) ..... 26
Results ..... 27
Pre-processing metrics ..... 27
Correlation Between Predictive Models ..... 27
Post-processing findings ..... 29
Gross Activity Patterns ..... 29
Standardisation ..... 30
Intra-model differences ..... 30
Conditions expected to be different. ..... 31
i) Low Vs. High Surprise. ..... 31
ii) Very Low Vs. Very High Surprise. ..... 32
iii) PE2 Vs. PWPE2 And PE3 ..... 33
Repetition effects. ..... 36
Regression Of Brain Activity ..... 37
GFPs ..... 37
i) $P E 2$ ..... 37
ii) PE3 vs. PWPE2. ..... 38
Inter-model Comparison ..... 40
i) HGF_Vanilla vs. Standard/Deviants ..... 40
ii) PS vs. BS vs. CS ..... 41
iii) HGF_Vanilla vs. CS ..... 42
iv) HGF_Vanilla vs. HGF_Pitch. ..... 43
HGF_Pitch Highlights Age Differences ..... 45
Age Comparison ..... 47
a) Gross differences (GFP). ..... 47
The Peaks In Young And Older People. ..... 48
A Closer Look At Latency And Amplitude. ..... 49
The First Peak. ..... 50
The Second Peak ..... 51
GFP Differences At The Second Peak ..... 52
b) ERF Differences. ..... 53
Repetition Effects, Again. ..... 56
Greater anticipatory activity. ..... 57
Condition-by-age Interactions ..... 58
Are The Differences Artefactual? ..... 60
Discussion ..... 61
Summary of results ..... 61
Interpreting the results ..... 63
Inter-model ..... 63
Intra-model. ..... 63
Repetition effects ..... 64
Age. ..... 65
Strengths And Weaknesses ..... 67
Strengths ..... 67
Weaknesses, And Potential Mitigators ..... 68
Future Research ..... 70
Other types of stimuli ..... 70
Other Models ..... 70
Explaining More Cognitive Phenomena ..... 71
Source Analysis ..... 73
Conclusions ..... 74
References. ..... 75
Appendix A ..... 83
Appendix B ..... 84
Appendix C ..... 85
Appendix D ..... 86
Appendix E ..... 87
Appendix $\mathbf{F}$ ..... 89
Appendix G ..... 90
Appendix H. ..... 97
Appendix I ..... 98
Appendix $\mathbf{J}$ ..... 102
600 Hz vs. 700 Hz ..... 102
650 Hz vs. 800 Hz ..... 105
Appendix K. ..... 107
Appendix L ..... 110


#### Abstract

\section*{Introduction}

Auditory mismatch responses (aMMRs) are neural signatures of learning evident when animals hear a surprising sound. Various mathematical models have been advanced to predict the amplitude, latency, and location of aMMRs.


## Method

Using magnetoencephalography (MEG), I played auditory tones to participants using the Roving Oddball paradigm. I contrasted the spatiotemporal pattern of aMMRs and of gross brain activity (as measured by Global Field Power [GFP]) in different age groups, including in an understudied group aged 10-16. I used various mathematical models including the Bayesian Hierarchical Gaussian Filter (HGF) - and a variant of this that I modified to incorporate physical stimulus characteristics - to quantify the neural surprise likely to be evoked by auditory tones, examining whether this predicted brain activity.

## Results

I found that aMMR amplitudes increased on trials where greater surprise was predicted. With age, anticipatory activity increased and aMMR amplitudes decreased. Unexpectedly, this decrease with age was not monotonic and older individuals' aMMRs occurred at a lower latency than younger individuals' aMMRs. The HGF predicted the magnitude of surprise better than traditional models did, especially on trials when the model predicted greater surprise.

Incorporating physical pitch characteristics further improved model predictions. Unexpectedly, more repetitions of a tone were not always associated with less surprise and incorporating the precision of predictions and re-training the model continuously were not unambiguously found to improve predictions.

## Discussion

My results broadly support the utility of Bayesian accounts of learning, with older individuals displays less evidence of neural surprise and greater anticipatory activity in relation to predictable stimulus features. However, the pattern of aMMRs found also suggests that brains may not implement Bayesian learning over purely categorical features, but also use continuous stimulus information, and that the role of precision is more nuanced. Some of the results with respect to age are arguably discordant with Bayesian predictions and more consistent with the developmental emergence of structure learning based on stimulus saliency.

## Statement Of Originality

## Declaration

I certify that the work in this thesis entitled "Neural signatures of surprise in auditory learning" has not previously been submitted for a degree nor has any part of it been submitted toward the requirements for a degree at any other university or institution other than Macquarie University.

I also certify that the thesis is an original piece of research written by me. Any help and assistance that I have received in my research work and the preparation of the thesis itself have been appropriately acknowledged.

I further certify that all information sources and literature used are indicated in this thesis.
The research presented in this thesis was approved by the Macquarie University Human Research Ethics Committee, under the following reference numbers (see Appendix $K$ ):

- Child Participants Approval: 5201600188
- Amendment (2021): 52021190235599
- Adult Participants Approval: 5201800226
- Amendment (2021): 52021297935883

SIGNED: Lance Abel (SID 45161769)

DATE: 15/7/22

## Author's Note

This thesis is designed to be read with Adobe Acrobat and navigated using the "Bookmarks" pane.

## Acknowledgements

I would like to acknowledge:

- My principal supervisor, Dr Paul Sowman, whose expertise was invaluable. Paul trained me to use the child MEG machine and explained how it works, and what the data represents in physical and cognitive terms. He also provided core Python and MATLAB code and generously debugged my Python (once on a weekend when he was away!). Thanks also for your teaching during the coursework period, for inspiring interest by letting us play with TMS and TDCS, and for sending so many interesting articles to me since I started this Masters (and reading and replying to my pageslong responses!).
- My co-supervisor and editor extraordinaire, Jordan Wehrman. Jordan was always on point about the areas of his own research most applicable for me, whilst providing me with many more data and general neuroscience resources of interest to explore: I'm sorry I ran out of time to implement the (much-better) graphs you showed me. Jordan also helped me recruit people and participated twice as a guinea pig to help me iron out my lab skills and speed, which was invaluable in collecting data from external participants. Thanks too for reining in my ambitions and making me laugh every week about retro-causation research.
- Dr Nick Benikos, who trained me to use the MEG software, to safely use the lab and digitising equipment and who steered me towards pre-processing algorithms.
- Markus Ockenden for his technical research and innovation in the MEG lab simplifying data collection, and for changing the system to let me to collect data in the late evening!
- Craig Richardson for responding fast when I had a problem with the digitising equipment.
- The CogSci department for their feedback on my proposal, for promoting good research practises and for fostering interest in Cognitive Science throughout my coursework.
- All of my participants, who travelled into the lab, trusted me to implement COVID-safe procedures, responded graciously to my "big brother" monitoring of their movements during their scans, and who competed with each other to sit as still for 35 minutes as possible (most moved less than 2 mm in any 10-minute period!).
- My wife Melissa Chin for her patience and cooking during this period. I coded ~12000 lines for this thesis, which involved some cursing at the screen and some neglect at times.
- The Commonwealth Government of Australia for funding Universities and student research and who provided me a Research Training Program (RTP) scholarship.
- The Wallumattagal clan of the Dharug people, on whose land Macquarie University is located.


## Abbreviations

| Abbreviation | Definition |
| :---: | :---: |
| A1 | Primary auditory cortices, located bilaterally at the superior temporal gyrus in the temporal lobe. |
| aMMR | Auditory mismatch response: the spatiotemporal pattern of difference in ERF between trial types. |
| BIC | Bayesian Information Criterion: a type of goodness-of-fit measure evaluating how well a model's predictions match a set of observations. |
| DC bias | Where the mean amplitude of a waveform is not equal to zero, it has DC bias. |
| Deviant | An auditory tone of a different pitch to the previously presented stimulus. |
| EEG | Electroencephalography, a method of measuring voltages at the scalp. |
| Epoch | A peri-stimulus window of time: I used a 500 ms window spanning from 100 ms before the tone onset to 400 ms after the tone onset. |
| ERF | Event-related field: magnetic field strength (in Tesla) time-locked to an event. |
| ERP | Event-related potential: a recorded voltage at the scalp time-locked to an event. |
| fT | Femtotesla, equal to $1.0 \mathrm{x} * 10^{-15} \mathrm{Tesla}(\mathrm{T})$ |
| GFP | Global Field Power: In EEG literature, the RMS of the ERF is the GFP. |
| GFP aMMR | The GFP of the aMMR signal, a temporal-only version of the aMMR. <br> Calculated as the GFP for condition "B" minus the GFP for condition "A". |
| HGF | Hierarchical Gaussian Filter, a mathematical model that predicts neural surprise. |
| HGF_Vanilla | The whatworld sub-model of the HGF predictive model. |
| HGF_Pitch | A modification to HGF_Vanilla which produces different aMMR predictions based on physical tone characteristics, via modifying all layers of the HGF. |
| ISI | Inter-stimulus interval: the (constant) time gap between tones in the experiment. |
| kNN | k-nearest neighbours algorithm which I used for mapping child to adult sensors. |
| Hz | Hertz: frequency (per second throughout this thesis). |
| ICA | Independent components analysis: used for identifying artefactual components in the MEG data unrelated to brain activity (e.g., data linked to cardiac activity). |
| MEG | Magnetoencephalography: the measurement of the magnetic field generated by the electrical activity in the brain. |
| MMN | Mismatch negativity: the negative ERP typically found for a surprising event. |
| MMR | Mismatch response: the entire spatiotemporal pattern of MMN. |

## Abbreviation Definition

MSR
N200 (N2
wave)
P3 (P300) A positive-deflection in the ERP typically elicited around 300ms after a Deviant tone
$\mathrm{PE}[\mathrm{n}] \quad$ Prediction error [at the nth layer of the HGF].
PP Predictive processing: A theory of brain function in which the brain is constantly generating and updating a mental model of the environment

Pre-deviant
PWPE[n]
RANSAC
RMS

ROI The (standard) auditory tone played prior to the Deviant tone. Precision-weighted prediction error [at the nth layer of the HGF]. Random Sample Consensus algorithm: used for identifying bad channels. Root-mean-square: this was approximately equivalent to the standard deviation in my data, which exhibited a mean of zero.
Region of interest: an area in the brain believed (based on prior research) to contain information or support behaviour relevant for the analysis of the questions of interest.

SQUID Superconducting quantum interference device artefacts where field artefacts

Standard A tone which is recognised and expected by virtue of very recent repetition.
Surprise
T (Tesla)
TOI
Times of interest: a window of time post-stimulus believed (based on prior research) to contain neural data relevant to the research questions of interest. In this thesis, the TOI of focus was $130-190 \mathrm{~ms}$.
vMMR Visual mismatch response.

## Introduction

Humans are adept at recognising changes in their environment. Deviations from a steady state are surprising, attracting attention and driving learning. This thesis examines how neural markers of surprise to auditory tones change with age, and how surprise and its variation with age are influenced by the statistics of stimulus sequences. Characterising surprise and its normative course with age is important for understanding atypical cognition as well as developmental and neurodegenerative disorders computationally. It will also help us explain phenomena such as how language is learnt, and music appreciated.

## Linking Homeostasis To Learning

All organisms react to their environment, adapting to it to survive and reproduce. Some - such as plants, or bacteria which respond to pH , temperature, and the surrounding salt concentrations (Krasensky \& Jonak, 2012; Taylor, 1988) do so without a nervous system - let alone a brain. In animals however, nervous systems facilitate faster and more complex responses: to engage in fight or flight, to exploit and modify the environment by gathering food and information, and to find mating partners. Brains are enhancements to the nervous system (Arendt et al., 2016; Martinez \& Sprecher, 2020) which specialise in planning, co-ordinating and optimising such activities over longer time horizons and at greater complexity and scale; brains must work together with the body to predict physiological parameters and maintain homeostasis.

Maintaining homeostasis, however, requires responding not only to the body's internal needs, but also predicting and adapting to the external sensory environment (including how rewards and dangers in the environment are likely to influence the body). Psychologists have already identified learning mechanisms and their molecular substrates underpinning animals' ability to respond behaviourally e.g., via operant conditioning (Lee et al., 2021) or via classical conditioning without conscious awareness (Rose et al., 2005). However, such cognitive processes operate at a large scale. Less progress has been made in understanding the low-level learning processes which allow animals (and which might allow artificial intelligences) to predict the sensory environment, and that support complex behavioural responses. Moreover, just as developmental psychologists describe and explain how human psychology changes with an individual's age and experience, so too should researchers attempt to understand how such variables influence low-level learning processes.

Lastly, as the brain requires energy, a key challenge for researchers is to understand how normally functioning brains learn adequately (or even optimally) in an energetically efficient way.

## Predictive Processing

To these ends, researchers have begun explaining the low-level cognitive processes of learning. One popular framework for understanding this is predictive processing (PP; Friston, 2010). PP posits that perceptions and learning arise from interactions between two processes (Rao \& Ballard, 1999). One is a top-down, intra-cortical process whereby the brain's 'model' of the environment generates sensory expectations (predictions) in a manner constrained by that organism's sensory abilities and its life experience. The second is a bottom-up process: whenever reality deviates from these predictions, prediction error (PE) signals travel up a thalamo-cortical neural hierarchy, creating a neural signature of surprise. Learning then - in the PP framework arises when PEs cause an update of the model's parameters or beliefs (Faraji et al., 2018). The ecological purpose of learning is therefore to adapt individuals to their environment by helping them to predict it (either passively or - where advantageous - through actively exploring and modifying it iteratively). Our expectations thus depend upon our previous beliefs, the speed of learning, and any actions that we take to control the environment.

Researchers have studied these (top-down) predictions and (bottom-up) surprise in various ways. A first, model-free method involves explicitly querying participants about their predictions and their phenomenological experience of surprise, noting that surprise is distinct from novelty (Barto et al., 2013) (although novelty can often prompt surprise). A second involves calculating participants' surprise on the basis of past probabilities: for example, the "Shannon" surprise (Tribus, 1961) that an individual should experience in response to a digit in a sequence increases when that digit has appeared less frequently before. A third involves deriving participants' surprise from behavioural data using some mapping (e.g., assuming faster reactions imply less surprise e.g., see Marshall et al., 2016 and Powers et al., 2017). A fourth involves calculating participants' expectations according to a model of their belief formation; these beliefs could range from potentially idiosyncratic and irrational (Griffiths et al., 2019) to Bayes-optimal (Stefanics et al., 2018). When a mismatch exists between a model and reality, updates to a model then occur (Gijsen et al., 2021) and this model can be confidence-corrected by specifying the confidence in an observation or belief. A fifth and final method is by inferring expectations and/or surprise from neural activity, which is the focus of the current study. I will however also link these signatures of neural surprise to underlying models of belief formation (the fourth method) and analyse the relationship between these.

## Mismatch Responses (MMRs)

A canonical neural index of surprise found using electroencephalography (EEG) is the mismatch negativity (MMN; Näätänen et al., 1978). The MMN is calculated by subtracting the event-related potential (ERP) voltage evoked by an unexpected ('deviant') stimulus (e.g., a tone of unexpectedly low or high pitch) from the ERP after an expected ('standard') stimulus (e.g., the same tone, after multiple repetitions); the magnetic equivalent of ERPs are referred to as eventrelated fields (ERFs). Researchers now refer to mismatch responses (MMRs) as a more general term which subsumes the MMN and which includes overlapping components (e.g., the $\mathbf{N} 200$, see Tsogli et al., 2019). MMRs have attracted attention as they are found across multiple modalities and contexts (Hedge et al., 2015), signifying their broad relevance to perception and learning in many brain networks. MMRs also occur for tasks which do not involve behavioural responses, suggesting their involvement in 'pure' (passive) learning (Meyniel et al., 2016). Finally, MMR amplitudes are known to correlate with how unexpected a stimulus should have been based on statistical likelihood (Sarasso et al., 2021) and with apparent surprise based on reaction times (McCarthy et al., 2002), making them an ideal neural marker of surprise to study.

## Auditory MMRs

Auditory MMRs (aMMRs) are one type of MMR and occur after a tone differing in pitch, intensity, location, or duration from the anticipated tone (Phillips et al., 2015). aMMRs relating to pitch (perceived frequency) are most evident in bilateral A1 (superior temporal gyrus), superior temporal sulcus, orbitofrontal cortex, and inferior frontal gyrus (Cheng, et al., 2013; Wacongne et al., 2011; King et al., 2014) and occur $130-200 \mathrm{~ms}$ after a deviant (Bonetti et al., 2018).

Researchers previously believed that aMMRs were artefacts of physiological adaptation (e.g., see the fresh afferent theory: May \& Tiitinen, 2010), aspects discussed by O'Reilly \& O'Reilly (2021), or other purely bottom-up processes arising from the mere detection of a change in the world. However, these hypotheses seem incompatible with three observations. Firstly, the repetition of a stimulus - which should cause adaptation - can instead increase surprise (Alain et al., 1999; Tervaniemi et al., 1994). Secondly, the omission of an expected stimulus can actually cause MMRs (Hughes et al., 2001; Ouden et al., 2009; Wacongne et al., 2011; Salisbury, 2012). This is significant because neural activity after an omitted stimulus must purely relate to the absent anticipated stimulus i.e., is in no way driven by bottom-up activity from an unanticipated sound. aMMR amplitudes are better predicted by computational models which assume observers generate predictions which are Bayes-optimal relative to their priors than by models which assume neural activity is driven by physiological adaptation or mere detection of tone changes e.g., bottom up sensory information (Lieder et al., 2013). These forms of evidence all suggest that aMMRs instead
index model-driven learning (a concept which also subsumes the familiar phenomenon of forgetting).
Measuring surprise neurally using aMMRs, SanMiguel et al., (2021) further established that aMMRs - like other MMRs - depend not only on the magnitude of PE but on the prediction's expected precision (i.e., confidence, or inverse variance: see Hsu et al., 2019). This precision is likely physiologically represented in the synaptic gain of superficial pyramidal neurons in auditory cortex (Moran et al., 2013). The precision of predictions thus modifies the PE into a precisionweighted prediction error (PWPE) governing the rate of learning (Friston, 2010; Lecaignard et al., 2021).


#### Abstract

Age Another important dimension influencing aMMRs is age (Kisley et al., 2005; Cheng et al., 2013). For example, smaller aMMRs were reported (Rapaport et al., 2022) in those aged 17-38 compared to those aged 3-9, and older individuals display reduced somatosensory MMRs compared to younger individuals (Strömmer et al., 2014) and reduced MMN amplitude in EEG experiments (Kisley et al., 2005). Potential causes for this have also been proposed: for example, (Moran et al., 2014) found an attenuation of synaptic connectivity associated with slower learning rates in older adults. This is consistent with a shift in cortical response from sensory (posterior) regions to executive (anterior) regions and increased aMMR latency found by van Dinteren et al., (2018) - although such research did not examine those under 20 or whether aMMR amplitudes reduce with age.


Such age-related changes have often been framed as exclusively an undesirable functional decline (Cheng, et al., 2013). Seemingly justifying this, aMMRs are known to reduce in amplitude in the presence of Alzheimer's disease (Pekkonen, 2000). Age-related changes may however also be conceived as a (Bayes)-optimal response to the available sensory evidence for two reasons. Firstly, reduced aMMRs might be mediated by the increased sensory experience accumulated by older individuals. Indeed, if prior data predicts the sensory environment well, reducing the speed of belief updating both reduces errors (by maintaining stable perceptions of reality robust to noise) and saves the energy required to update beliefs. A second explanation is advanced by Moran et al., (2014). They acknowledge that increasingly fixed expectations driven by top-down processing with age may be caused by a decline - in the form of noisier sensory processing (e.g., worse hearing) - but qualify that increasingly fixed expectations are an optimal response to this new data of lower quality: slower learning thus does not necessarily indicate poor cognition.

Age is also important because intriguing evidence is emerging that neurodevelopmental disorders may be explained by changes in the computational basis for sensory learning (Baldeweg et al., 2004). Moreover, some such disorders emerge at specific ages e.g., autism at 6-18 months (Tanner \& Dounavi, 2021) and schizophrenia at 18-25 (Gogtay et al., 2011). Therefore, if schizophrenia and autism indeed relate to such low-level changes in learning and perception, it is important to characterise the usual course of learning with age to better understand whether a person's cognition is consistent with a neurodevelopmental disorder or not.

Whilst there is evidence for a predictive coding account of neurocognitive development, there are age groups that have been inadequately studied (e.g., Kisley et al., 2005; Rapaport et al., 2022). For example, Rapaport et al., (2022) compared participants aged 17-38 with those aged 3-9. Because they lacked persons aged 10-16 and older than 39, it remains unclear from their data what trajectory the amplitude, latency and location of neural surprise follow with respect to age. For example, aMMRs might not change monotonically with age but may instead reverse over different age spans, just as height and fluid intelligence increase - but then decrease - with age (Hedden \& Gabrieli, 2004). Indeed, some ERP components like the P3 (van Dinteren et al., 2018) are already known to change over short time periods at ages less than 18, and indeed to peak at a certain age being more marked in younger children (Taylor, 1988). aMMRs might similarly be of greatest amplitude not at birth (where life experience is minimal, which is when a Bayesian might predict) but instead during adolescence as a result of an increase in predictive precision e.g., after more advanced verbal language training. To begin to be able to rule out such hypotheses, I also gather data from participants aged 10-16 and those over 39 .

## Sensory Memory

Age may influence aMMRs and learning speed via sensory memory specifically - a possibility which researchers have not considered thoroughly in predictive coding analyses of learning. Many environmental regularities are discoverable only by using the 'buffer storage' of working memory (Korovkin et al., 2018). It therefore is likely of significance that working memory declines with age (Pliatsikas et al., 2019). For example, in the Roving Oddball paradigm (detailed below), after a stimulus repeats the maximum number of times, it always changes: recognising this with high certainty requires a working memory for the previous seven tones in my experiment. To the extent that older individuals do not automatically partly or wholly learn such regularities, their aMMRs may differ to those of younger people and hence may be considered to index the duration of sensory memory in the auditory system (Pekkonen, 2000).

## Quantifying Expected Surprise

A crucial question in the literature has been how to predict neural surprise. Under the deviant/standard binary classification, all 'deviant' tones are expected to produce an equal amount of neural surprise. This is implausible, and research shows many other factors influence the neural response - such as how many repetitions of the 'standard' occurred prior to the 'deviant' tone; this is often called "repetition positivity' and thought to result from strengthened "memory trace" (Todd et al., 2013). Likewise, standards are defined either in a way which discards trials (if only the trial preceding the deviant becomes the standard) or are regarded as standards regardless of how many times they were repeated. The deviant/standard classification also discards other information e.g., how volatile the tones have been and the physical tone characteristics e.g., whether a specific deviant has ever been heard before. A solution to these problems is the method of maintaining a model of beliefs and updating its parameters.

## The Hierarchical Gaussian Filter (HGF)

One sequential Bayesian learning model capable of incorporating such factors into its prediction of neural surprise is the Hierarchical Gaussian Filter (HGF: Mathys et al., 2014), depicted in Figure 1:

Figure 1
A representation of a 3-layer Hierarchical Gaussian Filter (HGF)


Note: This figure is adapted from Lomakina (2014). See the update equations for this model in Appendix $B$.

The HGF is a generative PP model which computes belief trajectories trial-by-trial using the stimulus history. The pitch that we actually experience is the layer second from the bottom in Figure 1 above, i.e., perceived auditory tone. Whilst our perceptions are affected by the physical sound input from the layer below (i.e., the external world), they are also impacted by top-down predictions flowing downwards from layers (in the brain) above, which neurally encode beliefs and which are known to be responsible for auditory illusions (Groppe et al., 2010). In response to auditory tones and their perceived pitch, prediction errors occur on layer $i\left(\delta_{i}\right)$ and are multiplied by the layer's precision $\Psi_{\mathrm{i}}$ to produce precision-weighted PEs on each level (PWPEs, $\epsilon_{\mathrm{i}}$ ). These PWPEs flow upward, and represent the degree of model update (i.e., the speed of learning), which should predict aMMR amplitude. Importantly, as described by Mathys et al., (2011), the HGF's learning speed is dynamic, changing in response to the uncertainty around tone transitions. This is unlike models which assume a fixed rate of belief updating (e.g., the Rescorla-Wagner model: Rescorla, \& Wagner, 1972) and is likely superior as it allows modelling of approximately Bayes-optimal belief updating, consistent with the free-energy principle which has proven to have greater predictive power for aMMRs (Lieder et al., 2013). As shown in Figure 2, the HGF models several types of uncertainty:

Figure 2
The types of uncertainty modelled bv the HGF (also see Figure 1 above)


Notes: $\boldsymbol{\pi}_{\boldsymbol{\mu}}$ represents estimated precision (inverse variance). $\boldsymbol{\mu}_{\mathbf{n}}$ represents the mean estimate on the nth later, $\boldsymbol{\sigma}_{\mathbf{n}}$ represents the informational uncertainty. The meanings of $\boldsymbol{\kappa}$ and $\boldsymbol{\omega}$ are explained in Appendix B. Superscripts indicate the trial sequence number.

The first is informational uncertainty, which is uncertainty about the "rules of the game"; this uncertainty can be reduced with more observations. The second is environmental uncertainty, which are changes in these "rules of the game" over time. The third is outcome uncertainty: the irreducible uncertainty experienced every trial due to e.g., the inherent randomness in a roll of the dice.

The HGF's parameters also allow for more dimensions along which learning may differ between individuals. For example, the top layer outputs posterior beliefs about the volatility of tone frequency on the basis of prior beliefs about the volatility of tone frequency, uncertainty of these beliefs, and the observed volatility. The volatility estimate influences the second layer, which outputs beliefs about tone frequencies based on prior beliefs about tone frequencies and their volatility.

Applications. Supporting the utility of the PP framework - and the HGF especially - to model learning and neurodevelopmental disorders, prior research with the HGF concluded neural surprise may increase in schizophrenia due to abnormally imprecise priors on environmental volatility $\left(\Psi_{3}\right)$, causing perception of greater environmental uncertainty which yields greater delusion (Sterzer et al., 2018) and paranoia (Deserno et al., 2020; Reed et al., 2020). Researchers have also applied the HGF to suggest that autism may be characterised by abnormally low sensory noise (Kéita et al., 2011) or uncertainty on prediction errors (Van de Cruys et al., 2014), or alternatively/in combination with abnormally high precision of prior beliefs (Pellicano \& Burr, 2012). These may enhance observations (i.e., increase PWPE2 magnitudes) which speeds up learning, but speeded learning also implies potentially unstable cognition when the environment is volatile. Anatomical and synaptic causes corresponding to these computational-level explanations have been discovered, such as reduced brain connectivity in schizophrenia (Erdeniz et al., 2017) and altered lateral inhibition (Kéita et al., 2011; Puts et al., 2014) or laminar processing (Pak et al., 2021) in autism. Through the lens of the HGF, key behavioural features of autism e.g., repetitive behaviour may be seen as attempts to reduce environmental uncertainty. Furthermore, distinct neuromodulators in specific brain regions have been associated with surprise and belief updating processes (Schwartenbeck et al., 2016). Marshall et al., (2016) found HGF parameter changes are induced by the antagonism of noradrenaline and acetylcholine, molecules known to enhance bottom-up feedforward processing of sensory information relative to top-down feedback processing.

The HGF, deployed in various forms in the TAPAS toolbox (Frässle et al., 2021) for MATLAB (MathWorks, Inc) also offers a computationally inexpensive means of calculating Bayesian belief trajectories - which have no closed-form solution (Mathys et al., 2011) - using a mean field approximation. This fast calculation provides researchers with an (as yet, not fully exploited) opportunity to mimic the assumed recalibration of the brain's model of the auditory environment on a trial-by-trial basis when predicting aMMR magnitudes.

Trajectories and transitions. To simulate belief trajectories, a particular sub-model of HGF, hgf_whatworld, appears appropriate (hereafter referred to as HGF_Vanilla). This model computes the likelihood of all possible tone transitions (which to the observer are a "hidden Markov" process: Yoon, 2009) and the confidence associated with these transitions. Predicting transitions is important for two reasons. Firstly, the brain cares about the order of transitions (Squires et al., 1976; Maheu et al., 2019) and, consequently, unusually sequenced events are surprising even if the events themselves are familiar (Meyniel et al., 2016).
(e.g., the likelihood of a word in a sentence) are insensitive to word order (i.e., grammar) and so are inadequate to reflect speech and how it is learnt. Thus, understanding fluent speech requires segmenting words by their syllable transition probabilities (Saffran et al., 1996). Causal learning more broadly requires modelling transitions from potential causes to effects, not merely learning the frequency of associations, repetitions, or alternations (Kolossa et al., 2012). The HGF allows for dynamic learning of non-stationary patterns, leading to better performance than "fixed-belief" models that assume no change in the quantity they estimate (Meyniel et al., 2016). Such flexibility is important for learning applications beyond language (which is largely stable) i.e., for anything in a changing world. The transition model thus reflects real-world learning demands, including automatic attempts to learn the transitions in a sequence, such as in music (Brattico et al., 2006). Indeed, transition models cope better with 'local' patterns deviating from global frequency estimates than simpler models such as alternation models (McCarthy et al., 2002) - showing better goodness-of-fit and Bayesian Information Criterion (BIC, which adjusts this by penalising complex models: see Spolladore et al., 2021). Corroborating this, surprise predicted by transition models better correlates with neural surprise measures like the P3 and behavioural data like reaction times (Cho et al., 2002; McCarthy et al., 2002), including in non-auditory modalities (e.g., see Ramachandran et al., 2016).

A second reason to model learning of transitions is that neural surprise reflects more than just the summary statistics of sensory data. For example, if the brain expects only a low or a high-pitched tone with equal probability, it is surprised by a tone of intermediate pitch, despite such a tone being the expected value in frequency ( $\mathbf{H z}$ ) terms (Chumbley et al., 2014). Thus, surprise derives from more than the degree of deviation of a tone stimulus from the average sound in an auditory environment. Transition models reflect this by assuming that the auditory environment is modelled by the brain as only involving transitions to tones at discrete frequencies which have previously been heard.

## Modifying The HGF

However, by assuming the brain's auditory model specifies only discrete and pre-specified categories (and therefore only categories and their probabilities govern surprise and learning), transition models abstract too much away from continuous physical stimulus characteristics and the underlying neural architecture supporting their perception. For example, HGF_Vanilla predicts that deviants of any equally probable pitch change cause the same degree of surprise (if their precisions are also the same), regardless of how much the deviant differs in pitch from the preceding standard. This cannot be consistent with transitions of tiny magnitudes (smaller than the just-noticeable difference, $\mathbf{J N D}$ ), since parts of our perceptual machinery e.g., the basilar membrane respond to a range of frequencies (Ruggero, 1992) and would therefore respond identically to two similar tones. A potential solution to this problem involves incorporating the error in predicted acoustic frequency (pitch) as a multiplier to PE2. Indeed, research suggests aMMR amplitudes do depend on the pitch change (Picton et al., 2000), so this is well motivated. I modified HGF_Vanilla (producing what I call the HGF_Pitch model, available on Github) to achieve this: my model retains the advantages of the HGF_Vanilla categorical transition model whilst also assuming that the magnitude of pitch change during a transition influences aMMR amplitude. As seen in Figure 3, the HGF_Pitch model applies a multiplier proportional to the logarithm of the tone change in line with the logarithmic "place code" tonotopic mapping created by the basilar membrane at the applicable frequencies (Oxenham, 2013):

Figure 3
A simplified formula for the multiplier to the expected surprise in the HGF_Pitch model (in Hz )

$$
\text { multiplier }=\log _{\mathrm{e}}\left(\frac{\text { perceptual distance in units }}{\text { maximum perceptual distance }}\right)+\mathrm{c}
$$

Note: Predicted surprise is non-zero when a tone repeats (justifying the constant $c$ ). See Appendix $D$ for more detail.

The effect, as seen in this spreadsheet on Github, is to multiply the degree of expected surprise by a value that is specific to the tone transition displayed in Figure 4:

Figure 4
The multiplier applied to each tone transition for modifying the degree of expected surprise in HGF Pitch

| tone | 500 Hz | 550 Hz | 600 Hz | 650 Hz | 700 Hz | 750 Hz | 800 Hz |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 500 Hz | 0.69 | 1.01 | 1.18 | 1.27 | 1.33 | 1.37 | 1.40 |
| 550 Hz |  | 0.69 | 0.80 | 1.05 | 1.17 | 1.24 | 1.30 |
| 600 Hz |  |  | 0.69 | 0.69 | 0.98 | 1.12 | 1.20 |
| 650 Hz |  |  |  | 0.69 | 0.69 | 0.94 | 1.08 |
| 700 Hz |  |  |  |  | 0.69 | 0.69 | 0.90 |
| 750 Hz |  |  |  |  |  | 0.69 | 0.69 |
| 800 Hz |  |  |  |  |  |  | 0.69 |

As seen in Figure 5 below, HGF_Pitch predicts surprise, as shown by blue dots, that is smaller than that predicted by HGF_Vanilla when the pitch change is smaller (circled in red) and larger when the pitch change is greater (circled in green):

Figure 5
Prediction error (PE2) in the HGF_Pitch and HGF_Vanilla models


Notes: The models' PE2 differs on high-surprise trials. Over longer times, the lower band declines, suggesting surprise decreases with experience as participant model the tone generation process.

The modified PE2 in HGF_Pitch propagates, altering its other outputs (PWPE2, PE3, PWPE3). For example, when transitioning from the lowest (or highest) possible tone, the expected value and standard deviation of the magnitude of the pitch change from these extremities increases; consequently, after these tones, HGF_Pitch's volatility estimate increases relative to HGF_Vanilla. Moreover, such transitions occur only to higher or lower tones respectively (i.e., the sign of the change is known), so surprise probably relates more strongly to the magnitude of the pitch change).

Optimistically, the structure of the HGF may describe the brain's fundamental architecture. Even if this is true however, I had to consider how to set the HGF's four free parameter values ( $\mu_{2}, \sigma_{2}, \mu_{3}$, and $\sigma_{3}$ ) which determine the starting values of an individual's beliefs. For example, Weber et al., (2020) supports setting the initial value of $\mu_{2 \mathrm{ij}}$, the belief about tone transitions, to be neutral (i.e., assuming all transitions are equally likely) and I followed this practise. Researchers need more data, however, to understand whether these assumptions are valid, or whether understanding neurodevelopmental differences also requires an understanding of how priors and/or the brain architectures which may support them influence learning specific to an individual or a species.

## The Auditory Roving Oddball Paradigm

The auditory roving oddball paradigm (Garrido et al., 2008) provides a suitable stimulus train to test such models, and the independent contributions of prediction and precision to learning. As shown in Figure $\mathbf{6}$ below, it involves playing a 'standard' tone selected from 5-10 tones, that is drawn from a stationary distribution (usually a uniform one, so that the global probability of each tone is identical). This standard tone is repeated a number of times, this number of times being drawn from a frequency distribution - usually a skewed normal. Then, a different ('deviant' or 'oddball') tone - marked in orange below - plays, which then becomes the new standard (marked in blue) through repetition. For example, this stimulus sequence in my experiment, as described in Tones below, appears as follows:

Figure 6
The frequency of the first 30 tones that were played to all new recruits in this Roving Oddball experiment


Notes: Tones described as 'standards' are marked in blue and repeat a certain number of times.
Tones described as 'oddballs', or 'deviants' are marked in orange and should cause greater surprise.

The relatively small number of tones means it is at least a-priori feasible that the transition between all tone combinations (and uncertainty surrounding these) may be being modelled neurally. The tones typically employed are below 2 kHz , the frequency at which age-related hearing loss begins to occur (Salvi et al., 2018) and are sufficiently spaced in time and in apparent pitch (multiples of the just-noticeable-difference; Long, 2014) - so that anyone with intact hearing can distinguish them.

## Eliminating Behaviour, Minimising Attention

The roving oddball paradigm also allows us to examine aMMRs without requiring responses or attention, which helps to collect more data (as each trial is faster and less tiring for the subject) and is useful for studying populations who cannot respond (e.g., the physically disabled) or attend (e.g., those in a coma). As mentioned by Näätänen (2000), this will allow comparison of our data with those of such patients in future studies. The lack of response also serves to avoid the fraught task of mapping behavioural data (e.g., reaction times) to surprise (e.g., assuming fast reactions imply low surprise).

The paradigm uses a visual distractor task which tries to control for attention by (attempting to) eliminate it. This is unlike tasks which require attention to the stimuli (either due to low stimulus salience, high stimulus complexity, or spatial dependence e.g., visual tasks involving object tracking). This lack of attention means any findings will apply to automatic learning, increasing the importance of the phenomenon. Eliminating attention is feasible as pitch-aMMRs are known to occur despite distractions (Takegata et al., 2005) when the differences between standard and deviant tones are simple (Picton et al., 2000). Eliminating attention is necessary because attention nonetheless modulates aMMR amplitude (Auksztulewicz \& Friston, 2015; Woldorff et al., 1998), boosting aMMRs via enhanced precision. The presence of attention might thus otherwise confound comparisons between younger participants (who have e.g., reduced ability to maintain concentration: see Fortenbaugh et al., 2015) and older participants, especially given our data collection period of 15 minutes was not very short.

## Magnetoencephalography (MEG)

MEG has been widely used to study acoustic processing and plasticity in the auditory system (e.g., Kluge et al., 2011). This is because MEG can achieve better signal-to-noise ratio and readability than EEG in MMR experiments (Thönnessen et al., 2008; Hämäläinen et al., 1993; Strauss et al., 2015). MEG also shares EEG's excellent temporal resolution and exhibits superior spatial resolution (Singh, 2014).

## Statistical Analysis

Another opportunity to improve upon prior experiments is in the statistical analysis of aMMRs. Previous experiments have not fully utilised the information extracted from the brain. As mentioned, one practise has been to label every tone as either a deviant or a standard. Analysing tones using this simple dichotomy discards much information about factors which contribute to neural surprise: not all deviants or standards are created equal. Two examples of information lost which have been shown to matter are the proportion of trials that are standards and how many repetitions there had been of the standard before the deviant (Todd et al., 2013). Whilst such factors
can be tested as separate predictors, such model bloat is ill-advised and likely unnecessary. Indeed, the HGF natively incorporates such information through Bayes-optimal belief updating. With its single continuous class structure, the degree of surprise naturally changes on a trial-by-trial basis in response to such variables. For example, deviants become automatically less likely after many repetitions of the standard. To summarise, the HGF does not even require us to impose labels of deviant or standard on the data, nor to assume that each tone always belongs to one or the other class. Indeed, this lack of manual labelling may facilitate unsupervised explanation of learning in far less regular auditory environments in which humans struggle to label the features and may free us to discover what factors make the HGF's prediction more or less accurate.

A second practise has been to examine neural surprise too narrowly e.g., at a pre-specified location or latency. Rapaport et al., (2022) however, concluded that it was not appropriate to compare aMMRs associated with a single component across all participants, since the latency of the aMMR changes with age (Cooper et al., 2006; Schiff et al., 2008). Indeed, even within an age group, the components of brain signals (e.g., the N200 and MMN) overlap (Tsogli et al., 2019). The latency of MMRs also changes for other reasons, such as modality or type of tone mismatch (Bonetti et al., 2018), especially where MMRs summate (Takegata et al., 1999). This is both because of differences in how fast that modality is processed and due to inherent differences in the complexity of stimuli across and within modalities. For example, within a modality, some stimuli have greater complexity in the Shannon information sense. Such complexity is often modelled to influence surprise but could equally be argued to influence latencies to the extent that information is compressed, buffered, and otherwise processed before being compared to prior information stored in the brain. Indeed, different statistical regularities appear to be encoded at different latencies: stimulus likelihoods broadly appear to be encoded earlier than stimulus transition probabilities (Maheu et al., 2019). This raises issues with other analyses which have quantified surprise using narrow features e.g., the peak difference in ERP, or the area under the MMR curve (e.g., Beauchemin \& De Beaumont, 2005). It may instead be that a combination of features e.g., the shape of the whole waveform is required to describe one type of surprise. For example, low-latency stimulus-bound surprise may change longer-latency model updating (Gijsen et al., 2021). Overall, for the reasons described, it appears preferable to use an unrestricted analysis of the entire time (and possibly sensor space) to classify what constitutes neural surprise in response to a stimulus, such as in mass-univariate approaches (Groppe et al., 2011). As discussed in the Method, to determine the significance of results, I will use spatiotemporal cluster testing (Maris \& Oostenveld, 2007).


#### Abstract

Summary In summary, psychologists have learnt many broad facts about human learning, insights which were obtainable through observation with the naked eye. This knowledge has been exploited over decades e.g., in behavioural training programs, including in clinical settings. Research into aMMRs, analysed using the powerful explanatory framework of predictive processing is now yielding insights about much lower-level, higher-frequency (i.e., more pervasive and automatic) aspects of cognition and learning only discoverable by examining brain activity. In particular, it appears the brain uses generative models with hidden state variables to understand and predict environmental contingencies optimally and efficiently. Such mechanisms likely apply at various layers of the neural hierarchy, explaining different kinds of surprise and learning and behavioural data at various latencies. An improved understanding of how such generative models work may yield insights into neurotypical human perception (and aesthetics, e.g., music appreciation), surprise and learning. It may also contribute to our nascent understanding of the emergence of neurodevelopmental disorders - at least at the computational level-as resulting from a cascade of differences in priors, low-level surprise, and belief-updating. This may allow us to build better diagnostic tools for such conditions.

Given the large space of models which can describe learning, there is however a need to understand which models actually do so with some generality for persons of all ages. For this reason, I contrasted the existing HGF_Vanilla model with traditional Deviant vs. Standard classification, with other generative models, and with my HGF_Pitch model. My research with unstudied age groups will either provide further confirmatory evidence on the utility of Bayesian models in explaining automatic learning of auditory tones in human populations or will generate important counterexamples which motivate further modelling and research.


#### Abstract

Aims

Firstly, I wanted to replicate that the HGF would better predict aMMRs than the simple Deviant vs. Standard binary classification (Fitzgerald \& Todd, 2020). In seeking to understand why the HGF appears to offer superior predictive power, I also wanted to test which of the HGF's outputs (PE2, PWPE2, PE3) are most predictive of neural activity (in particular, of ERFs, of GFP and GFP aMMRs): I expected PWPE2 to be most predictive on the basis of prior literature. I also wanted to identify if predictive power is better on trials of low or high modelled surprise.

Secondly, I wanted to test a modified HGF to resolve whether the magnitude of pitch change influences the amplitude and/or latency of neural surprise: I expected that the amplitude of neural surprise would be larger for larger pitch changes. Similarly, I sought to classify and contrast trials along various dimensions in order to determine whether specific factors - such as the number of stimulus repetitions - influenced neural surprise.

Thirdly, I wanted to determine if older participants' aMMRs differ from younger participants with respect to their amplitude, latency, or gross location e.g., whether there is a shift in cortical response from sensory (posterior) regions to executive (anterior) regions. Relatedly, I wished to determine whether there is any interaction between age and how neural surprise varies with modelled neural surprise.

Lastly, I wanted to synthesise, with the aid of new data from an understudied group, whether the latency and amplitude of the GFP and ERFs varied smoothly with age.

To address these questions, I used MEG to determine the amplitude, latency, and location of aMMRs in response to specific types of trials. I analysed how these varied based on age and the surprise predicted by different generative models.


## Method

## Code I made for this thesis is available at https://github.com/LanceAbel/MQ MEG Analysis

## Participants

Past data from Rapaport et al., (2022) was combined with data from new participants. Past data was obtained from 81 individuals: 56 children (aged $\leq 10$ ), and 25 participants aged $\geq 17$. Of these, 10 children had been excluded due to excessive head movement. Nine children and six adults were excluded due to co-registration errors of $>5 \mathrm{~mm}$. Two children and one adult were excluded due to excessive noise in many sensors, one child and one adult due to inability to detect the auditory tones, and two children whose previously collected data appeared to be missing.

New recruits were paid $\$ 40 \mathrm{p} / \mathrm{h}$ to participate. They were screened for ADHD, Autism, and several other neurodevelopmental disorders. None reported having visual or auditory processing impairments, all showed normal hearing (see Hearing test. below) and none withdrew. Young participants were recruited using Macquarie's Neuronauts database and via word of mouth; older participants via Macquarie's Older Adult Participant Register and flyers at a dentist. Candidates were identified sequentially to equalise numbers by sex and in specific age ranges: participants were invited at random from the available candidates. Participants were excluded if they had nonremovable metal inside their body: four participants' sessions were discontinued due to the presence of metal in their mouth not declared during screening. An additional 23 participants were recruited, producing a final sample size of 72 ( 39 males, 33 females) with the age distribution ( $M=$ 13.72 years, $S D=11.84$ years) visible in Figure 7:

Figure 7
A histogram showing the number of participants in various age ranges


This included 65 right-handed individuals, 4 left-handed individuals, 1 ambidextrous individual and two individuals whose handedness had not been recorded.

## Equipment

## Head Shape Measurement

Participants wore a polyester MEG cap chosen to fit tightly, thus shifting minimally on the scalp. Three fiducials (the participants' nasion, and bilateral pre-auricular notches) were marked electronically using a digitiser pen (Polhemus Fastrak, Colchester, USA) or iPad application, along with the location of five head position indicator (HPI) 'marker' coils on their MEG cap and 20005000 points on the head and face.

## MEG

Adults used a 160-channel axial first-order gradiometer system (Model PQ1160RN2 by Kanazawa Institute of Technology/Yokogawa (Japan); see He et al., 2019). Children used a 125channel system (Model PQ1064R-N2m). Both systems recorded continuous MEG at 1000 Hz . The machines differ subtly e.g., the adult system has a greater concentration of sensors over frontal regions (where aMMRs occur; see Cheng, Baillet, et al., 2013), as the child helmet is designed to be more open (so the machine is less daunting). To compare the data from these systems, a relabelling process was conducted whereby the 160 channels in the adult system were transformed to a common 125-channel montage (i.e., mapped onto equivalent sensors in the 125 -channel child system, with the unmapped channels then dropped). To do this, a k-nearest-neighbours (kNN) clustering algorithm (with $k=1$ ) was used to select the sensor pairs with the shortest Euclidean distance. Paired channels were highly correlated; see Appendix L - Child vs Adult MEG systems for details.

Setup. Subjects lay supine in a magnetically shielded room (MSR) with their head in the helmet of the MEG system appropriate for their brain size. Participants positioned their head centrally within the MEG helmet, and their lower body as they wished to lie comfortably still. Coregistration of the marker coils was checked and up to two sensors were excluded as necessary to reduce this error to $<5 \mathrm{~mm}$. Participants had a median co-registration error of $3.28 \mathrm{~mm}(S D=$ 1.34 mm ).

Movement. Participants were given feedback on how much they moved in between blocks. Participants moved a median of only 1.5 mm per block ( $S D=5.0 \mathrm{~mm}$ ); those that moved $>5 \mathrm{~mm}$ were told they had moved more than others and were asked to re-centre themselves. The small degree of movement led to reliable information about the head position relative to the sensors.

## The Stimuli

## Visual Stimuli

Participants were given MEG-compatible glasses to aid vision where necessary whilst watching a silent video (with subtitles). The video was projected above their heads (1m away for the first 79 participants, and 20 cm for the final two participants due to a projector malfunction). Participants were instructed to minimise blinking, to attend to the video continuously, to ignore the tones and to return their focus to the video upon awareness of any sound. Consequently, the video was chosen by the participant with the goals of preventing boredom and controlling for (by eliminating) attention to the tones as far as possible - as per previously collected data. Visual checking of new recruits within the MSR confirmed none fell asleep and queries made of the participants after the experiment suggested that none had difficulty following the instructions.

## Tones

All participants were exposed to tones generated and played using MATLAB (Mathworks, Natick, MA, USA) with Cogent 2000 (http://www.vislab.ucl.ac.uk/cogent.php v1.32). All tones played had a duration of 70 ms with 10 ms rise and fall times and an inter-stimulus interval (ISI) of 500 ms . The tones were pure sinusoids in seven 50 Hz increments ranging from $500 \mathrm{~Hz}-800 \mathrm{~Hz}$ as within this range, participants aged up to 70 do not experience appreciable hearing loss (Park et al., 2016) - as was borne out in the results of the Hearing test shown below. Tones were played using a MEG-compatible 60x60cm high-quality speaker (Panphonics SSH Sound Shower) positioned centrally at the foot of the bed, 2 m away from the participant's chin where they registered at an intensity of 80 dB (as per Haenschel et al., 2005, Rapaport et al., 2022). It is further noted that differences in neural activity between tones of the same intensity but different pitch (i.e., pitch aMMRs) should vary less with absolute sound intensity than single-trial responses to tones of different intensity but identical pitch.

Hearing test. Participants' hearing thresholds were tested for both ears using pure-tone audiometric testing. A staircase procedure confirmed all participants had hearing thresholds $\leq 40 \mathrm{~dB}$ at both 250 Hz and 1000 Hz in both ears. Pilot tests also showed that eight participants aged 8,35 , $36,37,39,44,70$ and 71 were all $90 \%$ or more accurate in identifying the tone of higher pitch for 20 pairs of two tones adjacent in pitch within the $500-800 \mathrm{~Hz}$ range. This is consistent with JND data, which suggests tones are discriminable at $\leq 0.2-0.3 \%$ of the tone frequency in these ranges (Huanping \& Micheyl, 2011). Past participants. As per the roving auditory oddball paradigm (Garrido et al., 2008), tones were played in sequences with one to seven repetitions of a given tone
followed by a different tone drawn with uniform probability from the alternatives, with a distribution that is skewed with respect to the number of repetitions $(M=5.18, S D=1.69)$ as visible in Figure 8:

Figure 8
The distribution of the number of repetitions of a tone which occurred before the pitch changed


Notes: The data is shown over all 72 participants. The shape suggests the brain may be more surprised by a pitch change after four repetitions than one after five (see Results section).

New recruits. An identical sequence was played for all new recruits. This sequence however preserved the key statistics of those presented to prior participants: the number of repetitions of each tone was drawn from the same distribution and the tone transition probabilities were also uniform. Consequently, the frequency of each tone as well as the mean and standard deviation of the tone frequency were nearly identical

Number of tones. For all participants, the first 1730 tones played during the 15 -minute block were compared. This was a subset of the total number of tones shown to previous participants (which was not controlled but was usually $\sim 1770-1780$ depending on variations in PC speed), and to new recruits (which was capped at 1800).

## Post-processing

I coded a pipeline to process raw MEG data using the following sequence (broadly in keeping with practises employed by others e.g., see Andersen, 2018):

1. Head movement in the child MEG system was corrected for using data from a head monitoring system (ReTHM, which provides a combination of frequent hardware-based polling of the head-position and software-based correction; Rapaport et al., 2019; Cheour et al., 2004).
2. The TSPCA algorithm (de Cheveigné \& Simon, 2007) implemented in the MEG160 software (Yokogawa Electric Corp., Eagle Technology Corp., and the Kanazawa Institute of Technology) was run to filter out magnetic fields unrelated to brain activity (such as trains evident during recording), using three reference channels within the MSR away from the brain: the TSPCA algorithm also applied 0.3 Hz high-pass and 200 Hz low-pass filters.
3. MNE's annotate_flat function was used to automatically detect flat segments. Such segments include saturations (SQUID artifacts, which are extended periods where field measurements are larger than can be measured by the dynamic range of the system).
4. MNE's find_bad_channels_maxwell filter was used to automatically detect and interpolate noisy and flat channels caused by machine-specific problems. Additionally, channels noted previously as containing excessive high-frequency noise were manually excluded.
5. Noise from the electrical grid at 50 Hz and at harmonics of 50 Hz was removed. Data was bandpass filtered from $0.5-40 \mathrm{~Hz}$, to remove DC bias drifts due to perspiration as well as signals relating to muscle tension (e.g., see Shaw \& Bhaga, 2012).
6. Epochs were formed containing the data from 100 ms pre-stimulus to 400 ms post-stimulus.
7. Epochs were mean-centred by subtracting the mean of the entire epoch from each sample.
8. The latency of tone onset after the MEG trigger was calculated using the raw sound waveform, by identifying the first amplitude $\geq 3$ times above the peak immediately prior to the stimulus. Each Epoch's MEG trigger was then shifted by this Epoch-specific delay. The latency found was highly consistent across participants: ( $M=46 \mathrm{~ms}, S D=1 \mathrm{~ms}$ ) for the previous 60 participants and $(M=239 \mathrm{~ms}, S D=1 \mathrm{~ms}$ ) for the new 12 participants (who heard tones played using a modified sound card with a longer latency). For some previous participants, the tone was too weak to be detected: for these, a slower replacement algorithm was used which typically succeeded. Failing this, the median delay across all trials from the most recent participant was used given that the same equipment was generating the tones.
9. Raw data was downsampled to 200 Hz . This is thought to provide time buckets that were not too large, whilst speeding data processing.
10. RANSAC and Autoreject (Jas et al., 2017) were employed to repair or eliminate bad epochs and channels. RANSAC retains "inlier" sensors, removing those with implausibly low
correlation to "inlier" neighbouring channels. Autoreject removes bad trials based on a dynamic threshold detection, which can prevent outliers from distorting the mean response.
11. Independent (temporal) components analysis (ICA: Stone, 2002) was conducted on the entire raw data in a manner I automated; this detected 15 independent components and identified those unrelated to brain activity. These included phasic spatially segregated sources of physiological noise such as heartbeats and eye blinks (de Cheveigné \& Simon, 2007). Crosstrial phase statistics were used to identify and project out cardiac-related signals. As I lacked a dedicated ocular channel, for ocular components each channel was examined separately to check which components it matched using a (Pearson) correlation threshold of 0.8 , which was calibrated to detect approximately two components. The two components found most often across all channels were then removed. After these cardiac and ocular components were removed the fit was applied on Epoched data.
12. The data was again high-pass filtered to remove DC bias introduced by the ICA.
13. The aforementioned remapping of adult channels to the 125 -channel montage was performed.
14. For reasons explained in Appendix $I$, raw MEG data was standardised (i.e., z-transformed). Thereafter, all analyses were on standardised data only unless otherwise specified.

## Generation Of Predictors

Predictive surprise for each Epoch retained by the post-processing pipeline was calculated using several different models. In addition to the HGF models (Vanilla, and Pitch), these include three models for which code was available (Gijsen et al., 2021; see Appendix $A$ for code). The first is predictive surprise (PS), which is a Shannon-information measure derived from probabilities alone. The second is confidence-corrected surprise (CS) which in addition to probability, also considers confidence. The third is Bayesian Surprise (BS) with exponential forgetting; in Bayesian models, only observations which change beliefs create surprise, regardless of their probability of occurrence. This model was differentiated from the HGF in lacking the multiple layers of belief.

Unlike in the reviewed literature, on each participant I also fit HGF models trained only on subsets of the whole sequence (of lengths $346,692,1038,1384$ and 1730 tones). These should provide a purer estimate of the predictions that a Bayesian brain (with weak priors) trained on a continuous basis would make after accumulating different quantities of data than would training the model on the whole sequence (i.e., on data not yet seen by participants). These HGF parameter fits and goodness-of-fit metrics (e.g., BIC) are shown in Appendix E.

## Data Analysis Statistics

Subject-level Evoked responses were found by averaging the Epochs for the relevant condition (e.g., all trials expected to produce high surprise). Group-level Evoked responses (e.g., for different ages or conditions) were taken by computing the grand average over all participants.

## Condition Split And Regression

Trials were classified based on the magnitude of neural surprise likely to be evoked by the associated auditory tone - according to various models - either implicitly and categorically (for Deviants vs. Standards) or explicitly and continuously (e.g., in the HGF models). In the categorical analysis, separate low and high percentile cut-offs were used to classify trials as low surprise (condition 'A'), high surprise (condition 'B'), or neither. Various cut-offs were chosen which created contrasts of various strength; stronger contrasts came at the expense of reduced data. The conditions contrasted statistically are shown in Table 1 below: the cut-offs for the predictive models appear in the bottom row. For example, the "High vs. Low expected surprise" HGF contrast excludes trials with modelled surprise in the 20th-80th percentiles i.e., it compared the trials of the lowest and highest quintiles of expected surprise. These cut-offs were chosen to approximately match the proportions of trials in the traditional Deviant vs. Standard analysis. The continuous output of predictive models like the HGF also facilitated creating other partitions (e.g., contrasting "Very high vs. very low" expected surprise, the 90-100th vs 0 -10th percentiles respectively) as well as regression, analysed in Regression Of Brain Activity.

Table 1
The condition contrasts tested, and the proportion of all trials that fell into the low and high surprise categories

\left.| Condition Split | Percentage of all trials in each condition |  |
| :--- | :---: | :---: |
|  | Low surprise |  |$\right]$ High surprise

Notes: In the predictive models (bottom row), the percentages correspond to the cut-offs for expected surprise. In the control conditions, the low and high surprise conditions are arbitrary.

## Age Splits

Categorical analyses compared the youngest $20 \%$ of participants ( 15 people aged $<=5.1$ ) against the oldest $20 \%$ ( 15 people aged $>=22.1$ ) as well as the youngest $10 \%$ ( 7 people aged $<=4.3$ ) against the oldest $10 \%$ ( 7 people aged $>=34.7$ ). Contrasts were also made between the understudied group ( 8 people aged 10-16) and younger age groups. In my continuous analysis, age was also regressed against peak amplitude and its latency.

## Condition-by-age Interactions

Comparisons were made between the youngest and oldest $10 \%$ and $20 \%$ of participants' ERFs/GFPs for each condition separately. Differences found in the condition contrast between the age groups were noted.

## GFPs

For each participant, for each condition of interest, I calculated the Global Field Power (GFP; Lehmann \& Skrandies, 1980). I calculated this as the standard deviation of all measurement values, which provided an indicator of overall brain activity (especially synchronous brain activity) by compressing all spatial information (including for interpolated channels) down to a single timeseries. This compression means the GFP maintains statistical power which would be lost had I tested each sensor separately and corrected for multiple comparisons. As discussed by Files et al., (2016), the GFP also allowed me to test non-spatial hypotheses. This is not possible using ERFs, as ERFs do not sum to anything meaningful cognitively: by definition, in MEG, positive magnetic flux in one area of the brain offsets negative flux in another - at least when the head is spherical and aligned centrally within the helmet. Indeed, as all channels had been high-pass filtered, their mean (and sum) were $\sim 0$.

Rather than averaging the GFP per trial, I calculated the GFP on the ERF averaged over many trials, to enhance signals time-locked to the tones. The GFP for less-frequent conditions (e.g., deviants) is noisier than for more frequent measures (e.g., all non-deviants), which can cause bias (Files et al., 2016) but the rarest conditions still included $>12000$ trials at the group level.

The GFP also allowed me to select contrasts of interest which could then be temporally focused (Files et al., 2016). Effect sizes (Cohen's $d$; Cumming, 2013) were computed using the difference in GFP grouped by age or by condition and the pooled standard deviation of the GFP of each group.

## Regression On Brain Activity

I regressed age against the latency and magnitude of GFP at the time of peak GFP. Additionally, I considered whether the degree of HGF predicted surprise (the regressor) could be used to predict gross or localised brain activity (i.e., with the regressands being the GFP and the ERF). To do this, regressions were performed for each participant: I computed the (Pearson) correlation, the slope, and the $p$ value of slope. For the correlation, I calculated the Fisher $z$ value per participant, averaged these across participants, and then converted the average Fisher $z$ value into an average correlation. I applied a (conservative) Bonferroni correction for a non-directional statistical test, dividing the critical $p$ value ( 0.05 ) by 52 (because I examined the results of four different tests, each at the 13 timepoints of my $60 \mathrm{~ms}, 200 \mathrm{~Hz}$ sample). This produced a modified critical $p$ value of 0.00096 , which translated into a critical $z$ value of 3.3.

## Permutation Cluster Analyses

MEG data is subject to high spatial autocorrelation (see Appendix $F$ ), so the assumption of independence between observations is invalid and therefore Bonferroni (see Henry, 2015) or HolmBonferroni (Holm, 1979) adjustments for multiple comparisons would unnecessarily inflate the rate of false negatives. To determine the statistical significance of results, I used MNE's non-parametric testing suit, such as (two-tailed) spatiotemporal cluster tests (Maris \& Oostenveld, 2007) - which accounts for the spatial correlation of neural signals whilst correcting for multiple comparisons. Similar approaches to decoding neural surprise have been reported recently (e.g., by Modirshanechi et al., 2019 and Maheu et al., 2019). These methods also have the advantage of working on data which is not normally distributed. I did not select specific (brain) ROIs, but instead used nonparametric permutation tests over magnetometers to find significant clusters of difference globally (using GFP) or locally (for ERFs) and for evaluating the significance of statistics (such as the correlation between predictor and ERF) which were expected to cluster spatio-temporally. In such permutation tests, the null hypothesis is that data in one condition (e.g., "low surprise trials", or "older people"), and data in a second condition (e.g., "high surprise", or "younger people") were drawn from the same distribution. The largest cluster-level regions (temporal for GFPs or spatiotemporal for ERFs) where this was not the case were identified, using the spatial adjacency matrix extracted from the sensor locations in the child MEG system. The probability of these clusters occurring by chance was calculated using a subset of 10,000 permutations drawn from the possible $2^{72}$ permutations, with the $t$-threshold for the significance test chosen automatically to determine this at a $p=0.05$ significance level, where $p$ represents the proportion of random partitions that resulted in a larger test statistic than the observed one (Maris \& Oostenveld, 2007). As examples, the contrasts I performed using permutation clustering were as follows.

Firstly, I computed the difference in the GFP between conditions across all 72 participants. I will refer to this as the "GFP aMMR" (analogously to the ERF aMMR). Secondly, I computed the difference in the GFP between age groups (both by condition, and across all conditions) e.g., the youngest $20 \%$ ( 15 persons) minus the oldest $20 \%$ of participants ( 15 persons).

Thirdly, I computed the condition-by-age comparisons compared the difference in the GFP aMMR between age groups. Lastly, I computed the ERF comparison by subtracting the ERF in one condition (low surprise, or younger people respectively) from the other (high surprise, and older people respectively).

## Times Of Interest (TOIs)

The TOI was restricted to $130-190 \mathrm{~ms}$ (relative to tone onset) when running comparisons between low and high surprise conditions of ERFs, as this was the time during which prior research indicated that pitch aMMRs predominate (Bonetti et al., 2018). This allowed for greater statistical power in cases where ignoring times outside of this window appeared warranted. Otherwise, all analysis was performed in the $0-295 \mathrm{~ms}$ window.

## Results

## Pre-processing metrics

Per participant, the pre-processing pipeline identified and interpolated noisy channels ( $M=$ $0.56, S D=0.94)$, bad channels ( $M=0.15, S D=0.9$ ) and flat channels $(M=0.24, S D=1.03)$. Including these and channels manually labelled as containing high-frequency noise, on average 10.9 channels were marked as bad $(S D=6.24)$ - and were interpolated. Autoreject rejected a proportion of epochs ( $M=1.6 \%, S D=2.1 \%$ ). The automated ICA removed two components from every participant classified as reflecting cardiac activity and an average of 1.75 classified as reflecting oculomotor activity.

## Correlation Between Predictive Models

To determine which models were worthy of comparison by virtue of producing significantly different predictions - as models that are too well-correlated might be redundant - I calculated the (Pearson) correlations between models' predicted surprise (for PE2). As seen in Table 2, HGF_Vanilla was strongly correlated to HGF_Pitch, Predictive Surprise and surprise implied by the simple deviant/standard binary sequence (i.e., 1 or 0 ). It was however only moderately correlated to Bayesian surprise, and to Confidence-corrected surprise. This motivated comparing these models' predictive power with that of HGF_Vanilla.

Table 2
The correlations between HGF_Vanilla and other models' PE2/surprise for various sequence lengths.

| Model | Number of tones played |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\mathbf{3 4 6}$ | $\mathbf{6 9 2}$ | $\mathbf{1 0 3 8}$ | $\mathbf{1 3 8 4}$ | $\mathbf{1 7 3 0}$ |
| HGF Pitch | 0.95 | 0.94 | 0.93 | 0.94 | 0.94 |
| Predictive Surprise | 0.86 | 0.88 | 0.89 | 0.89 | 0.90 |
| Deviants vs. Standard (Binary) | 0.79 | 0.83 | 0.86 | 0.88 | 0.89 |
| Bayesian Surprise | 0.41 | 0.50 | 0.53 | 0.54 | 0.54 |
| Confidence-corrected Surprise | 0.34 | 0.42 | 0.49 | 0.49 | 0.49 |

Note: Similar relative correlations were evident between these models PWPE2, PE3 and PWPE3 (see Appendix C)
In deciding whether one model has better predictive power than another however, I also considered in which circumstances. Although the HGF_Pitch model is strongly correlated to HGF_Vanilla, recall that when the tone changes, HGF_Pitch outputs a PE2 that is smaller or larger than HGF_Vanilla in a manner which also depends on whether the acoustic frequency change was small or large. Importantly, this lowers the correlation of the two models significantly on deviant trials ( $\mathrm{to} \leq 0.48$, as seen in Table 3 compared to correlations $\geq 0.93$ over all trials as seen in Table 2):

Table 3
Correlation between PE2 in HGF_Vanilla and HGF_Pitch on deviants only for various sequence lengths.

| Length of sequence | Correlation |
| :---: | :---: |
| 346 | 0.48 |
| 1038 | 0.23 |
| 1730 | 0.18 |

Therefore, I conducted separate comparisons between the HGF models on high surprise trials. A similar argument justified comparing HGF_Vanilla with the simple deviant/binary sequence.

There was also a strong correlation between the expected surprise calculated by HGF models trained on different quantities of data ( $>=0.99$ on PE2 between all models trained on $>=346$ data points). Additionally, the correlation between predictions remained $\geq 0.99$ after changing the value of some HGF priors: the standard deviations of the logit of $\kappa$ (c.logitkasa) from 0 to 0.1 and of $\mu_{3}$ (c.mu3_0sa) from 0 to 0.05 in the Matlab file. These facts suggest that participants' model of the auditory environment - according to the HGF - should stabilise early in the experiment, regardless of the initial parameters. Consequently, I ran all HGF models with only one set of parameters (as seen in the Matlab file), trained on all data.

## Post-processing findings

Unless otherwise indicated, all plots show results for the HGF_Vanilla model, $\mathrm{t}=0$ refers to tone onset, and the low and high expected surprise trials (conditions "A" and "B") are defined as the 0-20th (lowest 20\%) and 80th-100th (highest 20\%) of trials ranked by modelled surprise, referred to as the " $\underline{\mathbf{2 0 / 2 0}} \mathbf{~ s p l i t " . ~ A r e a s ~ s h a d e d ~ i n ~ p i n k ~ a r e ~ s i g n i f i c a n t ~ a t ~ t h e ~} p<0.05$ level, using a two-tailed permutation cluster test. The $t$ and $d$ statistics reported here and in the Appendices refer to the average student's $t$ and Cohen's $d$ statistics respectively within the identified temporal/spatiotemporal clusters of interest.

## Gross Activity Patterns

Gross activity patterns are discussed in greater detail in Appendix F - Spatial autocorrelation. The key points are as follows:

1) As seen in Figure 9, two distinct peaks were seen in the GFP, at $\sim 110 \mathrm{~ms}$ and $\sim 240 \mathrm{~ms}$ :

Figure 9
GFPs for the grand average of trials of low surprise (condition A) and high surprise (condition B), and their difference



Note: $\quad$ Two distinct peaks are visible in the GFPs, with the contrast being maximal at 210 ms
2) As seen in Figure 9, these peaks were separated by a minimum (in GFP) at $\sim 170-180 \mathrm{~ms}$. As will be seen, at this point the polarity of the ERFs reverses in each brain hemisphere.
3) Anticipatory activity (see Discussion for further analysis of what is being anticipated) was evident around the time of tone onset, especially in adults.
4) GFP amplitudes reduced in the latter half of the Epoch before the anticipatory activity relating to the next stimulus, however different conditions (or, as shown later, age groups) were still distinguishable based on brain activity at 400 ms .

## Standardisation

The motivations for standardising (z-transforming) MEG data - and the impact of this on the results - are reported in greater detail in Appendix I - Standardisation. Generally, unstandardised data produced more temporal clusters of GFP difference between age groups. For example, using unstandardised data in HGF_Vanilla, there was a significant difference between age groups during throughout the period $55-295 \mathrm{~ms}, \mathrm{t}(28)=75.16, \mathrm{p}<0.001, \mathrm{~d}=3.07$; whereas in the standardised data there was no significant difference ( $\mathrm{p}>0.05$ ) between groups between $55-85 \mathrm{~ms}$ and between $150-215 \mathrm{~ms}$.

Conversely, between $0-55 \mathrm{~ms}$, there was a significant difference between the GFPs of the youngest and oldest using standardised data, $t(28)=-23.82, p<0.001, d=-1.75$, but no such effect using unstandardised data, $t(28)=1.17, p>0.05$. Using a $20 / 20$ split between conditions and age groups, during the TOI (between $155 \mathrm{~ms}-180 \mathrm{~ms}$ ) the contrast between the GFPs per condition was significant using standardised data only, $t(28)=6.44, p=0.023, d=0.42$. Following these analyses, comparisons all used standardised data unless otherwise indicated.

## Intra-model differences

Sanity checks. Choosing a condition boundary arbitrarily (based on a factor unlikely to influence expected surprise) should yield similar patterns of activity for both conditions. That is, two conditions delineated by an arbitrary boundary should evoke statistically indistinguishable GFP at any latency and indistinguishable ERF patterns (with indistinguishable correlations between the predictor and the ERF). I determined that this was in fact the case. The first contrast I calculated which serves as a control - contrasted tones with one frequency ( 600 Hz , condition ' A ') to another ( 700 Hz , condition 'B'). These tones are identical except in being 50 Hz below and above the average tone frequency respectively. Importantly, they were equally likely transitions from any other tone other than themselves, so I did not expect to observe any difference in the responses evoked. As seen in Appendix $J-600 \mathrm{~Hz}$ vs. 700 Hz , whilst not proving their equivalence, there was no evidence found of any difference between these conditions (all tests: $\mathrm{p}>0.05$ ). A comparison was also made between 650 Hz (the average physical tone frequency), and 800 Hz (which was the highest tone played). As seen in Appendix $J-650 \mathrm{~Hz}$ vs. 800 Hz - I again failed to find any evidence of a difference between these conditions (all tests: $p>0.05$ ).

## Conditions expected to be different.

i) Low Vs. High Surprise. As seen in Figure 10, a significant temporal cluster of GFP aMMR difference occurred from $155-180 \mathrm{~ms}, t(71)=6.44, p=0.024, d=0.36$; the high surprise condition showed greater activity.

Figure 10
The difference in GFP in the 20/20 split: high surprise trials (condition B) minus low surprise trials (condition A)



Note: Times of significant difference ( $p<0.05$ ) are shaded in pink. Note that confidence intervals cannot be directly displayed for permutation testing.

Examining the ERFs, as expected, the HGF models found spatiotemporal clusters containing significant differences between low and high surprise conditions both within and outside the TOI. For example, as seen in Figure 11, a significant spatiotemporal cluster of difference was found using HGF_Vanilla from $135-250 \mathrm{~ms}$ at 36 sensors around what is presumably left-hemisphere $\mathrm{A} 1, t(71)=-3.20, p=0.040$, and later around right-hemisphere A1 (at 34 sensors) from $170-280 \mathrm{~ms}, t(71)=4.22, p=0.041$.

Figure 11


Notes: Left: The mean ERF differences are displayed topographically.
Right: The mean ERF differences averaged over the spatiotemporal cluster amongst the significant sensors
ii) Very Low Vs. Very High Surprise. I also assessed whether greater modelled surprise was associated with greater neural activity. For example, the contrast between the $10 \%$ of tones expected to produce the least and most neural surprise (the " $10 / 10$ split") should be greater than this same contrast for the 20/20 split. However across models, instead I found that the 20/20 split showed a greater difference: in Figure 12 below, this manifested as the difference in the GFPs being statistically significantly for longer in the $20 / 20$ split ( $155-180 \mathrm{~ms}, t(71)=6.44, p=0.024$, $d=0.36$ ) than for the $10 / 10$ split ( $165-180 \mathrm{~ms}, p<0.05$ ). This did not however reflect a nominally greater difference between the GFPs throughout the $155-165 \mathrm{~ms}$ period: instead, it appears to be primarily due to reduced variance afforded by more data.

Figure 12
GFP differences between conditions during the TOI, with times of significant differences shown


Notes: Panel A shows trials with the lowest and highest $10 \%$ of expected surprise. There was a significant temporal cluster between 165 ms and 180 ms and a region approaching significance from $130-135 \mathrm{~ms}$.
Panel B shows trials with the lowest and highest $20 \%$ of expected surprise. There was a significant temporal cluster between 155 ms and 180 ms .
iii) PE2 Vs. PWPE2 And PE3. Unlike for PE2, the 20/20 split (analysed in Low Vs. High

Surprise) did not yield significant spatiotemporal clusters of ERF differences over trials with low vs. high PWPE2 ( $p>0.17$ ) or for low vs. high PE3 ( $p>0.51$ ). Further, as shown in Figure 13, the 20/20 split using PE2 created GFP aMMRs between conditions that were significantly different to zero $(t(71)=6.44, p=0.023, d=0.43)$, but the low vs. high PWPE2/PE3 splits did not $(p>0.05)$ :

Figure 13
The GFP aMMR in the 20/20 condition split for various HGF outputs, with times of significant difference shown


Notes: The GFP aMMR between low and high PE2 trials (blue), low and high PWPE2 (orange) and PE3 (grey). The GFP aMMR for PE2 was significantly different to zero in the temporal cluster shadowed with markers. The GFP aMMR for PWPE2 had one window approaching significance (thicker segment without markers). The GFP aMMR for PE3 had no significant spatiotemporal clusters.

As shown in Figure 14, visual inspection of the ERFs forming the basis of the GFPs for PE2, PWPE2 and PE3 provided further corroborating detail relating to the GFP differences:

Figure 14
aMMRs (for the 20\% most minus the 20\% least surprising trials) for PE2 (Panel A), PWPE2 (Panel B) and PE3 (Panel C)


Note: PWPE2 (and to an extent, PE3) appeared to delineate the conditions better at 295 ms than PE2.

For PE3 (Panel C) compared to PE2 (Panel A), between 195-200ms (marked as 197ms), the $20 \%$ of trials with the largest PE3 showed smaller-magnitude ERF differences to the $20 \%$ of trials with the lowest PE3 (significance not tested). Thus, PE3 appears to be less prognostic of neural activity.

For PWPE2 (Panel B) compared to PE2 (Panel A), between 195-200ms the aMMR ERF was again smaller in magnitude and more localised, likely explaining the lower (and not statistically significant) GFP reported in Figure 13 above at adjacent temporal regions. However, outside the TOI (at 295ms) the ERF for PWPE2 was of greater magnitude and less localised, suggesting that precisionweighting is applied later in the neural hierarchy. Indeed, as shown below in Figure 15, inspecting the GFP aMMR for PWPE2, I saw that it approached significance (shown in grey) at this longer latency, at $265-295 \mathrm{~ms}(p=0.071)$ :

Figure 15
The GFP aMMR for PWPE2 (20/20 split) and the associated t-statistic computed at each time-point


Repetition effects. The HGF models predict that successive repetitions cause decreasing neural surprise because repetitions increase both the probability of repetition - and the precision of this prediction. However, as seen in Figure 8, deviants after 4 repetitions are less than one third as likely as deviants after 5 repetitions. I reasoned that they therefore should - contrary to the predictions of the HGF - be more surprising. I indeed found that, across all participants, GFP was significantly different for deviants after five repetitions than for deviants after four repetitions in one cluster between $130-150 \mathrm{~ms}(t(71)=-5.34, p=0.038, d=-0.39)$, and the difference was nominally in the direction expected. As shown in Figure 16, the difference also approached significance between $165-175 \mathrm{~ms}(t(71)=-4.23, p=0.066, d=-0.35)$, and between $185-190 \mathrm{~ms}(t(71)=-4.81, p=0.07, d=-0.37)$ :

Figure 16
The difference in GFP for deviants occurring after five and four repetitions was significant between 130-150ms


Notes: In blue: The GFP aMMR of deviants after 5 repetitions minus deviants after 4 repetitions.
This contrast had temporal clusters of significant difference to zero (shown shadowed with markers).
The contrast also had regions approaching significance (shown as thick line segments).
By contrast, the difference in GFP between deviants which occurred after 5-7 repetitions and deviants which occurred after 1-4 repetitions was not significant ( $p>0.05$ ). This should however have been significant $i f$ the number of repetitions was the only thing the brain modelled, especially given that the number of trials in this contrast was much larger than in the contrast between four and five repetitions (and that this larger sample will have reduced variance).

## Regression Of Brain Activity

## GFPs.

i) PE2. As shown by Figure 17 below, for HGF_Vanilla, when PE2 was high (in the top $20 \%$ ), there was a significant positive (Pearson) correlation coefficient between the PE2 and the GFP throughout the period from $0-275 \mathrm{~ms}(p<0.05)$. By contrast, when the PE2 was low, the correlation between PE2 and the GFP was not significant ( $p>0.05$ ) at any times except for at 225 and 230 ms - were times which coincided with the second peak in the GFP.

Figure 17
The correlation of PE2 to GFP for low and high PE2 trials, during the TOI and over a wider window


Notes: In blue: On high surprise trials, PE2 was correlated to the GFP during most of the $0-300 \mathrm{~ms}$ period.
In red: On low surprise trials, the PE2 was rarely significantly correlated to GFP during the $0-300 \mathrm{~ms}$ period.
In green: The Bonferroni-adjusted confidence interval around the null of no correlation ([-0.021, 0.021]).
This was calculated from a sample of 24,545 and an unadjusted confidence interval of $[-0.013,0.013]$.

As seen, the correlation of the PE2 to the GFP was fairly stable for latencies up to 250 ms , but decreased sharply for all trial types thereafter, indicating that PE2 was not informative about the neural response at longer latencies.
ii) PE3 vs. PWPE2. As seen in Figure 18 below, for HGF_Vanilla, there was a significant positive correlation $(p<0.05)$ between the GFP and PWPE2 when PWPE2 was in the top $20 \%$ (or in the bottom $20 \%$, not shown), but not when PE3 was in the top $20 \%$ (or in the bottom $20 \%$, not shown):

Figure 18
The correlation of PWPE2 (blue) and PE3 (red) predictors to GFP, in cases when the PWPE2/PE3 were high


Notes: The Bonferroni-adjusted confidence interval for the (Pearson's) correlation is shown, [-0.021, 0.021]. The positive correlation coefficient is significant at all time points for PWPE2 but not for PE3 at any time point.

Using HGF_Vanilla, I also compared the strength of the correlation between the GFP and PE2 (on high PE2 trials) with the correlation between the GFP and PWPE2 (for high PWPE2 trials). To do this, Fisher's z-test for the difference between two correlation coefficients ( $r_{1}=$ 0.033 for PE2 and $r_{2}=0.072$ for PWPE2) for the 24,545 data points was performed. During the TOI, on trials when PWPE2 was high, PWPE2 was more (positively) correlated to the GFP than PE2 was to the GFP on trials when PE2 was high ( $z=4.32, p<0.001$ ). This suggests that high PWPE2 trials were better able (than PE2) to predict the GFP during the TOI. Furthermore, this same result was also obtained on trials where neural surprise was expected to be low.

As shown in Figure 19, these tended to occur in the same lateralised pattern and tended to show very similar patterns for correlation and slope:

## Figure 19

Topographic plot of the correlation (top row) and slope (bottom row) of the relationship between HGF_Pitch PE2 and ERFs


Note: This data was unstandardised data for HGF_Pitch for all trials over all participants.

These data appeared to indicate a significant correlation between PE2 as output by the HGF models and the neural responses recorded.

## Inter-model Comparison

In this section, I compare several models, identifying which is most predictive.
i) HGF_Vanilla vs. Standard/Deviants. In contrast to the HGF models (which, as described above, evidenced significant results in many contrasts), the GFPs for trials with standards and deviants were not statistically significantly different in any temporal cluster ( $p>0.05$ ). Moreover, the pattern of ERFs generated by the two trial types (in Figure 20 below, shown as aMMRs) appeared less distinctive than that seen for the low and high PE2 trials:

Figure 20
The aMMRs for HGF_Vanilla high minus low PE2 (top row) and for Deviants minus Standards (bottom row)


Closer examination revealed that unlike for the HGF, the comparison of deviants vs. standards did not yield any significant spatiotemporal clusters of ERF differences during the TOI ( $p>0.13$ ). Based on this, the generative HGF_Vanilla appears more predictive than the non-generative Deviant vs. Standard binary classification.
ii) PS vs. BS vs. CS. I compare three generative models (Predictive, Bayesian, and Confidence-corrected surprise). As shown in Figure 21 and Figure 22 respectively, when examining ERFs, only the confidence-corrected surprise model (but not Bayesian or Predictive surprise) evidenced the following:

1) aMMRs that were significantly different to zero (in one spatiotemporal cluster for the second peak that was shown in Figure 9):

Figure 21
The aMMR for all participants for the confidence-corrected surprise model


Note: 47 sensors in the left-hemisphere A1 were found to be significant at $225-295 \mathrm{~ms}$.
2) Differences between ages in the aMMR (using the $10 / 10$ and 20/20 splits), albeit at different time periods with ERFs of opposing polarity:

Figure 22
Significant spatiotemporally overlapping differences in the aMMRs between age groups using various age cut-offs


Notes: Panel A1: In the 20/20 split, a significant aMMR difference ( $65-120 \mathrm{~ms}$ ) at 30 sensors in left-hemisphere A1.
Panel A2: In the $20 / 20$ split, a significant aMMR difference ( $75-145 \mathrm{~ms}$ ) at 35 sensors in right-hemisphere A1.
Panel B: In the $10 / 10$ split, a significant aMMR difference (115-140ms) at 18 sensors in right-hemisphere A1.
I now compare this (confidence-corrected) generative model to HGF_Vanilla.
iii) HGF_Vanilla vs. CS. As seen in Figure 23, the GFP aMMR for HGF_Vanilla PE2 had a temporal cluster of significant difference at $155-180 \mathrm{~ms}, t(71)=6.44, p=0.023, d=0.43$.

Figure 23
The GFP aMMR for HGF_Vanilla (in blue) PE2 and for confidence-corrected surprise (in orange)


Note: The GFP aMMR was significant for HGF_Vanilla between 155 ms and 190 ms (shaded, with blue markers).

By contrast, no significant cluster was found for the confidence-corrected surprise model $(t(71)=$ $0.24, p>0.05)$. Using solely this criterion, the HGF_Vanilla was the 'best' model so far considered.
iv) HGF_Vanilla vs. HGF_Pitch. Using the expanded window, the GFP aMMR for HGF_Pitch PE2 was significant at latencies ranging between $260-295 \mathrm{~ms}, t(71)=12.00, p=0.017, d=0.58$. By contrast, as seen in Figure 24, HGF_Vanilla did not show any significant difference ( $p>0.12$ ):

Figure 24
GFP aMMRs for HGF_Pitch and HGF_Vanilla for the 20/20 split (trials with the highest vs. lowest 20\% of surprise)


Notes: In blue: HGF_Pitch had a statistically significant GFP aMMR at 260-295ms.
In orange: HGF_Vanilla did not approach significance over this larger time window.

The significant difference in GFP aMMRs for HGF_Pitch occurs at 260-295ms, which appears not to overlap with the periods seen in Figure 9. As will be discussed, this difference appears to relate not to the raw pitch characteristic (processed early in the neural hierarchy at the shortest latencies) or to updates of tone transition probabilities (which appear to occur at $200-240 \mathrm{~ms}$ ). Instead, this period of significance in HGF_Pitch likely reflects a difference in activity relating to a model's volatility update which should occur after a significant pitch change. This is consistent with the greater activity for PWPE2 at longer latency as discussed in Figure 25 below. HGF_Pitch (Panel A) appears to define conditions producing a more differentiated response than for HGF_Vanilla (Panel B). As shown in Figure 25, the greater contrast between the models can be seen both by noting the larger spatial extend of the aMMR at $98 \mathrm{~ms}, 150 \mathrm{~ms}$, and 170 ms latencies and the significantly different pattern at 295 ms as shown in the second row of each panel:

## Figure 25

The aMMR (ERF) for HGF_Pitch (Panel A) and HGF_Vanilla (Panel B) at various latencies for the 20/20 split

## A



B


Note: The aMMR for HGF_Pitch appears greater at short latencies as well as later at 295 ms (with reversed polarity).

HGF_Pitch Highlights Age Differences. A much broader discussion of how surprise varies with age occurs in the Age Comparison section below. One difference between HGF_Vanilla and HGF_Pitch with respect to age was that when examining the GFP aMMR between age groups, HGF_Pitch again was relatively more predictive than HGF Vanilla of activity later in the Epoch. Specifically, as seen in Figure 26, the GFP aMMR between age groups was significant from 195-260ms $(t(28)=7.69, p=0.0053, d=-1.02)$ for HGF_Vanilla but HGF_Pitch was significant between 225-295 ms $(t(28)=11.36, p<0.001, d=1.25)$ :

Figure 26

Contrasting the GFP aMMRs for older minus younger age groups for HGF_Pitch and HGF_Vanilla


Note: The areas shadowed with markers correspond to significant temporal clusters (at the $p<0.05$ level). The thicker orange segment without markers at the earlier period of $160-190 \mathrm{~ms}$ indicates an area which also approached significance $(t(28)=7.26, p=0.056, d=-1.01)$.

The presence of a period of significance was also more robust to the age cut-off used for HGF_Pitch than it was for HGF_Vanilla.

The interaction between model and age group could again be discerned from the pattern of ERFs. Notably, as seen in Figure 27, HGF_Pitch showed larger aMMR magnitudes and relatively more activity late in the Epoch:

Figure 27
HGF_Pitch (top row) showed a bigger difference between age groups in aMMRs than HGF_Vanilla (bottom row)


Note: The data displayed uses unstandardised data.

As seen in Figure 28, during the TOI, spatiotemporal cluster analysis confirmed statistically that the difference between the youngest and oldest $10 \%$ persons' aMMRs was significant at 50 fronto-central sensors at $125-185 \mathrm{~ms}$ in the HGF_Pitch model $(t(12)=4.97, p=0.011)$ and not significantly different for HGF_Vanilla ( $p>0.55$ ):

Figure 28
The difference in aMMR (oldest 10\% of persons minus youngest $10 \%$ of person) in HGF_Pitch


Note: This difference was significant at 50 sensors between 125 and 185 ms

## Age Comparison

a) Gross differences (GFP). Significant differences in gross brain activity were found between younger and older participants irrespective of the condition (i.e., even on control conditions). An example is shown in Figure 29 below comparing persons aged 3-5 with those aged 10-20 - but such differences occurred regardless of the exact age cut-off chosen:

Figure 29
Comparison of the GFPs between two age groups (left), with the differences in GFP shown at right.



The basic features consistently detailed in the sections that follow are as follows. Firstly, both age groups have two distinct peaks, which are more prominent in younger people. These peaks typically reach significance from $\sim 90$ to $\sim 150 \mathrm{~ms}$, and $\sim 220 \mathrm{~ms}$ to $\sim 270 \mathrm{~ms}$. Secondly, older groups have peaks of smaller magnitude than for younger groups. The effect size of the additional activity in the younger group (here, children aged 3-5 years) relative to the older group (here, 10-20-year-olds) is typically greater at the first peak. This first peak likely relates to physical stimulus characteristics and not transition probability updates, which are reflected neurally at longer latencies (Maheu et al., 2019). Thirdly, in contrast to at other times in the Epoch, the older group typically shows larger GFP amplitude at times before and proximal to the tone onset (here seen from -100 to $\sim 60 \mathrm{~ms}$ ); as discussed, this is likely anticipatory activity. Lastly, as shown below, the latency was unexpectedly longer for younger children for both peaks.

The Peaks In Young And Older People. Using all trials with HGF_Vanilla, there were condition-independent differences in GFP between younger and older age groups that were statistically significant during much of the Epoch. For example, comparison between the oldest and youngest $20 \%$ of participants was significant from 5-55ms $(t(28)=-23.80, p<0.001, d=$ -1.75 ) during which older participants had nominally greater GFP, presumably representing increased anticipatory (top-down) activity. It was also significant from $85-150 \mathrm{~ms}(t(28)=46.46$, $p<0.001, d=2.51$ ) and from 215-295ms $t(28)=27.8, p<0.001, d=1.96$ ), during which young persons had greater GFP. As seen in Figure 30 below, the results were significant in a manner insensitive to the exact age cut-off used, with the $10 / 10$ split also being significant from $5-60 \mathrm{~ms}(t(12)=-15.77, p=0.005, d=-1.97)$, from $85-150 \mathrm{~ms}(t(12)=38.56, p<0.001, d=$ 3.16) and from $230-295 \mathrm{~ms} t(12)=14.56, p=0.003, d=2.09)$.

Figure 30

The difference in GFP (younger minus older) over all trials, for the 10\%/10 (blue) or 20/20\% (orange) splits


Notes: Top: The difference in GFP between ages, for the 10/10 split (blue) and 20/20 split (orange).
Times of significance are shown with markers.
Bottom: $t$ values associated with the GFP differences.
The 20/20 split displays greater $t$ values at longer latency, owing to reduced variance.

A Closer Look At Latency And Amplitude. Figure 31 shows a closer look at the progression with age of aMMR amplitude and latency across quintiles of $\sim 15$ persons:

Figure 31
GFPs by condition (left, and centre) and GFP aMMR (right) per age quintile, calculated using HGF_Vanilla (20/20 split) for two different time windows (top and bottom)


Condition B (high surprise)



Condition B minus condition A



The First Peak. The overall pattern of results shown in Figure 32 below suggested that the latency of the first GFP peak shifts earlier with age. Over all trials, there was a significant negative correlation between age and the latency of the first peak, $r(70)=-0.64, p<0.001$. The correlation was not significant for low surprise trials $(r(71)=-0.15, p=0.19)$ but was significant for high surprise trials $(r(70)=-0.49, p<0.001)$. The shift was likely monotonic across the youngest four quintiles. Welch's $t$-test across all trials showed that the latency was different between the youngest quintile ( $M=122 \mathrm{~ms}, S D=8 \mathrm{~ms}$ ) and the second-youngest quintile $(M=111 \mathrm{~ms}, S D=11 \mathrm{~ms}), t(44.98)=4.08, p<0.001, d=1.12)$. Similarly, latency differed between the second and third youngest $(M=91 \mathrm{~ms}, S D=26 \mathrm{~ms})$ quintiles $(t(46.80)=$ 3.99, $p<0.001, d=0.95$ ) and between the third and fourth youngest ( $M=72 \mathrm{~ms}, S D=38 \mathrm{~ms}$ ) quintiles $(t(46.27)=2.10, p=0.041, d=0.55)$. Latency was not significantly different between the fourth and fifth youngest quintiles, $t(51.07)=-0.67, p=0.51)$.

Figure 32
The lines of best fit between age and i) latency (left) ii) max GFP amplitude (right) for the first peak of GFP


Notes: The data is obtained from the HGF Vanilla model. The GFP and latency were only examined in the range of $40-180 \mathrm{~ms}$ for the analysis of the first peak. Reducing the minimum latency would however have made the correlation more negative.

There was also a significant negative correlation between age and the maximum amplitude of GFP, $r(71)=-0.36, p=0.002$. However, as was seen previously in Figure 31, the GFP did not however decline monotonically with age: those aged 5.1-6.8 had greater GFP (if anything) and GFP aMMRs than those aged 3.2-5.1 at the first peak.

The Second Peak. The second peak appeared more ambiguous for the oldest two quintiles and was possibly shifted earlier in high surprise trials in these groups. Analysis of participants in the first three quintiles of age, who showed an unambiguous second peak, suggested that the latency of the second GFP peak also decreases with age; this is shown in Figure 33 below. Over all trials, there was a significant negative correlation between age and the latency of the second peak, $r(71)=-0.32, p=0.006$. The correlation was significant both for low surprise trials $(r(71)=-0.36, p$ $=0.007)$ and for high surprise trials $(r(71)=-0.40, p<0.001)$. The group averages for latencies were $256,251,249,241$ and 215 ms respectively across the age quintiles but the high variances across the quintiles (ranging from 13-40ms) meant the latency from quintile to quintile for the first four quintiles was not significantly different ( $p=0.48,0.72,0.42$ ). The latency was different between the fourth and fifth quintiles was significantly different to zero, $t(51.45)=2.75, p=0.008$, $d=0.74$

Figure 33
The lines of best fit between age and i) latency (left) ii) max GFP amplitude (right) for the second peak of GFP

and latency were only examined in the range of $180-300 \mathrm{~ms}$ for the analysis of the second peak.

There was also a significant negative correlation between age and the maximum amplitude of GFP, $r(71)=-0.55, p<0.001$. As was seen with the first peak, the GFP did not however decline monotonically with age: those aged 5.1-6.8 had greater GFP (if anything) and GFP aMMRs than those aged 3.2-5.1 at the second peak.

GFP Differences At The Second Peak. Differences in the GFP aMMR by age were also found to be significant or approach significance (using a variety of models) - but on the second peak only. For example, comparing those aged 3-5 to those aged 10-16 (in Figure 34, bottom), HGF_Pitch was significant from $260-295 \mathrm{~ms}(t(21)=6.59, p=0.008, d=1.21)$ and nowhere else $(p>0.05)$. This significance was robust to the age cut-off used; the $10 / 10$ split was significant from $255-285 \mathrm{~ms}$, $t(12)=18.7, p=0.046, d=3.8$, and the $20 / 20$ split from $250-295 \mathrm{~ms}, t(28)=15.3, p<0.001, d=2.3$.

Figure 34
The GFP aMMR between different age group contrasts for HGF_Vanilla (top) and for HGF_Pitch (bottom)


Notes: Areas with significant differences have markers shown.
Top: HGF_Vanilla GFP aMMR.
Bottom: HGF_Pitch GFP aMMR differences between age groups.
Unlike the GFP differences, these only occurred at longer latencies.
b) ERF Differences. The GFP differences mentioned must be driven by ERF differences. This appeared to be the case when conducting linear and quadratic regressions of age against the (most negative) amplitude of the aMMR during the TOI, shown in Figure 35. The positive linear and negative quadratic components did not however reach statistical significance ( $p>0.05$ ).

Figure 35

A positive but not significant linear coefficient and a negative non-significant quadratic coefficient was found between peak amplitude and age


Note: Two outliers in age (the oldest individuals) are not plotted, which did not influence the shape of the plot.

More sensitive measures do however detect the ERF differences between age groups. The distinct peaks in the difference of GFPs between age groups shown previously are associated with significant spatiotemporal clusters of ERF difference ( $p<0.05$ ). For example, on all trials, the GFP aMMR differences shown in Figure 34 appear to derive from the ERF differences, shown here in Figure 36:

Figure 36
Regions of significant spatiotemporal differences in ERF (over all trials) between age groups


Notes: The comparison made is the oldest $20 \%$ minus the youngest $20 \%$.
Left: early region of significant spatiotemporal differences at 110 sensors, between $5-165 \mathrm{~ms}$.
Right: late region of significant spatiotemporal differences at 105 sensors, between $160-295 \mathrm{~ms}$. In this later period, each cluster had opposite polarity.

Examining the difference between aMMRs in Figure 37 below, there was also statistically significant difference between older and younger people in the left hemisphere:

Figure 37
Regions of significant spatiotemporal differences in the aMMR (over all trials) between age groups


Notes: The comparison made is the oldest $20 \%$ minus the youngest $20 \%$. The difference is significant at 50 sensors around left-hemisphere A1 (and towards the mid-brain) from 180-295ms.

As seen in Figure 38 below, for HGF_Pitch, a similar spatiotemporal cluster was found in the $10 / 10$ split (at $125-185 \mathrm{~ms}, t(12)=4.97, p=0.011$, Panel A1) as well as with opposite lateralisation later in the period ( $200-295 \mathrm{~ms}, t(12)=4.46, p<0.001$, Panel A2). A similar significant difference in the $20 / 20$ split was found during this later period too (at $190-295 \mathrm{~ms}$, $t(28)=4.48, p=0.001$, visible in Panel B):

Figure 38
The aMMR difference of the oldest minus the youngest 10\% (Panels A1/A2) and 20\% (Panel B), using HGF_Pitch


Notes: Panels A1/A2: The differences in aMMR (oldest $10 \%$ minus youngest $10 \%$ ) were significant at 50 channels. Panel B: The difference in aMMR (oldest $20 \%$ minus youngest $20 \%$ ) was significant at 52 channels.

One source of these ERF differences between age groups is the polarity of the ERF varying between ages, as seen in Figure 39:

Figure 39
ERFS for the youngest (top row) and oldest (bottom row) 20\% of people for HGF_Pitch


Note: Unstandardised data. The youngest and oldest displayed opposite polarity at 130-150ms, which drove the early aMMR differences in Figure 38 .

Repetition Effects, Again. In Repetition effects, I saw that more gross activity was evoked by a deviant after four than to after five repetitions. Spatiotemporal cluster analysis, shown in Figure 40 below, also determined that the difference in the aMMR (ERF) signal between the oldest and youngest $10 \%$ of people was significant in two clusters: one with a positive ERF between $150-185 \mathrm{~ms}(t(12)=4.62, p=0.029)$ at (apparently) lefthemisphere fronto-central sensors and one slightly later with a negative ERF between $165-175 \mathrm{~ms}(t(12)=5.38, p=0.049)$ at regions in the righthemisphere:

Figure 40
Spatiotemporal clusters of significant ERF differences between the oldest and youngest (10\% of) participants


Notes: A cluster of 20 sensors of significant difference was fond at $150-185 \mathrm{~ms}$ (left, showing left-hemisphere) and of 24 sensors at 165 - 175 ms (right, for right hemisphere). The graph displays data for deviants after five repetitions minus deviants after four repetitions.

The GFP also differed between age groups ( $p<0.001$ ) and, unlike for all other condition contrasts, was nominally larger in younger people than in older people at all latencies. One interpretation of such findings is that younger and older individuals differ in their ability to track the number of repetitions - with younger individuals showing an advantage - which, as will be discussed, could result from working memory differences.

Greater anticipatory activity. As was seen in Figures 30, 31 and 34, older groups had reduced GFP relative to younger groups during most times, but at the first peak appeared to have more anticipatory activity. This was also seen in the pattern of ERFs (statistical significance not tested) shown in Figure 41:

Figure 41
The oldest quintile (bottom row) displayed event-related activity at the time of tone onset, unlike the youngest quintile (top)


Notes: This data is for the high surprise condition; a similar pattern occurs for all other trial types. The plot uses unstandardised data.

## Condition-by-age Interactions

A simple test for a condition-by-age interaction is to see if the difference in GFP for each condition differs temporally as a function of age. For example, as seen in Figure 42, when comparing low to high expected surprise trials for HGF_Pitch, no significant difference in GFP was observed between the youngest $10 \%$ (seven) and oldest $10 \%$ (seven) participants:

## Figure 42

The GFP difference (high surprise minus low surprise) for the youngest $10 \%$ minus the oldest $10 \%$ of participants


Note: The grey regions represented times with a GFP difference approaching statistical significance ( $p<0.3$ ).
This does not reject the hypothesis that a surprising tone generates the same amount of additional (or decreased) brain activity at each time-point in younger and older individuals. However, it may be that in this case, contrasting the GFP is not sufficiently powerful. Firstly, the GFP can be identical even with large differences in the underlying brain activity as the GFP calculation is indifferent to where activity occurs (Murray et al., 2008). Secondly, the GFP combines information from brain regions containing information not specific to auditory stimuli (e.g., in visual and olfactory regions) but which nonetheless contribute to variance. A potentially more sensitive measure for the difference between age groups is the difference in aMMRs. This indeed does reveal significant differences, as seen in Figure 43:

Figure 43
Spatiotemporal cluster of significant differences in ERF between the oldest and youngest $10 \%$ of participants


Note: This data is for HGF_Pitch. Differences were found at 50 sensors throughout the TOI (from 125-185ms).

Additionally, as seen in Figure 44, widening the TOI revealed significant differences with age in the GFP aMMR:

Figure 44
Differences in the GFP aMMR between different age groupings for HGF_Pitch


Notes: The comparisons are between the oldest and youngest $10 \%$ (in blue), and $20 \%$ (in orange). The $10 / 10$ split was significant from $255-290 \mathrm{~ms}(t(12)=8.52, p=0.05, d=1.58)$, shown with markers. The $20 / 20$ split was significant from $225-295 \mathrm{~ms}(t(28)=5.78, p=0.05, d=1.26)$, shown with markers.

Are The Differences Artefactual? It was considered whether the age-related affects described thus far might arise due to differences between the child and adult MEG systems. I believe this is unlikely for four reasons:

1. Spatial autocorrelation in the child system is of smaller magnitude than in the adult system (see Appendix $F$ ): this may reduce the odds of a spatiotemporal cluster forming.
2. The adult system has a greater concentration of sensors near the right inferior frontal gyrus (where aMMRs occur; see Cheng et al., 2013), but adult GFP amplitudes were reduced.
3. There seem to be age-related differences within systems e.g., the finding of less activity in older persons was found for the adult MEG system, as seen in Figure 45:

Figure 45
Differences between the youngest and oldest participants scanned within the same (adult) MEG system


Notes: The top row represents the youngest $50 \%$ of participants scanned in the adult system. The bottom row represents the oldest $50 \%$ of participants scanned in the adult system.
Differences in the ERFs between younger and older persons for subjects scanned within the adult MEG system implied that there were sources of age-related differences which were not related to the recording system used.

## Discussion

My research had two broad goals. Firstly, I aimed to compare models which predict neural surprise by evaluating how well they represented neural activity as recorded by MEG. The second was to analyse age-related differences in neural surprise (especially in younger teenagers compared to other studied groups), characterising how younger and older participants differed in neural surprise (e.g., in the amplitude and latency of GFP and ERF).

## Summary of results

Whilst not conclusively proving a lack of difference, the control contrasts provide us with more confidence that the differences found are not artefactual. The contrast between 650 Hz and 800 Hz tones being non-significant argues against the idea that the brain is modelling the physical pitch of tones and being surprised solely by extreme (very low or very high) pitches.

Several models - including the Confidence-corrected surprise, HGF_Vanilla and HGF_Pitch models - successfully characterised aspects of neural surprise based on modelled surprise. This was determined by analysing the correlation of modelled surprise to GFP, by identifying temporal clusters of difference in the GFPs for trials of low vs. high modelled surprise, and by contrasting the ERFs associated with these condition types spatiotemporally. These comparisons provided complementary, convergent evidence of a different brain response to trials expected to produce low surprise compared with trials expected to produce high surprise.

The HGF_Vanilla model - compared to the simpler (non-generative) Deviant vs. Standard binary classifier - additionally modelled estimated beliefs about tone transitions (and their uncertainty), tone transition volatility (and its uncertainty) and environmental uncertainty. For reasons likely related to these additional features, it was more predictive of GFPs and ERFs than the binary classifier and was also more predictive than the confidence-corrected surprise model.

Focusing on HGF_Vanilla, its outputs of the tone transition probability error (PE2) and precision-weighted prediction error (PWPE2) of tone transition probability were found to correlate with the GFP, whilst its PE3 output did not; PE3 also appeared to be less predictive of GFP aMMRs during the TOI than PE2 was. The comparison between PE2 and PWPE2 was ambiguous: PWPE2 had a higher correlation to the GFP than PE2, however the difference in GFP aMMR between low and high PWPE2 trials was less clear than it was between low and high PE2 trials. The clusters of statistically significant difference between low and high surprise trials in the HGF models indicated that neural activity differed based on modelled surprise at
what is (assumed to be) bilateral A1 at two latencies: an early response representing the brain's processing of physical sound characteristics and a later response representing the brain's model update. This later peak was better detected by PWPE2 and PE3 than by the PE2.

The novel modifications to the HGF to incorporate physical tone frequency improved upon the HGF's ability to predict the degree of neural surprise, with this relative advantage occurring at latencies longer than traditionally identified as part of the MMN signal and longer than the two peaks in GFP identified: thus, it may relate to the update of volatility estimates.

Robust differences were found between younger and older people. In older participants, smaller-magnitude ERFs and GFPs were found that were independent of the condition classifications created (i.e., on all trial types). Condition-dependent decreases in the magnitude and latency of GFP peaks were also found with age via regression, as were smaller GFP aMMR amplitudes in older compared to the younger people (especially to conditions of high surprise) at longer latencies. aMMR differences were partly driven by a difference in the polarity of the ERF with age at the early GFP peak, with younger participants showing a negative (positive) field value in the left (right) hemisphere and older individuals showing the opposite (see Figure 39). Older persons' brains also displayed more anticipatory activity temporally locked to sound onset and apparently less ability to track the number of repetitions.

Overall, my results are consistent with relatively greater top-down activity in older persons and bottom-up activity in younger persons. Some effects were however found which do not disprove but which may challenge Bayesian brain theories. Specifically, deviants after four repetitions were more surprising than deviants after five repetitions - and the progression of GFP amplitude with age was not monotonic. These results may imply that transitions are modelled over more complex features than just tone transition probability and/or that factors such as shortterm memory - which are known to vary with age - may impact the pattern of evoked responses.

## Interpreting the results

## Inter-model

The HGF_Vanilla model made better predictions than the binary output of the Deviants vs. Standards binary classification. For example, only the HGF classified trials as low or high surprise in a manner which produced statistically significant spatiotemporal clusters of difference between the ERFs for these conditions. This suggests that information relevant to surprise is relatively better captured in the HGF models - such as the number of recent repetitions and/or the transition volatility - and that these factors contribute to neural activity.

The novel modifications made to HGF_Vanilla to incorporate physical tone frequency information was reflected in the greater ability of HGF_Pitch than HGF_Vanilla to discriminate between low and high surprise trials and between younger and older persons. This was seen in the GFP aMMRs and in the difference in GFP aMMRs between age groups, both of which were significantly different to zero for HGF_Pitch (but not for HGF_Vanilla) at longer latencies. This may be interpreted several ways. This could occur because the actual neural response reflects the addition of multiple processes (e.g., because neural activity serves both to update tone transition probabilities and reflects surprise at the magnitude of the tone change, which is only captured in the HGF_Pitch model). Alternatively - or perhaps additionally - it could be because the surprise modelled by HGF_Pitch incorporates information about the update of a different quantity - e.g., estimated volatility - in a way that is somehow better than PWPE2 and which is ultimately reflected in neural activity (perhaps at longer latencies).

For the intra-model and age-related differences discussed below, these were explored through the HGF_Vanilla model to aid comparison to the results found by others.

## Intra-model

The impact of precision-weighting was ambiguous. As seen in the Confidence-corrected model, the Bayesian surprise model - based on Kullback-Leibler (KL) divergence (Kullback \& Leibler, 1951) - was improved by additionally considering the confidence of the generative model (by scaling it by the negative entropy of the prior distribution). This suggests that confidence influences brain activity. Within HGF_Vanilla however, precision weighting had a mixed impact. On the one hand, PWPE2 was more strongly correlated to the GFP when PWPE2 was high than PE2 was to GFP when PE2 was high. On the other hand, in Figure 13 it was seen that precision-weighting the PE2 reduced the discriminability of low and high surprise trials using the GFP aMMR - at least during the TOI. It may not be warranted, however, to conclude
that the brain does not form estimates of precision: for example, it is possible that precision weighting is applied not over transition probability, but instead in some other space (e.g., in physical frequency space, and/or in number-of-repetitions space etc.) which simply is not detected using GFP aMMRs. Alternatively, precision estimates may be neurally instantiated only at a longer latency. Indeed, the pattern of ERFs and GFPs appears to be better separated by low vs. high PWPE2 values (than by low vs. high PE2 values) later in the Epoch (see Figure 14). Thus, the overall pattern of results for PWPE2 suggests that precision weighting might modify bottom-up signals not at the lowest (earliest) levels of the neural hierarchy but instead may do so only further up and at a longer latency. It remains to be seen how this occurs, though some (e.g., Friston, 2010) have considered that precision-weighting is implemented anatomically by greater density (weighting of) lateral and intrinsic connections between neural units and neurochemically via neuromodulators like dopamine and acetylcholine.

As detailed in Appendix C, for HGF_Vanilla, unexpected tone transition volatility (PE3) had a moderate to strong correlation to PE2, depending on the trial type. However PE3 appears to be an inferior differentiator of gross brain activity (GFP) relative to PE2, as seen in the lower correlation between high PE3 and GFP shown in Figure 17 (relative to that between PE2 and GFP, shown in Figure 18). This may also be seen in the pattern of GFP aMMRs in Figure 13, and may relate to the reduced predictive power of the PWPE2 seen in the same figure, given that in the HGF models it is PWPE2 (and not PE2) that flows up to higher layers, affecting the estimate of volatility.

## Repetition effects

The larger neural surprise after four repetitions than five repetitions - as quantified by the GFP aMMRs in Figure 16 - reflected the reality that there was a high probability of a further repetition following four repeated tones but an appreciably smaller chance of a repetition after five repeats. This result however conflicts with the predictions of the HGF, in which surprise (to a deviant) should be greater when the number of repetitions preceding the deviant increases.

Interestingly, the comparison of the GFP between deviants which occurred after 5-7 repetitions and of the GFP for deviants occurring after 1-4 repetitions was not significant ( $p>$ $0.05)$. This should however have been significant $i f$ the number of repetitions was the only thing the brain modelled, especially given that the number of trials in this contrast was much larger than in the contrast between four and five repetitions (and that this larger sample will have reduced the variance). The dissociation lends some support to the idea that the brain learnt about the specific discontinuity in the probability of a tone change after four repetitions.

These findings - alongside the higher predictive power for HGF_Vanilla compared to the Deviants vs. Standards binary classification - imply that information about the number of repetitions and information about tone transition probability as modelled by the HGF compete in combination, with both influencing neural surprise in opposite directions. This is especially likely when you further consider that after five repetitions, deviants become more surprising after each repetition according to the HGF, but should produce less and less surprise given that tones never repeat more than seven times.

## Age

With respect to age, the results were only partly as expected. One set of results which had been anticipated was that GFP, ERF, and aMMR (ERF) amplitudes declined with age, and that this occurred at both of the two peaks in GFP identified. These reduced amplitudes indicated that older participants' brains engaged less in both the processing of physical pitch characteristics (likely represented by the short-latency peak) and in model updating (likely represented by the long-latency peak) respectively along the neural hierarchy. An unexpected result was that the decline in GFP amplitudes with age appears not to be monotonic. If neural surprise is largest not at birth but during early childhood, this leaves open the possibility that Bayesian accounts are wrong in the sense that any additional sensory data accumulated between birth and early childhood does not lead to slower belief updating. However, the results may still be consistent with Bayesian belief updating. For example, the precision of sensory observations may increase throughout early childhood with language training, leading to greater neural surprise.

With regards to the decline in GFP amplitude with age at both peaks evidenced by the significant negative slope on the linear regression and the decreasing amplitude of the GFP aMMR - these may indeed index reduced learning. As discussed in the Introduction, such results may be mediated by inferior sensory memory, as revealed by performance differences between models of learning that are windowed or 'leaky' (Gerstner \& Kistler, 2002) and models which are memory-less (Meyniel et al., 2016). Similarly, contrary to Bayesian models, I expect that human learning does not proceed in a memory-less fashion (in which a model's beliefs always fully reflect the whole history at any given snapshot of time). Instead, as some have previously modelled (e.g., see Kolossa et al., 2012), our beliefs and predictions likely vary idiosyncratically as evidence accumulates in - or falls out of - memory. Indeed, metrics such as reaction time have been shown to be dependent upon multiple memory traces (Los et al., 2014). It may be that older brains encode less information in sensory memory in the first place, or encode it more slowly or less precisely. Alternatively (or in addition), adults may retain less information over time, or
discard or 'down-sample' (in precision terms) old information when the quantity of data accumulated in memory structures increases (especially when near capacity). For example, older memories can become 'decontextualised' (Yassa \& Reagh, 2013) i.e., stored with less precision over time unless they are re-activated. Importantly, in my experiment, any of these mechanisms would have reduced the ability of older individuals' brains to model the 'hazard function' (see Tavano, 2019) which describes the probability of a transition to any deviant (not to a specific deviant). This would result in greater irreducible uncertainty on each trial, smaller PWPE2 magnitude and reduced belief updating. Declining sensory memory should also have flow-on consequences for surprise captured by HGF's predictions. For example, if sensory memory limitations make older individuals' brains less likely to automatically draw the inference that a deviant after the maximum number of repetitions is certain, they should form an erroneous belief that the environment is less predictable (i.e., more volatile) than it really is; this should then have induced greater erroneous surprise when the environment exhibits periods of stability. I did not statistically test whether this occurred, but it must be said that this appears unlikely given the main pattern was of reduced surprise in older subjects at all latencies.

I also saw that older participants had greater GFP than younger participants before the first peak, from 100 ms prior to the tone to $\sim 50 \mathrm{~ms}$ after it. It is possible that this reflects relatively more top-down anticipatory activity in relation to tone pitch. This appears unlikely though, as greater anticipation of pitch should also subsequently have caused greater GFP amplitude in older persons at longer latencies in response to surprise about the tone's pitch. Instead, it seems more likely that the anticipatory activity was related to another characteristic of the tones which would be reflected only at shorter latencies with high accuracy, such as the tone's timing via greater entrainment to beat (Thompson et al., 2015); predictive coding accounts of more complex musical features are described elsewhere (e.g., see Vuust et al., 2018). Greater entrainment to beat could be seen as an optimal adaptation with age, since the pitch varies randomly and unpredictably and thus attempting to predict the pitch is fruitless. That is, older brains may be better equipped to recognise the futility of attempting to predict tone transitions. A Bayesian framing of this would be that the impossibility of predicting random tone transitions in a roving oddball paradigm leads to the earlier activation of a reduced precision context when older brains recognise the random nature of the tones and disengage. Within the framework of the HGF, the entire context being modelled with very low precision leads to low PWPE2 outputs with flow-on to estimates of PE3. It is also possible that changes with age occur along more dimensions than simply aMMR amplitude and latency e.g., there may be a dissociation between neural surprise
and belief updating which depends on age, but which was not observable in the present experiment.

Latency was unexpectedly longer for younger participants for the first peak, which should relate to raw physical stimulus characteristics rather than to transition probability updates (Maheu et al., 2019). This conflicted with other findings reported (e.g., Bertoli et al., 2002; Schiff et al., 2008; Cooper et al., 2006). This is unlikely to be due to 'spill-over' from the aforementioned increased anticipatory activity near tone onset, since the latency also declined with age at the second peak. This second peak is more likely to represent a model update and thus the difference cannot reasonably be attributed to entrainment to beat. Why this should be so remains to be explained.

## Strengths And Weaknesses

## Strengths

The use of the HGF to predict neural surprise had many advantages. Firstly, it produced a dynamic learning rate which approximated Bayes-optimal belief updating, meaning that it no longer discarded information about e.g., the number of stimulus repetitions. This was reflected in greater predictive power of the HGF model compared to the traditional Deviant vs. Standard (implicit) classification model. Secondly, the use of the HGF allowed the comparison of high vs. low predicted surprise trials using any specified cut-off (and the demonstration of a difference between these categories): such analysis was not possible with the traditional Deviants vs. Standards model, with its binary class structure. Thirdly, the continuous class structure of the HGF also allowed for regression of the modelled surprise against indices of neural surprise (e.g., the peak GFP amplitude). Fourthly, the HGF provided a scaffold which allowed further modifications to improve predictive power. Indeed, the novel modifications that I made to the HGF to incorporate physical tone frequency information further improved HGF_Vanilla's ability to predict the degree of neural surprise. This recommends that future researchers control for absolute pitch change.

Secondly, I addressed a shortcoming in the existing literature by measuring aMMRs in individuals aged 10-16 and demonstrating that age-related differences previously found between young children and older adults using EEG were also detectable between individuals aged 10-16 and younger or older groups using MEG. The gathering of data from more participants further
supported the segregation of all participants by age quintile and the analysis of within and between-age group comparisons.

Lastly, I also improved upon previous statistical practises, in which only specific features of the aMMR were typically analysed (e.g., its peak amplitude, latency and area under the curve) by conducting permutation clustering analyses temporally (of the GFP) and spatiotemporally (of the ERF). In engaging a non-parametric permutation analysis, I was able to limit the family-wise error rate and thus search for significant MMRs at all latencies in all locations (allowing analysis of my questions of interest, such as whether activity differed by age) whilst controlling statistically for multiple comparisons. The clustering method maximised statistical power relative to independent t -tests/confidence intervals at each time point and brain location, which in turn facilitated the discovery of significant differences between predictive models and the trial types defined by these models.

## Weaknesses, And Potential Mitigators

I did not select brain regions of interest (ROIs), which lowered statistical power. This was both because regions are harder to identify and compare in MEG-only analysis (relative to EEG analysis, or MEG analysis augmented by source analysis) and because I had not wanted to potentially exclude regions which may differentiate the new conditions and models analysed. In future, I could engage in preliminary source analysis to select ROIs. This would involve playing auditory tones to select a ROI specific to each participant using some other independent criterion, e.g., by selecting the channels where the amplitude of the N 1 was found to be largest; such a method was for example used in (Luo et al., 2010).

It is also important to qualify that I have only examined auditory learning. It remains unclear whether the insights derived here apply to sensory systems of different modalities (e.g., whether a larger spatial error in a visual prediction task would analogously lead to larger prediction error), or to unexpected transitions in higher cognitive systems processing more abstract categories (e.g., the sudden presence of a turtle in an inappropriate context, such as inside a car). Prediction errors may not even apply in all auditory contexts, e.g., they may not generalise to predict neural activity well in situations with very long inter-stimulus intervals (ISIs) (over which information may be lost, and which rely more on long-term memory systems than sensory memory). Prediction errors might similarly manifest quite differently when the ISI is very short, as it would be unlikely that predictions can be formed quickly enough, which would then lead to bottlenecking in updating the neural model. As a result of such bottlenecks,
during these short ISIs, updates and learning consequently would be attenuated/impaired and observations would necessarily combine in some fashion (with a corresponding loss of sequencing information) to produce less frequent, 'chunked' model updates. In summary, speculatively, belief updating probably also involves non-Bayesian 'threads' which will reduce the predictive power of a Bayesian model like the HGF trial by trial. Subsequent research could vary the ISI to determine whether this occurs.

The specific MEG channels used were also not directly comparable between child and adult systems, which invites the question of whether the adult channels selected for comparison were 'correct'. The examined paired channels did show a high correlation (implying a similar orientation) and appear to be drawn from equal numbers of (equivalent contra-lateral) locations, and the gains of the adult and child systems were equal. However, channels were not paired deliberatively based on their orientations, and the correlations between all paired sensors (bar one) were not examined. Were any of these pairings inappropriate, this might bias my age comparison analysis. This limitation could be addressed through further examination of the pairings but was mitigated in several ways. Firstly, analysing the standardised signal reduced inter-individual differences as described. Secondly, the comparisons of GFP should be more robust than any comparison between specific regions of each system - as orientation differences may cancel out at a global scale. Finally, the subtraction of signals in response to different trial types should be yet more robust to machine-specific differences.

There are also some important theoretical weaknesses in the Bayesian models used. Firstly, the HGF predicts a one-dimensional surprise magnitude per trial (somewhat analogously to the peak aMMR signal amplitude) - and not any features of the aMMR such as its latency, sign, or location. Secondly, in Bayesian learning, it is always unclear what the appropriate scale is at which learning occurs; for example, language learning may occur at the level of phonemes or words. In this case, the sequential Bayesian learner (HGF) modelled transitions from the smallest stimulus packet size of a single tone to other single tones only, so I am unable to report on whether transition learning occurs over larger groupings of stimuli and, if so, whether this interacts with trial-by-trial transition learning. Thirdly, when fitting the whole tone sequence, the HGF 'cheats' by incorporating knowledge of the number of tones (and by seeding sensible estimates of parameters such as initial tone volatility). Fourth, it did not disambiguate precision decreasing with age vs. the attenuation of model (belief) updates. Fifth, the HGF assumes that two equally surprising events (i.e., events with the same HGF outputs) yield the same neural activity and that this purely reflects model adjustment (learning). But the neural activity
observed is the joint result of (at least) expectation, surprise, learning, inhibition, and salience. It can seen that surprise and learning may be dissociated (e.g., when inhibition is active, such as in 'denial') and that learning probably depends on saliency too (Treviño, 2015), e.g., if the tones are deemed irrelevant or threatening by higher cognitive systems. More basically, there is still little understanding of physiological basis (e.g., separate neurons for) prediction and surprise (Ficco et al., 2021) or how neurons could feasibly encapsulate both.

## Future Research

## Other types of stimuli

It is not clear if the statistics of the tones presented or if the models designed to predict surprise reflect ecologically realistic learning, in which the brain must encounter novel tones and may dynamically assemble new informative features (e.g., as in maximum entropy hidden Markov models, see Toutanova \& Manning, 2002). As examples, whilst I varied the number of repetitions of 'standards', it is unclear whether my findings will generalise to situations where the distribution of stimuli is non-uniform (e.g., where high-pitched tones are more likely). Similarly, my results may not hold where the transition probabilities (to different tones) are not uniform, and/or where the transition probabilities are non-stationary - either because new tones are introduced (i.e., previously unheard tones of lower or higher pitch are played) or because the probability of existing tones being played changes (e.g., the pitch 'trends' higher slowly, via increased likelihood of higher tones). To address these potential limitations, additional data was gathered e.g., one block used a non-uniform distribution of tones. However, it was beyond the scope of this thesis to analyse this data or comment on what it implies about the suitability of the HGF model for significantly different types of stimuli. Future researchers should synthesise findings from different auditory paradigms like these in order to evaluate whether the neural signatures of surprise to sequences with different statistics can all be reasonably described by the saode ks inierarchical Bayesian).

## Other Models

Future research should further investigate other models of surprise. As one example, we may model different brain sources using separate HGF models with different parameters, or may scale the prediction error at different channels using some metric e.g., based on the distance from auditory processing centres. As a second example, HGF_Pitch could instead calculate the expected magnitude (absolute value of) the pitch change and then scale the PE2 up or down depending on whether the magnitude of the actual pitch change was smaller or larger than this.

Indeed, I could also have regressed aMMR characteristics (e.g., GFP amplitude) against factors which appear to influence neural surprise, such as the specific stimulus transition (e.g., the magnitude of the pitch change) and latency e.g., see the rERP approach (Smith \& Kutas, 2015). After hearing the tone that we understand to be the highest pitched, we should only experience surprise about the magnitude of the frequency change but not its sign. In such a case we may expect a change in the latency at which the MMR becomes significant, since expectations for lower tones can be strongly activated. Further, tone transition models have several advantages which were not leveraged in this analysis. The HGF model could optionally involve separate precisions - one for the stored category (from which predictions are made, which may depend on memory) and one for the perceptual process (the perceived tone being the thing upon which surprise relative to a stored category is calculated). Such a separation may help explain the impact of deficits that are selective to perception (e.g., in damaged hearing) or to memory.

With regards to entrainment to pitch - if older individuals become more entrained to beat than pitch, future research may attempt to determine whether entrainment to beat is responsible for this greater GFP by making the tone onset less predictable and thus discouraging such entrainment in older individuals, but this may come at the cost of introducing other distortions in the neural signal.

Lastly, in seeking a predictor of neural activity, future researchers may instead seek to use a model designed to match the mathematical structure of the tones presented. When presenting tones using this Roving Oddball paradigm (which has tones which are non-random in the senses of being within a specific range, and discretised to specific values), researchers may seek a model which unlike the HGF, does not assume a Gaussian random walk of states. Conversely, future users of the HGF may expect that were it to perform well, it should do so when using a stimulus paradigm in which the tones also move along random walks (of pitch) themselves implying that there can be an unbounded number of tones (with a specific assumption about how unheard tones would be handled) spanning an unbounded range of frequencies. The caveat to this is that based on my results, the HGF would only be expected to work better if the magnitude of the pitch change formed a larger component of the predicted neural surprise.

## Explaining More Cognitive Phenomena

The HGF could also potentially be further improved in a way that it could start to explain broader cognitive phenomena involving beliefs which are biased towards particular points in
memory, or which change less slowly than is optimal - such as in anchoring (Tversky \& Kahneman, 1974) - or in confirmation bias (Nickerson, 1998). This could be accomplished were the HGF modified to have a faster rate of learning in the direction of previously held beliefs on single trials, independent of beliefs about precision. Conversely, different modifications could be made which would allow the HGF to describe learning which occurs faster than is Bayesoptimal (as generative models generally do, e.g., in "one-shot learning", see Weaver, 2015) and/or in a less biased manner in response to a single data point. Although it remains to be seen whether basic perceptual learning could also be pulled towards prior beliefs in the manner that more complex beliefs seemingly are, this appears plausible in light of phenomena found in lowlevel perceptual psychology. For example, in bistable vision (Rodríguez-Martínez \& CastilloParra, 2018), perception can be strongly biased by top-down activity towards a previous perception. Alternatively, perceptions which are more consistent with things which have been observed before (e.g., a series of black and white dots which are suddenly recognised as depicting a Dalmatian dog) might be suddenly stored with greater precision. Should this be the case, then there would an elegant theoretical continuity between learning at different scales. The attraction of such models is even greater when one considers that they should be able to begin to incorporate theories of learning concerning exemplars and prototypes (Bowman et al., 2020). An example might be a 'clustering' model, which would re-contextualise the relevance of the PE and PWPE in circumstances when there are multiple feasible predictions. In such circumstances, we may show small aMMRs, even for a previously unheard tone: indeed, aMMRs are less evident when there is no one dominant expectation (Garrido et al., 2013; Hsu et al., 2019). In such situations, a new tone may instead simply instantiate a new category. To the extent that alterations are made to existing categories, this may be modifications to that cluster's boundary without a canonical aMMR response (e.g., the mere incorporation of a new tone into an existing cluster which best resembles the new tone). Such cognitive operations may occur on the basis of events' relative surprise with respect to existing categories. This would be a form of higher-level learning involving neural activity which reflects a subtraction of PWPEs: this would still be underpinned by PE and PWPE computations, but the subtraction would be harder to detect neurally as a specific peak or trough in the GFP or ERF. If this model describes reality, greater experience in aged individuals may result in learning new categories and a proliferation of models and predictions which are valid in different contexts, but which produce smaller PEs and PWPEs, i.e., less strong adjustment of existing beliefs and categories. There may also be a dissociation between neural surprise and belief updating which depends on age.

## Source Analysis

I did not take advantage of the high spatial resolution afforded by MEG. Given that I have found differences between e.g., the HGF_Vanilla and HGF_Pitch model, it would be natural to follow this up with an analysis of where in the brain differences between these modes of processing appear to occur. The results from this can also then be reconciled with the latency differences described in this thesis to provide a more complete functional picture of the progression of surprise in relation to various tone features through a spatially organised neural hierarchy.

## Conclusions

Overall, my results suggest the responses evoked to Roving Oddball auditory stimuli are consistent with the Bayesian view of increased top-down activity in older persons and increased bottom-up activity in younger persons.

One conclusion from my research is that robust differences exist between younger and older people in the degree of anticipatory activity and in the latency and amplitude of the GFP and ERF. This pattern was also found to occur across those aged 10-16 in comparison to older and younger participants. Bayesian accounts of neural surprise may however need to explain in greater detail why such changes appear to be discontinuous with age, by testing which components of such models explain variations in neural surprise with age and by testing predictions about how the latencies and brain regions at which age-related differences should manifest.

A second conclusion is that surprise and learning vary in response to dimensions other than stimulus transition probability. In particular, both stimulus characteristics (such as the magnitude of pitch change and the 'hazard' function) and participant factors (age, and factors which vary with it such as short-term memory) appear to impact the pattern of evoked responses. Researchers building computational models of surprise should begin to incorporate such characteristics in their predictive models of surprise.

## References

Alain, C., Achim, A., \& Woods, D. L. (1999). Separate memory-related processing for auditory frequency and patterns. Psychophysiology, 36(6), 737-744. https://doi.org/10.1111/14698986.3660737

Andersen, L. M. (2018). Group Analysis in MNE-Python of Evoked Responses from a Tactile Stimulation Paradigm: A Pipeline for Reproducibility at Every Step of Processing, Going from Individual Sensor Space Representations to an across-Group Source Space Representation [Protocols]. Frontiers in Neuroscience, 12. https://doi.org/10.3389/fnins.2018.00006
Arendt, D., Tosches, M. A., \& Marlow, H. (2016). From nerve net to nerve ring, nerve cord and brain-evolution of the nervous system. Nature reviews. Neuroscience, 17(1), 61-72. https://doi.org/10.1038/nrn.2015.15
Auksztulewicz, R., \& Friston, K. (2015). Attentional Enhancement of Auditory Mismatch Responses: a DCM/MEG Study. Cerebral cortex (New York, N.Y. : 1991), 25(11), 4273-4283. https://doi.org/10.1093/cercor/bhu323
Baldeweg, T., Klugman, A., Gruzelier, J., \& Hirsch, S. R. (2004). Mismatch negativity potentials and cognitive impairment in schizophrenia. Schizophr Res, 69(2-3), 203-217. https://doi.org/10.1016/j.schres.2003.09.009
Barto, A., Mirolli, M., \& Baldassarre, G. (2013). Novelty or Surprise? [Hypothesis and Theory]. Frontiers in Psychology, 4. https://doi.org/10.3389/fpsyg.2013.00907
Beauchemin, M., \& De Beaumont, L. (2005). Statistical analysis of the mismatch negativity: To a dilemma, an answer. Tutorials in quantitative methods for psychology, $1(1)$, 18-24. https://doi.org/10.20982/tqmp.01.1.p018
Bertoli, S., Smurzynski, J., \& Probst, R. (2002). Temporal resolution in young and elderly subjects as measured by mismatch negativity and a psychoacoustic gap detection task. Clinical neurophysiology, 113(3), 396-406. https://doi.org/10.1016/S1388-2457(02)00013-5
Bonetti, L., Haumann, N. T., Brattico, E., Kliuchko, M., Vuust, P., Särkämö, T., \& Näätänen, R. (2018). Auditory sensory memory and working memory skills: Association between frontal MMN and performance scores. Brain Res, 1700, 86-98. https://doi.org/10.1016/j.brainres.2018.06.034
Bowman, C. R., Iwashita, T., \& Zeithamova, D. (2020). Tracking prototype and exemplar representations in the brain across learning. eLife, 9, e59360. https://doi.org/10.7554/eLife. 59360
Brattico, E., Tervaniemi, M., Näätänen, R., \& Peretz, I. (2006). Musical scale properties are automatically processed in the human auditory cortex. Brain Research, 1117(1), 162-174. https://doi.org/10.1016/j.brainres.2006.08.023
Cheng, C.-H., Baillet, S., Hsiao, F.-J., \& Lin, Y.-Y. (2013). Effects of aging on neuromagnetic mismatch responses to pitch changes. Neuroscience letters, 544, 20-24. https://doi.org/10.1016/j.neulet.2013.02.063
Cheng, C.-H., Hsu, W.-Y., \& Lin, Y.-Y. (2013). Effects of physiological aging on mismatch negativity: A meta-analysis. International Journal of Psychophysiology, 90(2), 165-171. https://doi.org/10.1016/j.ijpsycho.2013.06.026
Cheour, M., Imada, T., Taulu, S., Ahonen, A., Salonen, J., \& Kuhl, P. (2004). Magnetoencephalography is feasible for infant assessment of auditory discrimination. Experimental neurology, 190, 44-51. https://doi.org/10.1016/j.expneurol.2004.06.030
Cho, R. Y., Nystrom, L. E., Brown, E. T., Jones, A. D., Braver, T. S., Holmes, P. J., \& Cohen, J. D. (2002). Mechanisms underlying dependencies of performance on stimulus history in a twoalternative forced-choice task. Cognitive, affective, \& behavioral neuroscience, 2(4), 283-299. https://doi.org/10.3758/CABN.2.4.283
Chumbley, J. R., Burke, C. J., Stephan, K. E., Friston, K. J., Tobler, P. N., \& Fehr, E. (2014). Surprise beyond prediction error. Human brain mapping, 35(9), 4805-4814. https://doi.org/10.1002/hbm. 22513
Cooper, R. J., Todd, J., McGill, K., \& Michie, P. T. (2006). Auditory sensory memory and the aging brain: A mismatch negativity study. Neurobiology of Aging, 27(5), 752-762.
https://doi.org/https://doi.org/10.1016/j.neurobiolaging.2005.03.012

Cooper, R. J., Todd, J., McGill, K., \& Michie, P. T. (2006). Auditory sensory memory and the aging brain: A mismatch negativity study. Neurobiol Aging, 27(5), 752-762. https://doi.org/10.1016/j.neurobiolaging.2005.03.012
Cumming, G. (2013). Understanding the new statistics: Effect sizes, confidence intervals, and metaanalysis. Routledge.
de Cheveigné, A., \& Simon, J. Z. (2007). Denoising based on time-shift PCA. Journal of neuroscience methods, 165(2), 297-305. https://doi.org/10.1016/j.jneumeth.2007.06.003
Deserno, L., Boehme, R., Mathys, C., Katthagen, T., Kaminski, J., Stephan, K. E., Heinz, A., \& Schlagenhauf, F. (2020). Volatility Estimates Increase Choice Switching and Relate to Prefrontal Activity in Schizophrenia. Biological psychiatry : cognitive neuroscience and neuroimaging, 5(2), 173-183. https://doi.org/10.1016/j.bpsc.2019.10.007
Erdeniz, B., Serin, E., İbadi, Y., \& Taş, C. (2017). Decreased functional connectivity in schizophrenia: The relationship between social functioning, social cognition and graph theoretical network measures. Psychiatry Res Neuroimaging, 270, 22-31. https://doi.org/10.1016/j.pscychresns.2017.09.011
Faraji, M., Preuschoff, K., \& Gerstner, W. (2018). Balancing new against old information: The role of puzzlement surprise in learning. Neural computation, 30(1), 34-83. https://doi.org/10.1162/NECO_a_01025
Ficco, L., Mancuso, L., Manuello, J., Teneggi, A., Liloia, D., Duca, S., Costa, T., Kovacs, G. Z., \& Cauda, F. (2021). Disentangling predictive processing in the brain: a meta-analytic study in favour of a predictive network. Scientific Reports, $11(1)$, 16258. https://doi.org/10.1038/s41598-021-95603-5
Files, B. T., Lawhern, V. J., Ries, A. J., \& Marathe, A. R. (2016). A Permutation Test for Unbalanced Paired Comparisons of Global Field Power. Brain Topography, 29(3), 345-357. https://doi.org/10.1007/s10548-016-0477-3
Fitzgerald, K., \& Todd, J. (2020). Making Sense of Mismatch Negativity. Frontiers in Psychiatry, 11, 468-468. https://doi.org/10.3389/fpsyt. 2020.00468
Fortenbaugh, F. C., DeGutis, J., Germine, L., Wilmer, J. B., Grosso, M., Russo, K., \& Esterman, M. (2015). Sustained Attention Across the Life Span in a Sample of 10,000: Dissociating Ability and Strategy. Psychological Science, 26(9), 1497-1510. https://doi.org/10.1177/0956797615594896
Frässle, S., Aponte, E. A., Bollmann, S., Brodersen, K. H., Do, C. T., Harrison, O. K., Harrison, S. J., Heinzle, J., Iglesias, S., Kasper, L., Lomakina, E. I., Mathys, C., Müller-Schrader, M., Pereira, I., Petzschner, F. H., Raman, S., Schöbi, D., Toussaint, B., Weber, L. A., . . . Stephan, K. E. (2021). TAPAS: An Open-Source Software Package for Translational Neuromodeling and Computational Psychiatry. Frontiers in Psychiatry, 12, 680811-680811. https://doi.org/10.3389/fpsyt.2021.680811
Friston, K. (2010). The free-energy principle: a unified brain theory? Nature Reviews Neuroscience, 11(2), 127-138. https://doi.org/10.1038/nrn2787
Garrido, M. I., Friston, K. J., Kiebel, S. J., Stephan, K. E., Baldeweg, T., \& Kilner, J. M. (2008). The functional anatomy of the MMN: a DCM study of the roving paradigm. NeuroImage, 42(2), 936944. https://doi.org/10.1016/j.neuroimage.2008.05.018

Gerstner, W., \& Kistler, W. M. (2002). Spiking Neuron Models: Single Neurons, Populations, Plasticity. Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9780511815706
Giard, M. H., Perrin, F., Pernier, J., \& Bouchet, P. (1990). Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study. Psychophysiology, 27(6), 627-640. https://doi.org/10.1111/j.1469-8986.1990.tb03184.x
Gijsen, S., Grundei, M., Lange, R. T., Ostwald, D., \& Blankenburg, F. (2021). Neural surprise in somatosensory Bayesian learning. PLoS computational biology, 17(2), e1008068-e1008068. https://doi.org/10.1371/journal.pcbi. 1008068
Gogtay, N., Vyas, N. S., Testa, R., Wood, S. J., \& Pantelis, C. (2011). Age of onset of schizophrenia: perspectives from structural neuroimaging studies. Schizophrenia Bulletin, 37(3), 504-513. https://doi.org/10.1093/schbul/sbr030
Griffiths, O., Shehabi, N., Murphy, R. A., \& Le Pelley, M. E. (2019). Superstition predicts perception of illusory control. Br J Psychol, 110(3), 499-518. https://doi.org/10.1111/bjop. 12344

Groppe, D. M., Choi, M., Huang, T., Schilz, J., Topkins, B., Urbach, T. P., \& Kutas, M. (2010). The phonemic restoration effect reveals pre-N400 effect of supportive sentence context in speech perception. Brain Res, 1361, 54-66. https://doi.org/10.1016/j.brainres.2010.09.003

Groppe, D. M., Urbach, T. P., \& Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. Psychophysiology, 48(12), 1711-1725. https://doi.org/10.1111/j.1469-8986.2011.01273.x
Haenschel, C., Vernon, D. J., Dwivedi, P., Gruzelier, J. H., \& Baldeweg, T. (2005). Event-Related Brain Potential Correlates of Human Auditory Sensory Memory-Trace Formation. The Journal of Neuroscience, 25(45), 10494. https://doi.org/10.1523/JNEUROSCI.1227-05.2005
Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., \& Lounasmaa, O. V. (1993). Magnetoencephalography-theory, instrumentation, and applications to noninvasive studies of the working human brain. Reviews of modern physics, 65(2), 413-497. https://doi.org/10.1103/RevModPhys.65.413
He, W., Sowman, P. F., Brock, J., Etchell, A. C., Stam, C. J., \& Hillebrand, A. (2019). Increased segregation of functional networks in developing brains. NeuroImage (Orlando, Fla.), 200, 607620. https://doi.org/10.1016/j.neuroimage.2019.06.055

Hedden, T., \& Gabrieli, J. D. E. (2004). Insights into the ageing mind: a view from cognitive neuroscience. Nature reviews. Neuroscience, 5(2), 87-96. https://doi.org/10.1038/nrn1323
Hedge, C., Stothart, G., Todd Jones, J., Rojas Frías, P., Magee, K. L., \& Brooks, J. C. W. (2015). A frontal attention mechanism in the visual mismatch negativity. Behavioural Brain Research, 293, 173-181. https://doi.org/https://doi.org/10.1016/j.bbr.2015.07.022
Henry, R. (2015). Etymologia: Bonferroni correction. Emerging infectious diseases, 21(2), 289-289. https://doi.org/10.3201/eid2102.et2102
Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics, 6(2), 65-70. http://www.jstor.org/stable/4615733
Hsu, Y.-F., Waszak, F., \& Hämäläinen, J. A. (2019). Prior Precision Modulates the Minimization of Auditory Prediction Error [Original Research]. Frontiers in Human Neuroscience, 13. https://doi.org/10.3389/fnhum.2019.00030
Huanping, D. A. I., \& Micheyl, C. (2011). Psychometric functions for pure-tone frequency discrimination. The Journal of the Acoustical Society of America, 130(1), 263-272. https://doi.org/10.1121/1.3598448
Hughes, H. C., Darcey, T. M., Barkan, H. I., Williamson, P. D., Roberts, D. W., \& Aslin, C. H. (2001). Responses of Human Auditory Association Cortex to the Omission of an Expected Acoustic Event. NeuroImage (Orlando, Fla.), 13(6), 1073-1089. https://doi.org/10.1006/nimg. 2001.0766
Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., \& Gramfort, A. (2017). Autoreject: Automated artifact rejection for MEG and EEG data. NeuroImage, 159, 417-429. https://doi.org/10.1016/j.neuroimage.2017.06.030
Kéïta, L., Mottron, L., Dawson, M., \& Bertone, A. (2011). Atypical Lateral Connectivity: A Neural Basis for Altered Visuospatial Processing in Autism. Biological psychiatry (1969), 70(9), 806-811. https://doi.org/10.1016/j.biopsych.2011.07.031
King, J.-R., Gramfort, A., Schurger, A., Naccache, L., \& Dehaene, S. (2014). Two distinct dynamic modes subtend the detection of unexpected sounds. PLoS ONE, 9(1), e85791. https://doi.org/10.1371/journal.pone. 0085791
Kisley, M. A., Davalos, D. B., Engleman, L. L., Guinther, P. M., \& Davis, H. P. (2005). Age-related change in neural processing of time-dependent stimulus features. Brain research. Cognitive brain research, 25(3), 913-925. https://doi.org/10.1016/j.cogbrainres.2005.09.014
Kluge, C., Bauer, M., Leff, A. P., Heinze, H.-J., Dolan, R. J., \& Driver, J. (2011). Plasticity of human auditory-evoked fields induced by shock conditioning and contingency reversal. Proceedings of the National Academy of Sciences, 108(30), 12545. https://doi.org/10.1073/pnas. 1016124108
Kolossa, A., Fingscheidt, T., Wessel, K., \& Kopp, B. (2012). A model-based approach to trial-by-trial p300 amplitude fluctuations. Frontiers in Human Neuroscience, 6, 359-359. https://doi.org/10.3389/fnhum.2012.00359

Korovkin, S., Vladimirov, I., Chistopolskaya, A., \& Savinova, A. (2018). How Working Memory Provides Representational Change During Insight Problem Solving [Original Research]. Frontiers in Psychology, 9. https://doi.org/10.3389/fpsyg.2018.01864
Krasensky, J., \& Jonak, C. (2012). Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. Journal of Experimental Botany, 63(4), 1593-1608. https://doi.org/10.1093/jxb/err460
Kullback, S., \& Leibler, R. A. (1951). On information and sufficiency. The annals of mathematical statistics, 22(1), 79-86.
Lecaignard, F., Bertrand, O., Caclin, A., \& Mattout, J. (2021). Neurocomputational underpinnings of expected surprise. bioRxiv, 501221. https://doi.org/10.1101/501221
Lee, I.-S., Jung, W.-M., Lee, Y.-S., Wallraven, C., \& Chae, Y. (2021). Operant and classical learning principles underlying mind-body interaction in pain modulation: a pilot fMRI study. Scientific Reports, $11(1)$, 1663. https://doi.org/10.1038/s41598-021-81134-6
Lehmann, D., \& Skrandies, W. (1980). Reference-free identification of components of checkerboardevoked multichannel potential fields. Electroencephalography and clinical neurophysiology, 48(6), 609-621. https://doi.org/10.1016/0013-4694(80)90419-8
Lieder, F., Daunizeau, J., Garrido, M. I., Friston, K. J., \& Stephan, K. E. (2013). Modelling trial-by-trial changes in the mismatch negativity. PLoS Comput Biol, 9(2), el002911. https://doi.org/10.1371/journal.pcbi. 1002911
Lomakina, E. Computational and Physiological Models. Computational Psychiatry Seminar, Universität Zürich, March 7, 2014. https://slideplayer.com/slide/6269399/
Long, M. (2014). Human Perception and Reaction to Sound. In (pp. 81-127). https://doi.org/10.1016/B978-0-12-398258-2.00003-9
Los, S. A., Kruijne, W., \& Meeter, M. (2014). Outlines of a multiple trace theory of temporal preparation. Frontiers in Psychology, 5, 1058-1058. https://doi.org/10.3389/fpsyg.2014.01058
Luo, H., Liu, Z., \& Poeppel, D. (2010). Auditory Cortex Tracks Both Auditory and Visual Stimulus Dynamics Using Low-Frequency Neuronal Phase Modulation. PLoS biology, 8(8), e1000445. https://doi.org/10.1371/journal.pbio. 1000445
Maheu, M., Dehaene, S., \& Meyniel, F. (2019). Brain signatures of a multiscale process of sequence learning in humans. eLife, 8 . https://doi.org/10.7554/eLife. 41541
Maris, E., \& Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods, 164(1), 177-190. https://doi.org/10.1016/j.jneumeth.2007.03.024
Marshall, L., Mathys, C., Ruge, D., de Berker, A. O., Dayan, P., Stephan, K. E., \& Bestmann, S. (2016). Pharmacological Fingerprints of Contextual Uncertainty. PLoS biology, 14(11), e1002575e1002575. https://doi.org/10.1371/journal.pbio. 1002575
Martinez, P., \& Sprecher, S. G. (2020). Of Circuits and Brains: The Origin and Diversification of Neural Architectures [Review]. Frontiers in Ecology and Evolution, 8. https://doi.org/10.3389/fevo.2020.00082
Mathys, C., Daunizeau, J., Friston, K. J., \& Stephan, K. E. (2011). A bayesian foundation for individual learning under uncertainty. Frontiers in Human Neuroscience, 5, 39-39. https://doi.org/10.3389/fnhum.2011.00039
Mathys, C. D., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., \& Stephan, K. E. (2014). Uncertainty in perception and the Hierarchical Gaussian Filter [Original Research]. Frontiers in Human Neuroscience, 8. https://doi.org/10.3389/fnhum.2014.00825
MATLAB. (2018). 9.7.0.1190202 (R2019b). Natick, Massachusetts: The MathWorks Inc.
McCallum, A., Freitag, D., Pereira, F. (2000). Maximum Entropy Markov Models for Information Extraction and Segmentation. Proc. ICML 2000. pp. 591-598.
May, P. J. C., \& Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. Psychophysiology, 47(1), 66-122. https://doi.org/10.1111/j.14698986.2009.00856.x

McCarthy, G., Huettel, S. A., \& Mack, P. B. (2002). Perceiving patterns in random series: dynamic processing of sequence in prefrontal cortex. Nature Neuroscience, 5(5), 485-490. https://doi.org/10.1038/nn841

Meyniel, F., Maheu, M., \& Dehaene, S. (2016). Human Inferences about Sequences: A Minimal Transition Probability Model. PLoS computational biology, 12(12), e1005260-e1005260. https://doi.org/10.1371/journal.pcbi. 1005260
Modirshanechi, A., Kiani, M. M., \& Aghajan, H. (2019). Trial-by-trial surprise-decoding model for visual and auditory binary oddball tasks. NeuroImage (Orlando, Fla.), 196, 302-317. https://doi.org/10.1016/j.neuroimage.2019.04.028
Moran, R., Symmonds, M., Dolan, R., \& Friston, K. (2014). The Brain Ages Optimally to Model Its Environment: Evidence from Sensory Learning over the Adult Lifespan. PLoS computational biology, 10, e1003422. https://doi.org/10.1371/journal.pcbi. 1003422
Moran, R. J., Campo, P., Symmonds, M., Stephan, K. E., Dolan, R. J., \& Friston, K. J. (2013). Free energy, precision and learning: the role of cholinergic neuromodulation. J Neurosci, 33(19), 8227-8236. https://doi.org/10.1523/jneurosci.4255-12.2013
Murray, M. M., Brunet, D., \& Michel, C. M. (2008). Topographic ERP analyses: a step-by-step tutorial review. Brain Topogr, 20(4), 249-264. https://doi.org/10.1007/s10548-008-0054-5
Näätänen, R. (2000). Mismatch negativity (MMN): perspectives for application. International Journal of Psychophysiology, 37(1), 3-10. https://doi.org/10.1016/S0167-8760(00)00091-X
Näätänen, R., Gaillard, A. W. K., \& Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. Acta psychologica, 42(4), 313-329. https://doi.org/10.1016/0001-6918(78)90006-9
Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. Review of general psychology, 2(2), 175-220.
O'Reilly, J. A., \& O'Reilly, A. (2021). A Critical Review of the Deviance Detection Theory of Mismatch Negativity. NeuroSci, 2(2), 151-165. https://www.mdpi.com/2673-4087/2/2/11
Ouden, H. E. M. d., Friston, K. J., Daw, N. D., McIntosh, A. R., \& Stephan, K. E. (2009). A dual role for prediction error in associative learning. Cerebral cortex (New York, N.Y. 1991), 19(5), 11751185. https://doi.org/10.1093/cercor/bhn161

Oxenham, A. J. (2013). Revisiting place and temporal theories of pitch. Acoustical science and technology, 34(6), 388-396. https://doi.org/10.1250/ast.34.388
Pak, A., Kissinger, S. T., \& Chubykin, A. A. (2021). Impaired Adaptation and Laminar Processing of the Oddball Paradigm in the Primary Visual Cortex of Fmr1 KO Mouse. Front Cell Neurosci, 15, 668230. https://doi.org/10.3389/fncel.2021.668230

Park, Y. H., Shin, S.-H., Byun, S. W., \& Kim, J. Y. (2016). Age- and Gender-Related Mean Hearing Threshold in a Highly-Screened Population: The Korean National Health and Nutrition Examination Survey 2010-2012. PLoS ONE, 11(3), e0150783-e0150783. https://doi.org/10.1371/journal.pone. 0150783
Pekkonen, E. (2000). Mismatch negativity in aging and in Alzheimer's and Parkinson's diseases. Audiology and neurotology, 5(3-4), 216-224. https://doi.org/10.1159/000013883
Phillips, H. N., Blenkmann, A., Hughes, L. E., Bekinschtein, T. A., \& Rowe, J. B. (2015). Hierarchical Organization of Frontotemporal Networks for the Prediction of Stimuli across Multiple Dimensions. The Journal of Neuroscience, 35(25), 9255. https://doi.org/10.1523/JNEUROSCI.5095-14.2015
Picton, T. W., Alain, C., Otten, L., Ritter, W., \& Achim, A. (2000). Mismatch negativity: different water in the same river. Audiol Neurootol, 5(3-4), 111-139. https://doi.org/10.1159/000013875
Pliatsikas, C., Veríssimo, J., Babcock, L., Pullman, M. Y., Glei, D. A., Weinstein, M., Goldman, N., \& Ullman, M. T. (2019). Working memory in older adults declines with age, but is modulated by sex and education. Q J Exp Psychol (Hove), 72(6), 1308-1327. https://doi.org/10.1177/1747021818791994
Powers, A. R., Mathys, C., \& Corlett, P. R. (2017). Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. Science, 357(6351), 596-600. https://doi.org/10.1126/science.aan3458
Puts, N. A., Wodka, E. L., Tommerdahl, M., Mostofsky, S. H., \& Edden, R. A. (2014). Impaired tactile processing in children with autism spectrum disorder. J Neurophysiol, 111(9), 1803-1811. https://doi.org/10.1152/jn.00890.2013

Ramachandran, S., Meyer, T., \& Olson, C. R. (2016). Prediction suppression in monkey inferotemporal cortex depends on the conditional probability between images. J Neurophysiol, 115(1), 355-362. https://doi.org/10.1152/jn.00091.2015
Rao, R. P. N., \& Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature Neuroscience, 2(1), 79-87. https://doi.org/10.1038/4580
Rapaport, H., Seymour, R., Sowman, P., Benikos, N., Stylianou, E., Johnson, B., Crain, S., \& He, W. (2019). Studying Brain Function in Children Using Magnetoencephalography. Journal of Visualized Experiments, 146, 58909. https://doi.org/10.3791/58909
Rapaport, H., Seymour, R. A., He, W., Benikos, N., Sun, Y., Pellicano, E., \& Sowman, P. (2022). Investigating the Mismatch Field in Young Children and Adults Using MEG and a Roving Auditory Oddball Paradigm. Manuscript in preparation.
Reed, E. J., Uddenberg, S., Suthaharan, P., Mathys, C. D., Taylor, J. R., Groman, S. M., \& Corlett, P. R. (2020). Paranoia as a deficit in non-social belief updating. eLife, 9, e56345. https://doi.org/10.7554/eLife.56345
Rescorla, R. A., \& Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black \& W. F. Prokasy (Eds.), Classical conditioning II: Current research and theory (pp. 64-99). New York: Appleton-Century Crofts
Rodríguez-Martínez, G. A., \& Castillo-Parra, H. (2018). Bistable perception: neural bases and usefulness in psychological research. Int J Psychol Res (Medellin), 11(2), 63-76. https://doi.org/10.21500/20112084.3375
Rose, M., Haider, H., \& Büchel, C. (2005). Unconscious detection of implicit expectancies. J Cogn Neurosci, 17(6), 918-927. https://doi.org/10.1162/0898929054021193
Ruggero, M. A. (1992). Responses to sound of the basilar membrane of the mammalian cochlea. Current opinion in neurobiology, 2(4), 449-456. https://doi.org/https://doi.org/10.1016/0959-4388(92)90179-O
Saffran, J. R., Aslin, R. N., \& Newport, E. L. (1996). Statistical learning by 8-month-old infants. Science, 274(5294), 1926-1928. https://doi.org/10.1126/science.274.5294.1926
Salisbury, D. F. (2012). Finding the missing stimulus mismatch negativity (MMN): emitted MMN to violations of an auditory gestalt. Psychophysiology, 49(4), 544-548. https://doi.org/10.1111/j.1469-8986.2011.01336.x
Salvi, R., Ding, D., Jiang, H., Chen, G. D., Greco, A., Manohar, S., Sun, W., \& Ralli, M. (2018). Hidden Age-Related Hearing Loss and Hearing Disorders: Current Knowledge and Future Directions. Hearing, balance and communication, 16(2), 74-82. https://doi.org/10.1080/21695717.2018.1442282
SanMiguel, I., Costa-Faidella, J., Lugo, Z. R., Vilella, E., \& Escera, C. (2021). Standard Tone Stability as a Manipulation of Precision in the Oddball Paradigm: Modulation of Prediction Error Responses to Fixed-Probability Deviants [Original Research]. Frontiers in Human Neuroscience, 15. https://doi.org/10.3389/fnhum.2021.734200
Sarasso, P., Neppi-Modona, M., Rosaia, N., Perna, P., Barbieri, P., Del Fante, E., Ricci, R., Sacco, K., \& Ronga, I. (2021). Nice and easy: Mismatch negativity responses reveal a significant correlation between aesthetic appreciation and perceptual learning. J Exp Psychol Gen. https://doi.org/10.1037/xge0001149
Schiff, S., Valenti, P., Andrea, P., Lot, M., Bisiacchi, P., Gatta, A., \& Amodio, P. (2008). The effect of aging on auditory components of event-related brain potentials. Clin Neurophysiol, 119(8), 17951802. https://doi.org/10.1016/j.clinph.2008.04.007

Schiff, S., Valenti, P., Andrea, P., Lot, M., Bisiacchi, P., Gatta, A., \& Amodio, P. (2008). The effect of aging on auditory components of event-related brain potentials. Clinical neurophysiology, 119(8), 1795-1802. https://doi.org/https://doi.org/10.1016/j.clinph.2008.04.007
Schwartenbeck, P., FitzGerald, T. H. B., \& Dolan, R. (2016). Neural signals encoding shifts in beliefs. NeuroImage, 125, 578-586. https://doi.org/10.1016/j.neuroimage.2015.10.067
Shaw, L., \& Bhaga, S. (2012). Online EMG Signal Analysis for diagnosis of Neuromuscular diseases by using PCA and PNN. International Journal Of Engineering Science and Technology 0975-5462, 4, 4453-4459.

Singh, S. P. (2014). Magnetoencephalography: Basic principles. Annals of Indian Academy of Neurology, 17(Suppl 1), S107-S112. https://doi.org/10.4103/0972-2327.128676
Smith, N. J., \& Kutas, M. (2015). Regression-based estimation of ERP waveforms: I. The rERP framework. Psychophysiology, 52(2), 157-168. https://doi.org/10.1111/psyp. 12317
Spolladore, L., Gelfusa, M., Rossi, R., \& Murari, A. (2021). Improved Treatment of the Independent Variables for the Deployment of Model Selection Criteria in the Analysis of Complex Systems. Entropy (Basel, Switzerland), 23(9), 1202. https://doi.org/10.3390/e23091202
Squires, K. C., Wickens, C., Squires, N. K., \& Donchin, E. (1976). The Effect of Stimulus Sequence on the Waveform of the Cortical Event-Related Potential. Science (American Association for the Advancement of Science), 193(4258), 1142-1146. https://doi.org/10.1126/science. 959831
Stefanics, G., Heinzle, J., Horváth, A. A., \& Stephan, K. E. (2018). Visual Mismatch and Predictive Coding: A Computational Single-Trial ERP Study. J Neurosci, 38(16), 4020-4030. https://doi.org/10.1523/jneurosci.3365-17.2018
Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., Petrovic, P., Uhlhaas, P., Voss, M., \& Corlett, P. R. (2018). The Predictive Coding Account of Psychosis. Biological psychiatry (1969), 84(9), 634-643. https://doi.org/10.1016/j.biopsych.2018.05.015
Stone, J. V. (2002). Independent component analysis: an introduction. Trends in Cognitive Sciences, 6(2), 59-64. https://doi.org/https://doi.org/10.1016/S1364-6613(00)01813-1
Strauss, M., Sitt, J. D., King, J.-R., Elbaz, M., Azizi, L., Buiatti, M., Naccache, L., van Wassenhove, V., \& Dehaene, S. (2015). Disruption of hierarchical predictive coding during sleep. Proceedings of the National Academy of Sciences - PNAS, 112(11), E1353-E1362. https://doi.org/10.1073/pnas. 1501026112 (PNAS Plus)
Strömmer, J. M., Tarkka, I. M., \& Astikainen, P. (2014). Somatosensory mismatch response in young and elderly adults [Original Research]. Frontiers in Aging Neuroscience, 6. https://doi.org/10.3389/fnagi.2014.00293
Takegata, R., Brattico, E., Tervaniemi, M., Varyagina, O., Näätänen, R., \& Winkler, I. (2005). Preattentive representation of feature conjunctions for concurrent spatially distributed auditory objects. Brain research. Cognitive brain research, 25(1), 169-179. https://doi.org/10.1016/j.cogbrainres.2005.05.006
Takegata, R., Paavilainen, P., Näätänen, R., \& Winkler, I. (1999). Independent processing of changes in auditory single features and feature conjunctions in humans as indexed by the mismatch negativity. Neurosci Lett, 266(2), 109-112. https://doi.org/10.1016/s0304-3940(99)00267-0
Tanner, A., \& Dounavi, K. (2021). The Emergence of Autism Symptoms Prior to 18 Months of Age: A Systematic Literature Review. Journal of Autism and Developmental Disorders, 51(3), 973-993. https://doi.org/10.1007/s 10803-020-04618-w
Tavano, A., Schröger, E., Kotz, S. (2019). Beta power encodes contextual estimates of temporal event probability in the human brain. PLoS ONE 14(9): e0222420-e0222420.
Taylor, M. J. (1988). Developmental changes in ERPs to visual language stimuli. Biological psychology, 26(1), 321-338. https://doi.org/https://doi.org/10.1016/0301-0511(88)90027-0
Tervaniemi, M., Saarinen, J., Paavilainen, P., Danilova, N., \& Näätänen, R. (1994). Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity. Biological psychology, 38(2), 157-167. https://doi.org/10.1016/0301-0511(94)90036-1
Thompson, E. C., White-Schwoch, T., Tierney, A., \& Kraus, N. (2015). Beat Synchronization across the Lifespan: Intersection of Development and Musical Experience. PLoS ONE, 10(6), e0128839e0128839. https://doi.org/10.1371/journal.pone. 0128839
Thönnessen, H., Zvyagintsev, M., Harke, K. C., Boers, F., Dammers, J., Norra, C., \& Mathiak, K. (2008). Optimized mismatch negativity paradigm reflects deficits in schizophrenia patients A combined EEG and MEG study. Biological psychology, 77(2), 205-216.
Todd, J., Harms, L., Schall, U., \& Michie, P. T. (2013). Mismatch negativity: translating the potential. Frontiers in Psychiatry, 4, 171-171. https://doi.org/10.3389/fpsyt.2013.00171
Toutanova, K., \& Manning, C. (2002). Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger. https://doi.org/10.3115/1117794.1117802
Treviño, M. (2015). Associative Learning Through Acquired Salience. Front Behav Neurosci, 9, 353. https://doi.org/10.3389/fnbeh.2015.00353

Tribus, M. (1961). Information Theory as the Basis for Thermostatics and Thermodynamics. Journal of applied mechanics, 28(1), 1-8. https://doi.org/10.1115/1.3640461
Tsogli, V., Jentschke, S., Daikoku, T., \& Koelsch, S. (2019). When the statistical MMN meets the physical MMN. Scientific Reports, 9(1), 5563. https://doi.org/10.1038/s41598-019-42066-4
Tversky, A., \& Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. Science, 185(4157), 1124-1131.
Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., \& Wagemans, J. (2014). Precise Minds in Uncertain Worlds: Predictive Coding in Autism. Psychological review, 121(4), 649-675. https://doi.org/10.1037/a0037665
van Dinteren, R., Huster, R. J., Jongsma, M. L. A., Kessels, R. P. C., \& Arns, M. (2018). Differences in Cortical Sources of the Event-Related P3 Potential Between Young and Old Participants Indicate Frontal Compensation. Brain Topography, 31(1), 35-46. https://doi.org/10.1007/s10548-016-0542-y
Vuust, P., Dietz, M. J., Witek, M., \& Kringelbach, M. L. (2018). Now you hear it: a predictive coding model for understanding rhythmic incongruity: Now you hear it. Annals of the New York Academy of Sciences, 1423(1), 19-29. https://doi.org/10.1111/nyas. 13622
Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., \& Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. Proceedings of the National Academy of Sciences - PNAS, 108(51), 20754-20759. https://doi.org/10.1073/pnas. 1117807108
Weaver, J. (2015). How one-shot learning unfolds in the brain. PLoS Biol, 13(4), e1002138. https://doi.org/10.1371/journal.pbio. 1002138
Weber, L. A., Diaconescu, A. O., Mathys, C., Schmidt, A., Kometer, M., Vollenweider, F., \& Stephan, K. E. (2020). Ketamine Affects Prediction Errors about Statistical Regularities: A Computational Single-Trial Analysis of the Mismatch Negativity. The Journal of neuroscience : the official journal of the Society for Neuroscience, 40(29), 5658-5668. https://doi.org/10.1523/JNEUROSCI.3069-19.2020
Woldorff, M. G., Hillyard, S. A., Gallen, C. C., Hampson, S. R., \& Bloom, F. E. (1998). Magnetoencephalographic recordings demonstrate attentional modulation of mismatch-related neural activity in human auditory cortex. Psychophysiology, 35(3), 283-292. https://doi.org/10.1017/s0048577298961601
Yassa, M. A., \& Reagh, Z. M. (2013). Competitive trace theory: A role for the hippocampus in contextual interference during retrieval. Frontiers in behavioral neuroscience, 7, 107-107. https://doi.org/10.3389/fnbeh. 2013.00107
Yoon, B.-J. (2009). Hidden Markov Models and their Applications in Biological Sequence Analysis. Current genomics, 10(6), 402-415. https://doi.org/10.2174/138920209789177575

## Appendix A

## Code Repository

## The code is hosted at: https://github.com/LanceAbel/MQ_MEG_Analysis

String documentation is included.
The primary files are:

- participant.py for analysing individuals' trial data
- experiment.py for analysing group data

There is also a Readme.md file which contains links to model code (PS, CS and BS) as well.

The codes representing each condition in the scripts are as follows:

## Table A4

Scripting codes used for each condition

| Condition contrast | Codes |
| :---: | :---: |
| Controls |  |
| 600 Hz tones vs. 700 Hz tones | 600 vs. 700 |
| 650 Hz tones vs. 800 Hz tones | 650 vs. 800 |
| Deviant vs. standard (all) | 99 vs. 88 |
| Deviant after 7 standards vs. pre-deviant ( $7^{\text {th }}$ repetition) | 997 vs. 887 |
| Deviants after 5 standards vs. deviants after 4 repetitions | 995 vs. 994 |
| Deviants after 5-7 standards vs. deviants after 1-4 repetitions | 99x vs. 99y |
| Predictive Surprise (PS) | 72 vs. 71 |
| Confidence-corrected surprise (CS) | 82 vs. 81 |
| Bayesian Surprise (BS) | 92 vs. 91 |
| HGF_Vanilla (HGF_Pitch) | PE2: 12 vs. 11 (16 vs. 15) |
|  | PE3: 22 vs. 21 (24 vs. 23) |
|  | PWPE2: 32 vs. 31 ( 34 vs. 33) |
|  | PWPE3: 42 vs. 41 (44 vs. 43) |

## Appendix B

## HGF: Update Equations

Figure B46
The (vanilla) HGF's update equations (Lomakina, 2014)
The resulting update equations are not only efficient and easy to compute but also resemble results from the field of reinforcement learning.

Update equations for expectations

| Level 3 | $\Delta \mu_{3}=\sigma_{3} \cdot \frac{\kappa}{2} \cdot w_{2} \cdot \delta_{2}$ <br> Expectation update (Unweighted) learning rate Weighting factor | with | $\begin{gathered} \Delta \mu_{3}=\mu_{3}^{(k)}-\mu_{3}^{(k-1)} \\ \sigma_{3}=\sigma_{3}^{(k)} \\ w_{2}=\frac{\mathrm{e}^{\alpha \mu_{3}^{(k-1)}+\omega}}{\sigma_{2}^{(k-1)}+\mathrm{e}^{\kappa \mu_{3}^{(k-1)}+\omega}} \\ \delta_{2}=\frac{\sigma_{2}^{(k)}+\left(\mu_{2}^{(k)}-\mu_{2}^{(k-1)}\right)^{2}}{\sigma_{2}^{(k-1)}+\mathrm{e}^{\alpha \mu_{3}^{(k-1)}+\omega}}-1 \end{gathered}$ |
| :---: | :---: | :---: | :---: |
| Level 2 | $\Delta \mu_{2}=\sigma_{2} \cdot \delta_{1}$ | with | $\begin{gathered} \Delta \mu_{2}=\mu_{2}^{(k)}-\mu_{2}^{(k-1)} \\ \sigma_{2}=\sigma_{2}^{(k)} \\ \delta_{1}=\mu_{1}^{(k)}-s\left(\mu_{2}^{(k-1)}\right) \end{gathered}$ |

Rescorla-Wagner model: prediction ${ }^{(k)}$ - prediction ${ }^{(k-1)}=$ learning rate x prediction error
(i) $\kappa$ determines the degree of coupling between the second and third level in the hierarchy ( $x_{2}$ and $x_{3}$ )
(ii) $\omega$ represents a constant (tonic) component of the log-volatility of $x_{2}$, capturing the subjectspecific magnitude of the belief update about the stimulus-outcome probabilities that is independent of $x_{3}$ (i.e., regardless of how volatile the environment is).
(iii) $\vartheta$ is a meta-volatility parameter and determines the evolution of $x_{3}$, or how rapidly the volatility of the associations changes in time. I used the default $\mu_{3(\mathrm{k}=0)}$, the subject's initial belief about volatility of the outcome probabilities.

## Appendix C

## HGF: Layers of the model

Looking more closely at how the outputs on each layer differ between HGF_Vanilla and HGF_Pitch, two observations emerge. Firstly, as seen by comparing the top vs. bottom rows of Table C5, correlations at the same level intra-model (i.e., between PE2/PWPE2 and PE3/PWPE3), again appear to be higher on high expected surprise trials than on all other trials (the statistical significance of this comparison was not however tested). Secondly, as seen when comparing the left vs. the right-hand side of the table, the HGF_Pitch model tends to produce the same or lower correlations, especially on deviants:

## Table C5

The correlations between the outputs of HGF_Vanilla and HGF_Pitch models
On all trials:

| HGF_Vanilla | HGF_Pitch |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | PE3 | PWPE2 | PWPE3 |  | PE3 | PWPE2 | PWPE3 |
| PE2 | 0.48 | 0.73 |  | PE2 | 0.47 | 0.71 |  |
| PE3 |  | 0.72 | 0.31 | PE3 |  | 0.71 | 0.29 |
| PWPE2 |  |  | 0.45 | PWPE2 |  |  | 0.41 |

On deviants only:

| HGF_Vanilla | HGF_Pitch |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | PE3 | PWPE2 | PWPE3 |  | PE3 | PWPE2 | PWPE3 |
| PE2 | 0.64 | 0.79 |  | PE2 | 0.55 | 0.78 |  |
| PE3 |  | 0.33 | 0.46 | PE3 |  | 0.16 | 0.38 |
| PWPE2 |  |  | 0.55 | PWPE2 |  |  | 0.6 |

[^0]
## Appendix D

## The HGF_Pitch model

The spreadsheet used to build this is available here

The formula used to modify PE2 in the HGF_Pitch model is shown (for A, B in range [1,N]).

$$
\text { multiplier }_{\text {tone }_{A \rightarrow B}}=\ln \left(\frac{\max \left(c, \mid \text { perceptual } \operatorname{gap}_{A \rightarrow B} \mid\right)}{\mid \text { perceptual } \operatorname{gap}_{A \rightarrow B} \mid}\right)+d
$$

Where:

- perceptual gap $_{1 \rightarrow 1}=0$ (from $500 \mathrm{~Hz}->500 \mathrm{~Hz}$ )
- perceptual gap $_{1 \rightarrow 2}=1$ (from $500 \mathrm{~Hz}->550 \mathrm{~Hz}$ )
- perceptual $\operatorname{gap}_{X \rightarrow X+1}=e * \frac{1}{X-1}+$ perceptual gap $p_{(X-1) \rightarrow X}+(1-e)$ for $2 \leq X<N$
$N=7$, the number of unique sound tones played
$c$ is set $>0$ to produce a non-erroneous logarithm for transitions to the same tone, e.g., as:

$$
\text { perceptual gap }{ }_{500 \mathrm{~Hz} \rightarrow 500 \mathrm{~Hz}}=0
$$

d was set in conjunction with c to produce multipliers which increase smoothly with perceptual gap. $\boldsymbol{e}$ controls to what extent this increase is logarithmic; the remaining weight (1-e) controls to what extent the perceptual gaps are assumed to be linear (as was done by e.g., Lieder et al., 2013).

The multiplier is then exponentiated by a power factor $\boldsymbol{f}$ (which decreases multipliers $>1$ and increases multipliers $<1$ ) and multiplied by a constant $\boldsymbol{g}$ to yield an average multiplier of 1 .

$$
\text { multiplier }_{\text {tone }_{A \rightarrow B}}=\left(\text { multiplier }_{\text {tone }_{A \rightarrow B}}\right)^{f} * g
$$

Only a single value was tested for each constant: $\mathrm{c}=0.5, \mathrm{~d}=2.5, \mathrm{e}=0.75, \mathrm{f}=0.5, \mathrm{~g}=0.88$.

## Appendix E

## HGF: Parameter Estimates and fit

The parameters, and metrics of fit are shown in Figure E47 and Figure E48:

1) For the 346 -tone sequence:

Figure E47
Parameter estimates and model quality metrics for the 346-tone sequence for HGF_Vanilla (top) and HGF_Pitch (bottom)

```
Parameter estimates for the perceptual model:
    mu2_0: [ [-1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918
```



```
    mu3_0: 1.0755
    sa3_0: 0.1051
        ka: 2.3443
        om: -5.9013
        th: 0.0731
Model quality:
        LME (more is better): -296.4178
    AIC (less is better): 596.6597
    BIC (less is better): 615.8774
    AIC and BIC are approximations to -2*LME = 592.8356.
Parameter estimates for the perceptual model:
    mu2_0: [-1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918
    sa2_0: [1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1]
    mu3_0: 1.0691
    sa3_0: 0.1049
            ka: 2.3402
            om: -5.8396
            th: 0.0736
Model quality:
    LME (more is better): -294.647
    AIC (less is better): 593.0938
    BIC (less is better): 612.3115
    AIC and BIC are approximations to -2*LME = 589.2939.
```

2) For the 2000-tone sequence:

Figure E48
Parameter estimates and model quality metrics for the 2000-tone sequence for HGF_Vanilla (top) and HGF_Pitch (bottom)

```
Parameter estimates for the perceptual model:
    mu2_0: [-1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918 -1.7918
```



```
    mu3_0: 0.9976
    sa3_0: 0.0999
        ka: 2.2150
        om: -14.7508
        th: 0.0448
Model quality:
    LME (more is better): -4222.2737
    AIC (less is better): 8450.4851
    BIC (less is better): 8483.2421
    AIC and BIC are approximations to -2*LME = 8444.5474.
```

```
Parameter estimates for the perceptual model:
    mu2_0: [lll.7918 -1.7918 -1.7918 -1.7918 -1.7918
    sa2_0: [1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1]
    mu3_0: 0.9682
    sa3_0: 0.0999
        ka: 2.2298
        om: -14.9904
        th: 0.0458
Model quality:
    LME (more is better): -4202.0893
    AIC (less is better): 8410.3349
    BIC (less is better): 8443.0919
    AIC and BIC are approximations to -2*LME = 8404.1786.
```

Whilst the BIC measure is only strictly interpretable relative to other models' BICs, they appear to indicate a poor model fit relative to when I trained the HGF on simple patterns (e.g., deterministic sequences or those with trend lines).

## Appendix F <br> Spatial autocorrelation of MEG data

As seen in Figure F49 below, the spatial autocorrelation for children (top row) and adults (bottom row) was significantly positive for all participants (red lines represent quadratic fits for each participant). This invalidated the assumption of independence between channels, justifying the statistical approach of spatiotemporal clustering.

Figure F49
Spatial autocorrelation between MEG sensors in the child system (Panel A) and the adult system (Panel B)


## Appendix G

## Gross activity patterns

Two distinct peaks in gross activity were present for each condition (low vs high surprise trials), as seen in Figure G50 below. At the latter peak, the difference in GFP nearly between conditions breached the confidence interval.

## Figure G50

GFPs for low surprise (blue line) and high surprise (orange line) conditions (top) and their differences (bottom)



Notes: Top: In blue: Condition A (low surprise trials). Shaded: GFPs of all persons, with $20 \%$ outliers removed. In orange: Condition B (high surprise trials)
Bottom: Condition B minus Condition A. Shaded: Confidence interval, with $20 \%$ of outliers removed. The bounds of the confidence interval approach zero at 235 ms , indicating the GFP MMR is significantly different to zero.

The difference between the GFPs was not significant at either peak, as seen in Figure G51:

## Figure G51

The comparison between the GFP in high minus low surprise trials did not show any significant differences


The presence of two peaks reflecting processing of the auditory tones is confirmed by plotting the ERF per condition topographically, as seen in Figure G52. Note the changed polarity in the ERFs that was found just prior to 180 ms (which is in between the peaks in GFP shown in Figure 29):

Figure G52
ERF topographic plots for standards (top row) and deviants (bottom row).


Note: No right-lateralisation was found, in contrast to that found by (Giard et al., 1990).

As seen in Figure G53, there was greater activity in high expected surprise trials (condition B), e.g., in the youngest $20 \%$ :

Figure G53
The response of young people appeared greater (see scale) for high than low PE2 at times of significant activity


Note: The bottom row represents high expected surprise conditions and shows greater activity than the top row (the low expected surprise condition).

As seen in Figure G54, the difference between conditions can also be seen by viewing the aMMR directly:

Figure G54

The difference in ERFs between conditions, across all participants


As seen in Figure G55, activity reduced as the sounds receded into the past within an Epoch, especially for tones with lower predicted surprise (condition A, shown in blue):

Figure G55
GFPs in low/high surprise conditions (blue and orange) as the sound event recedes into the past


As seen in Figure G56, this was visible from ERFs that moved toward a zero average in both conditions, too (note the rest of the graphs in this appendix are generated from unstandardised data).

Figure G56
ERF magnitudes for low (top row) and high (bottom row) surprise trials reduced as the sound receded into the past.


Caution is warranted, however, before concluding that different conditions (or, as shown later, between age groups) are indistinguishable at high latency. For example, at 395 ms , whilst gross activity is diminished in both conditions, differences between the conditions (as shown in the aMMR in Figure G57) remain and may even be stronger than at earlier latencies:

Figure G57
Differences between low and high surprise conditions at longer latencies


The gross patterns described so far appear consistent with the sound stimulus progressing up a neural hierarchy where it is processed at various stages. Whilst a reduction in brain activity as the sound recedes is expected during this process, it might be expected that there should be an ERF at the time of the tone presentation (i.e., at a lower latency that is needed for bottom-up processing), as a result of anticipatory brain activity. If this is the case then, as shown in Figure G58, it is not visible immediately before the tone but occurs at or immediately after its intensity rises $(0-0.01 \mathrm{~s})$ whilst it is plays $(0-0.07 \mathrm{~s})$ :

## Figure G58

If anticipatory brain activity occurs, the timing of this does not precede the tone:


Note: Condition A (low surprise trials; top row) and Condition B (high surprise; bottom row), unstandardised data.

This does not strongly argue against the existence of model-driven gross anticipatory activity (including activity earlier than at 0 ms ) occurring in the brain. Indeed, GFP is not a minimum at the time of tone onset, as shown in Figure G59:

Figure G59
The GFP appears to be rising already even as the tone is played (and before it is processed)


Note: Unstandardised data.
It does however suggest that any such top-down predictive activity is not as strong as in response to the tone and that model modifications influence surprise rather than manifesting through brain activity indicating a strong expectation. However, activity at 0 s does appear in individual participants' data, and was not attributable to trigger leakage (see Appendix $H$ Trigger Leakage). Moreover, it may be relatively more pronounced in the oldest individuals (if not as an absolute score) as seen by the final row in Figure G60 below:

Figure G60
ERFs in the youngest $20 \%$ (top) / 40\% (bottom) (Group A, Panel A) and oldest $20 \% / 40 \%$ of people (Group B, Panel B)
A


- B

fT $\begin{array}{r}\text {-0000 } \\ -20000 \\ -20000 \\ -0000 \\ -30000 \\ -20000 \\ -30000\end{array}$


## Appendix H

## Trigger leakage evaluation

There was no evidence of trigger leakage (an artefact induced in a MEG sensor coil due to an electric pulse sent via parallel port to the MEG system by MATLAB to record an event when a tone is played), with most channels having a value of $\sim 0$ at the time of the MEG trigger, as seen in Figure G61:

Figure G61


## Appendix I

## Standardisation of data

It was thought that standardising (z-transforming) continuous MEG data would be useful for three reasons. Firstly, standardising will remove any low-frequency drifts remaining after ICA, which is important to do as any DC bias remaining in the data will make the GFP uninterpretable. Secondly, standardising should reduce inter-individual differences and differences between age groups by extracting a signal-to-noise ratio. For example, some participants laid further away from the MEG helmet, which would result in their brain regions of interest (ROIs) being further from the relevant MEG sensor(s), reducing signal power for all of their trials. Thirdly, standardising would adjust for the brain to helmet size ratio, as superficial ROIs are further away from the sensors when this is low e.g., in the youngest participants scanned in the adult system. Overall, standardisation should thus especially assist in the comparison of data sets taken from the child and adult systems and consequently, to discover condition-by-age interactions.

When comparing standardised vs. unstandardised data in Figure I62, several findings emerged. Firstly, standardisation reduced but did not eliminate differences between age groups in GFPs:

Figure I62
The differences in GFP (young minus old) between the youngest and oldest 20\% (top row) and 10\% (bottom row)


Notes: Unstandardised data (at left) and standardised data (at right) for all trials.
Associated t -values are shown in the second row of each graph with significant differences ( $p<0.05$ ) shown in pink.

This manifested as a shorter range of times at which there were significant differences between age groups in the standardised data. For example, in the 20/20 split, using unstandardised data there was a significant difference in the GFPs between age groups throughout the period of $55-295 \mathrm{~ms}$ $(t(28)=75.16, p<0.001, d=3.07)$ however in the standardised data there was not a significant difference between the groups in the windows spanning 55-85ms $(t(28)=3.23, p>$ $0.05)$ or $150-215 \mathrm{~ms}(t(28)=2.6, p>0.05)$. Interestingly, the opposite occurred early in the Epoch (at $0-55 \mathrm{~ms}$ ). In the $20 / 20$ split, there was a significant difference between the GFPs using standardised data $(t(28)=-23.82, p<0.001, d=-1.75)$ but there was no effect for unstandardised data $(t(28)=1.17, p>0.05)$.

Secondly, as seen in Figure I63, the same difference between age groups was evident in the GFP difference between conditions. For example, in the window $95-135 \mathrm{~ms}$, for unstandardised data there was a significant temporal cluster $(t(28)=8.41, p=0.028, d=1.1)$ but not for standardised data $(t(28)$ $=3.02, p>0.05)$ :

Figure I63
The difference in GFP between conditions (for the youngest 20\% minus the oldest 20\%) for unstandardised (left) and standardised (right) data.


Note: The difference in brain activity (i.e., a larger GFP for surprising compared to unsurprising trials) was greater at for younger people earlier on (e.g., at $\sim 100 \mathrm{~ms}$ ), and was greater for older people later in the Epoch (e.g., between 200-250ms).

Thirdly, standardisation generally did not alter the broad spatiotemporal pattern of ERFs but occasionally boosted the contrast between conditions as measured by the GFP aMMR. As can be seen in the topographic plots of Figure I64 below, the spatiotemporal pattern of ERFs was not easily visually easily distinguished based on whether the data had been standardised or not. Rather, standardisation looked like a scaling operation within-subject, indicating the ERF had a zero mean.

## Figure 164

Examples of unstandardised (top row) and standardised and scaled (bottom row) data for low surprise trials


Notes: The unstandardised data (top row) resembles the standardised data (bottom row) in its spatial distribution The magnitudes are not comparable as standardised data was multiplied by an arbitrary constant.

The apparent similarity between the datasets was preserved when taking the difference between conditions, as shown in Figure I65:

Figure I65
The aMMR ERF patterns for unstandardised (top row) and standardised (bottom row) data


However, as seen in Figure I66, the GFP $a M M R$ between the conditions (across all ages) was significant using the focused TOI: the standardised comparison was significantly different (during $155 \mathrm{~ms}-180 \mathrm{~ms}, t(71)=6.44, p=0.023, d=0.42$ ):

Figure 166
The GFP aMMRs between conditions using unstandardised (left) and standardised (right) data


Indeed, when comparing age groups, the GFP aMMRs seen above in Figure I66 can be seen to be driven by different magnitudes in the ERFs, even if the spatial distribution of the aMMR is again seen to be similar, as shown in Figure I67:

Figure 167
The group difference in the aMMR of ERFs for unstandardised (top row), standardised (bottom row) data.


Given the apparent qualitative and quantitative differences between unstandardised and standardised data discovered and the theoretical advantages of standardised data discussed, all data analysed consists of standardised data only unless otherwise specified.

## Appendix J

## Control Conditions

There are controls conditions I expected would produce a pattern of results which could not be differentiated statistically. This absence of differentiability turned out to be the case.

600 Hz vs. 700 Hz

Specifically, the GFP vs. time graphs appeared nearly identical between conditions, as seen in Figure J68 below. Whilst this does not amount to evidence of their equivalence, a permutation cluster test on these GFPs did not reveal any differences in brain activity ( $p>0.05$ ):

Figure J68
GFPs for a tone 50 Hz below (blue) or above (orange) the average tone ( 650 Hz ), and their differences


Additionally, as see in Figure J69, no spatiotemporal clusters of any significant differences in the ERFs were found:

Figure J69
ERFs for 600 Hz tones ( $1^{\text {st }}$ row), for 700 Hz tones ( $2^{\text {nd }}$ row) tones and the difference ( $3^{\text {rd }}$ row)


Notes: This plot uses unstandardised data.
No spatiotemporal clusters of ERF difference were found to significantly differ from zero

Finally, the (Pearson) correlation between the predictor (codes as 0 for 600 Hz , and 1 for 700 Hz ) and the GFP was not significantly different from zero at any time point, as seen in Figure J70:

Figure J70
There was no latency at which the physical tone frequency (binarized as 0 or 1) had a correlation with the GFP which was significantly different to zero $(\mathrm{p}>0.05)$


Nor was the correlation to the ERF significantly different from zero at any time point, as seen in Figure J71:

Figure J71
There was no latency at which the physical tone frequency (binarized as 0 or 1) had a correlation with the ERF which was significantly different to zero ( $\mathrm{p}>0.05$ )
-_Grand average $(\mathrm{n}=12)$


As shown in Figure J72, the GFP aMMR for 650 Hz and 800 Hz tones were not statistically distinguishable,

Figure J72
The GFP for the pitch outlier of 800 Hz (Condition B) compared to the GFP for the average tone ( 650 Hz )


Note: The GFPs were not significantly different at any time at any time

As seen in Figure J73, the ERFs again appeared similar (top two rows) and the difference (bottom row) was not significantly different in any spatiotemporal region:

Figure J73
The ERFs for 650 Hz (top row) and $800 H z$ (2 $2^{\text {nd }}$ row) tones, and the difference ( $3^{\text {rd }}$ row).


Note: The data shown are unstandardised.

As seen in Figure J74, the ERF difference in the control contrast was not significant in the restricted TOI either:

Figure J74
The ERFs for condition A ( 650 Hz , top row), condition B ( 800 Hz , middle row) and the difference (bottom row)

0.150 s
0.170 s


Note: The difference (bottom row) was not significantly different at any spatiotemporal cluster.

## Appendix K

## Ethics Approval

Adult Ethics approval

Office of the Deputy Vice-Chancellor (Research)

MACQUARIE
University
Research Services
Research Hub, Building C5C Eas
Research Hub, Buil
Macquarie University
NSW 2109 Australia
T: +61(2) 98504459
htps $/ / M m w$ w...q.e.edu auresear
$\xrightarrow{\text { ABN } 90} 9552801237$

16 May 2018

Dear Associate Professor Sowman

Reference No: 5201800226
Title: Magnetoencephalography and combined Magnetoencephalography with Electroencephalography studies of Adult Cognition

Thank you for submitting the above application for ethical and scientific review. Your application was considered by the Macquarie University Human Research Ethics Committee (HREC (Medical Sciences))

I am pleased to advise that ethical and scientific approval has been granted for this project to be conducted at:

- Macquarie University

Date Approved: 07 May 2018
This research meets the requirements set out in the National Statement on Ethical Conduct in Human Research (2007 - Updated May 2015) (the National Statement).

## Standard Conditions of Approval

1. Approval is contingent on continuing compliance with the requirements of the National Statement, which is available at the following website:
http://www.nhmrc.gov.au/book/national-statement-ethical-conduct-human-research
2. This approval is valid for five (5) years, subject to the submission of annual reports. Please submit your reports on the anniversary of the approval for this protocol.
3. Proposed changes to the protocol and associated documents must be submitted to the Committee for approval before implementation

It is the responsibility of the Chief investigator to retain a copy of all documentation related to this project and to forward a copy of this approval letter to all personnel listed on the project.

Should you have any queries regarding your project, please contact the Ethics Secretariat on 98504194 or by email ethics.secretariat@mq.edu.au

The HREC (Medical Sciences) Terms of Reference and Standard Operating Procedures are available from the Research Office website at:
https://www.mq.edu.au/research/ethics-integrity-and-policies/ethics/human-ethics
The HREC (Medical Sciences) wishes you every success in your research

Yours sincerely

Professor Tony Eyers
Chair, Macquarie University Human Research Ethics Committee (Medical Sciences)

This HREC is constituted and operates in accordance with the National Health and Medical Research Council's (NHMRC) National Statement on Ethical Conduct in Human Research (2007) and the CPMP/ICH Note for Guidance on Good Clinical Practice.

Child ethics approval:

```
Office of the Deputy Vice-Chancello
(Research)
Research Office
Research Office
Macquarie University
\(\mathrm{T}:+61\) (2) 98504459
nitp \(/ / 1\) mum research mq edu au
```

MACQUARIE
University
sydney-australia

19 May 2016

Dear Dr Brock
Reference No: 5201600188
Title: MEG studies of child brain development
Thank you for submitting the above application for ethical and scientific review. Your application was considered by the Macquarie University Human Research Ethics Committee (HREC (Medical Sciences))

I am pleased to advise that ethical and scientific approval has been granted for this project to be conducted at:

- Macquarie University

This research meets the requirements set out in the National Statement on Ethical Conduct in Human Research (2007 - Updated May 2015) (the National Statement).

## Standard Conditions of Approval:

1. Continuing compliance with the requirements of the National Statement, which is available at the following website:
http://www.nhmrc.gov.au/book/national-statement-ethical-conduct-human-research
2. This approval is valid for five (5) years, subject to the submission of annual reports. Please submit your reports on the anniversary of the approval for this protocol.
3. All adverse events, including events which might affect the continued ethical and scientific acceptability of the project, must be reported to the HREC within 72 hours.
4. Proposed changes to the protocol and associated documents must be submitted to the Committee for approval before implementation

It is the responsibility of the Chief investigator to retain a copy of all documentation related to this project and to forward a copy of this approval letter to all personnel listed on the project.

Should you have any queries regarding your project, please contact the Ethics Secretariat on 98504194 or by email ethics.secretariat@mq.edu.au

The HREC (Medical Sciences) Terms of Reference and Standard Operating Procedures are available from the Research Office website at:
http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human research ethics

The HREC (Medical Sciences) wishes you every success in your research

Yours sincerely

Professor Tony Eyers
Chair, Macquarie University Human Research Ethics Committee (Medical Sciences)

This HREC is constituted and operates in accordance with the National Health and Medical Research Council's (NHMRC) National Statement on Ethical Conduct in Human Research (2007) and the CPMP/ICH Note for Guidance on Good Clinical Practice.

## Review references

## Adult:

Magnetoencephalography and combined Magnetoencephalography with Electroencephalography studies of Adult Cognition

Note: There is a newer version of the project. (Please contact the project owner to update this form).


Child:
MEG studies of child brain development
Note: There is a newer version of the project (Please contact the project owner to update this form).

| Project Tree |  |  |  |
| :---: | :---: | :---: | :---: |
| - MEG studies of child drain development |  |  |  |
| Human Eilics Data Transfer Form v1. 1 |  |  |  |
| Action Required on Form | Status | Review Reference | Date Modified |
| No | HE-Approved | 52021190235599 | 02/12/2021 09:24 |

## Appendix L

## Child vs. Adult MEG systems

As a form of verification of the sensibility of the mapping of adult to child sensors, I recorded empty-room data simultaneously using both child and adult MEG systems (from slightly different start times). Two channels which had been mapped to each other (MEG004 in the child system and MEG007 in the adult system) were compared, as shown in Figure L75. The data produced by the child system was shifted temporally until the correlation of the two sources was maximised (at $\sim 0.96$ ): this strong correlation is suggestive of high signal-to-noise ratios in both machines, facilitating comparison between age groups.

Figure L75
The line of best fit (and correlation) between data from similarly-aligned MEG sensors in child and adult systems.


[^1]Figure L76
The location of all sensors in the adult system (left), and of those ultimately paired with the child sensors (right)

Adult system


Sparse adult system



[^0]:    Notes: Correlations are displayed at various layers of the hierarchy.
    Correlations are shown for all trials (top row) as well as for deviants only (bottom row).

[^1]:    Notes: The mapped channels shown were MEG004 in the child system and MEG007 in the adult system. Correlation is low when the data are not synchronised (red at left, and yellow at centre).
    Correlation is strong (peaking at $\sim 0.96$ ) when the sources are correctly aligned temporally (green, right).

