Hierarchical Learning of Statistical Regularities over Multiple Timescales of Sound Sequence Processing: A Dynamic Causal Modeling Study

Kaitlin Fitzgerald¹, Ryszard Auksztulewicz², Alexander Provost¹, Bryan Paton¹,³, Zachary Howard¹, and Juanita Todd¹,³

Abstract

Our understanding of the sensory environment is contextualized on the basis of prior experience. Measurement of auditory ERPs provides insight into automatic processes that contextualize the relevance of sound as a function of how sequences change over time. However, task-independent exposure to sound has revealed that strong first impressions exert a lasting impact on how the relevance of sound is contextualized. Dynamic causal modeling was applied to auditory ERPs collected during presentation of alternating pattern sequences. A local regularity (a rare \( p = .125 \) vs. common \( p = .875 \) sound) alternated to create a longer timescale regularity (sound probabilities alternated regularly creating a predictable block length), and the longer timescale regularity changed halfway through the sequence (the regular block length became shorter or longer). Predictions should be revised for local patterns when blocks alternated and for longer patterning when the block length changed. Dynamic causal modeling revealed an overall higher precision for the error signal to the rare sound in the first block type, consistent with the first impression. The connectivity changes in response to errors within the underlying neural network were also different for the two blocks with significantly more revision of predictions in the arrangement that violated the first impression. Furthermore, the effects of block length change suggested errors within the first block type exerted more influence on the updating of longer timescale predictions. These observations support the hypothesis that automatic sequential learning creates a high-precision context (first impression) that impacts learning rates and updates to those learning rates when predictions arising from that context are violated. The results further evidence automatic pattern learning over multiple timescales simultaneously, even during task-independent passive exposure to sound.

INTRODUCTION

Time is a very important contextual factor in determining how we make sense of the world around us. When determining whether or not a given event is surprising, we call upon past experience as a reference. However, selecting the right reference period is key. One principle to apply in determining the reference period is to employ that which will minimize surprise (Friston, 2005). In doing so, behavior is thought to be optimized by determining what learning might be necessary in this instance—what is new that should be used to update our prior knowledge, and, conversely, what is already known that can be largely ignored while pursuing other aspects of experience that do provide new information. In other words, a process that requires a memory search for when, where, and how this event was encountered in the past. This contextualization of learning can be experienced as a “conscious” deliberate process, but also occurs automatically in many instances such as perceptual inference (the use of patterns in sensory input to anticipate future experience; Hohwy, 2012; Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001). In this paper, a computational modeling approach is applied to test the hypothesis that first impressions of event likelihood exert long-lasting influences over the future estimates of event likelihood when the environment changes. These first impressions appear to change the reference for determination of likelihood, such that events that should be equally surprising on a local timescale are actually treated differently.

Automatic determinations of how surprising an event is are often studied in laboratories using sequential learning (Kujala, Tervaniemi, & Schroger, 2007). The particular learning environment in this case is passive (task-independent) exposure to sequences of sound containing patterning. Using changes in the response to sound over time, it is possible to infer both the learning of regularities within a patterned sequence, and the surprise that occurs when an event violates predictions that arise from the learned pattern (Baldeweg, 2007). Early work in this field revealed the reference period to be determined by the contents of auditory sensory memory, which is generally considered to be confined to an upper...
limit of around 30 sec (Näätänen, 1992; Näätänen, Paavilainen, Alho, Reinvikainen, & Sams, 1989). If this were the case, responses that index learning and surprise should be updated very dynamically. A good example of this dynamic responsiveness can be observed from sound sequences called “roving standard paradigms” (Garrido et al., 2008; Baldeweg, 2007; Haenschel, Vernon, Dwivedi, Gruzeller, & Baldeweg, 2005). In a roving standard paradigm, a repeating string of sounds of one frequency is suddenly interrupted by a sound that changes in frequency, and the new frequency subsequently starts repeating. The sounds that represent the first change or pattern interruption are called deviants, and the repetitions are called standards. By averaging over many instances of deviants, and subsets of the standards, the roving paradigm has demonstrated that longer sequences of repetition produce more positive going or suppressed standard responses and lead to higher levels of negativity to the deviations. That is, the ERP changes from having clear morphology with a negative peak at 100 msec (N1), positive peak around 200 msec (P2), and a negative peak around 250 msec (N2) to having these components imposed over a positive shift in the ERP. The negativity to deviations in the window of 100–250 msec after the sound deviation onset contains the output of a neuronal comparator process referred to as MMN, the amplitude of which is prosed to index the degree of surprise in relation to the deviation (Näätänen, Jacobsen, & Winkler, 2005).

Based on the above observations and the proposed reliance on auditory sensory memory, the degree of surprise evident in response to a deviant sound should be determined in reference to a recent past and should be dynamically updated when a regularity changes (as observed for the roving standard paradigm). However, this is not the case in results obtained from sequences in which two sounds alternate as the repeating and rare event (Todd, Provost, Whitson, Cooper, & Heathcote, 2013; Todd, Provost, & Cooper, 2011). Indeed, the degree of surprise evident to a locally rare pattern deviation appears to be unduly influenced by the impressions formed at the sequence onset. Assuming MMN amplitude to be an index of surprise, results show that we are quickly surprised by events that are rare during early learning but take longer to be surprised by events that are initially common and become rare later. These differences occur despite the equivalence of probabilistic information about these sounds on a local timescale.

Examples of this distinctive variation in surprise are presented in Figure 1. In this figure, the two different block colors represent periods in which the same two sounds have opposite probabilities; for example, where the two sounds are a longer and shorter sound, a longer sound is rare in the dark blocks and a shorter sound is rare in the white blocks. There is a high level of surprise (large MMN amplitude) to rare sounds at the beginning of blocks matching how the sequence starts (hereafter called first deviants), and this MMN amplitude stays large during the blocks. In contrast, MMN is smaller initially for the sounds that are first common then become rare (hereafter called alternate deviants) in the second block type (see Figure 1). Although the MMN amplitude to the alternate deviant typically increases over the duration of the block to reach an equivalent amplitude to that of the first deviant in the latter half of blocks, it is reliably and significantly smaller at the start of blocks relative to the first deviant (Mullens et al., 2016; Todd, Whitson, et al., 2014). Furthermore, Figure 1 left shows that this is true whether the sequence starts with a rare longer sound among shorter sounds (Pictorial Depiction 1) or a rare shorter sound among longer sounds (Pictorial Depiction 2), so it is clearly order-dependent and not tied to tone properties. Based on the way MMN amplitude is weighted by stability in the environment (Lieder, Stephan, Daunizeau, Garrido, & Friston, 2013; Garrido, Kilner, Stephan, & Friston, 2009), the amplitude of MMN should increase as the block composition remains stable. This appears to occur more
rapidly for the first deviant than the alternate deviant despite equivalence in local probability. However, this difference between the first and alternate rare sounds reverses when a second sequence with a different regular block length occurs. Figure 1 right shows how MMN amplitude is now small at the beginning of blocks that are consistent with how the sequences starts but is much larger for the alternate deviant. As a result, only deviants within the first block type show a pattern of amplitude modulation that we might expect based on block length (i.e., MMN is larger is longer block than shorter block sequences for the first deviant only (Todd, Provost, Whitson, & Mullens, 2018; Todd, Heathcote, et al., 2014; Todd et al., 2013).

In every case, the group mean amplitudes that are displayed in Figure 1 represent a response to a sound that has a local probability of \( p = .125 \), and yet the degree of surprise, as indexed by MMN amplitude, is systematically different as a function of order effects. In fact, the amplitudes appear to bare little relationship to local stability with the alternate deviants in early periods of the short blocks (deviants within the first 0.4 min from the six blocks) actually eliciting a larger amplitude MMN than early period of the longer blocks (deviants within the first 1.2 min from each of the two blocks). These MMN amplitude variations can therefore not be dependent on a reference to the contents of sensory memory alone and do not bear any simple relationship to probability or stability period in general.

How can these differences in the weighting of surprise be explained? In this paper, an analysis technique known as dynamic causal modeling (DCM) is used to test the hypothesis that the variation in MMN amplitude is actually tied to differences in the precision associated with predictions in the first and alternate block types of a sequence, and how those precisions are affected when the block lengths suddenly change (Garrido et al., 2008; Stephan et al., 2007; David et al., 2006). This hypothesis is based on the assumption that the patterned learning during exposure to these sequences is not just the local regularity but also the regularity in block lengths. The patterns in MMN amplitude in Figure 1 do not occur if the same blocks of sound are presented with no regularity in the block length (Todd, Petherbridge, Speirs, Provost, & Paton, 2018) or if participants have been told and shown images of the sequence composition (Frost, Haasnoot, McDonnell, Winkler, & Todd, 2018). Furthermore, Mullens et al. (2016) found that the patterns observed in Figure 1 depend on the order of the sequences too. By testing participants with either the longer block sequence first (as per Figure 1) or the shorter block sequence first (reverse of Figure 1), Mullens et al. (2016) found that the first deviant MMN was significantly smaller for blocks in the second sequence heard independent of whether the second sequence was a longer block or shorter block sequence. In other words, the results appear to be reliant on the presence of regular block structure and a sequential experience of the sound patterns.

DCM is applied here to EEG data obtained during exposure to sound sequences to identify neurobiologically plausible mechanisms at the source level (i.e., regions within the brain) that explain the time course and scalp topography of responses to the sounds (David et al., 2006). In DCM, the brain is assumed to differentiate the activation states induced by the sounds through a temporal hierarchy where different brain regions are sensitive to dynamics at different temporal scales. At the lowest level, sensory cortices encode faster timescale dynamics underlying simple sensory processing, whereas the highest level involves the pFC engaging the more complex functions required to represent slower-changing environmental states (Hasson, Yang, Vallines, Heeger, & Rubin, 2008; Kiebel, Daunizeau, & Friston, 2008). These hierarchical models describe the neural dynamics implemented in canonical microcircuits, which are compatible with predictive coding (i.e., a theory that poses that the brain is constantly generating and updating internal models of the environment; Friston, 2005). Each processing point in the network undertakes a comparison between the state predicted on the basis of regularities (reflected in activity communicated from higher to lower brain regions), with the actual state resulting from sensory input (reflected in activity communicated from the lower to higher network regions). This comparison will, for example, result in current input either being largely explained in the case of a highly repeated common sound or generating a large “prediction error” in the case of a rare deviant. Errors drive for an update to the existing predictions in order to iteratively improve predictions to minimize surprise. Under predictive coding accounts, the MMN component evident in responses to deviant sounds is purported to be an example of a precision-weighted error signal where the amplitude will be inversely proportional to the level of variance in the sensory input (Lieder et al., 2013; Friston, 2005). Precision-weighting of errors occurs at each point in the network, which reflects how reliable predictions are (inversely how surprising errors are), which in turn is a function of relative stability (higher precision) or volatility (lower precision) in the patterns of sound.

DCM is used in this study to test specific assumptions about how the order effects on MMN amplitude in sequences like those in Figure 1 are reflected in the underlying network. It has elsewhere been hypothesized that the sequences enable us to see the effects of hierarchical learning (simultaneously learning on local and longer-term timescales) by creating two key points of surprise that have a strong impact on precision-weighting and model updating (Todd, Petherbridge, et al., 2018; Mullens et al., 2016; Todd et al., 2015; Frost, Winkler, Provost, & Todd, 2015; Todd et al., 2013). The first is at the juncture between the first two block types when the sound that is first rare starts to repeat generating a sudden increase in the rate of prediction errors. This could be considered a local-level or shorter timescale change where the current internal model begins to fail quite
dramatically. The frequent errors would initially reduce model certainty triggering remodeling in the effort to explain these errors moving the model into the direction of the properties of the tone that is now frequent in this context (Winkler, Karmos, & Näätänen, 1996). The second major point of surprise would occur when the first block type violates block length predictions. Assuming the brain learns the block length via the changing rate of prediction errors, an unexpected early change in error density (when longer blocks become shorter) or a delayed change (when shorter blocks become longer) should trigger an update to predictions about the rates at which error densities change. In each case, these points of surprise (unexpected change in a given regularity) should reduce model precision and increase learning rates (Yu & Dayan, 2002b), however, the effect of the surprise should impact models at different timescales.

Hierarchical models of learning suggest that the precision-weighting associated with an error will be affected by the precision associated with the local prediction, but also the precision associated with beliefs/models at the level above over a longer timescale (Mathys et al., 2014; Mathys, Daunizeau, Friston, & Stephan, 2011). In the case of these sequences, we suggest that the brain formulates predictions about the next sound (driven by local sequence statistics) that are weighted by how likely it is that the next sound might violate those predictions (that errors occur), which in turn is weighted by how likely it is that the likelihood of change varies (that error rates/density change). Such models theoretically link nested predictions within a hierarchical inferential network in the brain where more rostral areas reflect longer timescales predictions (and updates) and more caudal areas represent shorter timescale predictions (and updates) and the violation of predictions at higher levels will in turn impact the precision-weighting of errors at lower levels (Hasson et al., 2008; Kiebel, Daunizeau, & Friston, 2008). Specifically, DCM is used here to test the following hypotheses: (1) that the precision-weighting associated with deviants will be higher in the first block type than the alternate block type; (2) that the error signal generated to deviants in the alternate block type will trigger more change in network connectivity reflecting higher levels of new learning in general (model updating); and (3) that the precision-weighting associated with first block and alternate block deviants will be associated with hierarchical differences in precision modulation after the change in block length.

METHODS

Participants

Participants were 19 healthy adults (15 women; aged 18–53 years, M = 25.26 years, SD = 11.44 years) recruited from undergraduate psychology students at the University of Newcastle and community volunteers. Exclusion criteria included current diagnosis of, or treatment for, a mental disorder per Diagnostic and Statistical Manual of Mental Disorders–Fifth Edition American Psychiatric Association (2013) criteria, history of head injury or neurological disorder, hearing loss, regular recreational drug use, heavy alcohol use, or a first-degree relative with schizophrenia. All participants provided written informed consent to participate in the study protocol as approved by the University of Newcastle Human research ethics committee prior to participating. Reimbursement was provided in the form of course credit for students or gift vouchers for volunteers as compensation for time and expenses incurred.

Stimuli and Sequences

The data used in this study were previously published in a traditional auditory ERP analysis (Fitzgerald & Todd, 2018). The sound sequences were compatible with those presented diagrammatically in Figure 1 and comprised two sounds: 30- and 60-msec tones created with a 5-msec rise/fall time and a 20- and 50-msec pedestal, respectively. Tones were 1000-Hz pure tones, presented over binaural Sennheiser HD280pro headphones at 75 dB with a regular 300-msec SOA. Tones were arranged into two different probability blocks; a first block type in which the 30-msec tone was a common standard (p = .875) and the 60-msec tone was a rare deviant (p = .125) and an alternate block in which the tone probabilities reversed. In other words, in this data set, the 60-msec tone was always the first deviant and the 30-msec tone always the second deviant. Blocks always began with five occurrences of the relevant standard, and no two deviants ever occurred consecutively within a block.

There were two different block lengths as per Figure 2, with participants hearing 3840 tones with the first 1920 arranged into the 12 faster alternating blocks of 160 tones, and the second 1920 arranged into four blocks of 480 tones. After a brief (5 min) silent break, the
participants were presented with the same 3840 tones, with the first 1920 arranged into four blocks of 480 tones and the second 1920 arranged into 12 faster alternating blocks of 160 tones.

EEG Data Collection and Preprocessing

Fitzgerald and Todd (2018) obtained a continuous EEG recording during presentation of the sound sequences via a SynAmps2 Neuroscience system using a 1000-Hz sampling rate, high-pass 0.1 Hz, low-pass 70 Hz, notch filter 50 Hz, and fixed gain of ×10. The EEG setup consisted of 64 electrodes in accordance with the International 10 ± 10 system with Modified Combinatorial Nomenclature (“Guideline 5: Guidelines for Standard Electrode Position Nomenclature”; American Clinical Neurophysiology Society, 2006) and included one electrode at the nose and each of the bilateral mastoids for use as reference. Additional EOG electrodes were placed 1 cm from the outer canthi of each eye and directly above and below the left eye to monitor eye movements. Impedances were reduced to below 5 kΩ prior to recording. The continuous EEG recordings (Fitzgerald & Todd, 2018) were processed using Neuroscan Edit software for suitability to the current DCM analysis. Adjustments involved bandpass filtering to a range of 0.1–40 Hz with 12-dB drop-off and zero phase. Manual artifact rejection and bad channel exclusions were carried over from the previous analysis. Eye blink corrections were also completed in the previous analysis using an EEG–VEOG covariance analysis, linear regression, and point-by-point subtraction procedure (Semlitsch, Anderer, Schuster, & Presslich, 1986). Data were epoched from 50-msec pre-stimulus to 300-msec post-stimulus, and any epochs containing amplitudes exceeding ± 70 μV discarded prior to averaging. Subsequent processing steps were undertaken using Statistical Parametric Mapping (SPM) software (Version 12, Revision 6906). SPM is a freely available academic software package specialized for the spatially extended statistical analysis of brain imaging data and which is suitable for DCM analyses (“Statistical Parametric Mapping”; Functional Imaging Laboratory, 2016).

Data were common-average referenced as recommended for the application of DCM of EEG data (Litvak et al., 2011), and a baseline correction from 25-msec prestimulus to 25-msec poststimulus was applied to prevent late prestimulus baseline contamination because of isochronous presentation: auditory event-related responses typically starting later than 25-msec poststimulus onset. Single-subject and grand averages were subsequently generated for the response to each tone (60 msec, 30 msec) as deviant for periods before the block length changed and after the block length changed. As reviewed in the Introduction section and in Figure 1, the effects on MMN are order-dependent, and not tone-property- or sequence-dependent. The data were therefore arranged to test the effects of block deviant (first vs. alternate) and the effect of a block length change (before vs. after change) on each deviant. The resultant averages for the first and alternate deviants before and after the block change therefore contained a maximum of 2 × 120 samples each (i.e., deviants from each of the sequence orders in Figure 2). This provided a high signal-to-noise ratio sample from each participant for each period of interest in the sequence.

Sensor Space Analyses

The common-average referenced auditory ERPs waveforms were analyzed to confirm the presence of a difference in response amplitude between the first and alternate deviant type and a difference between the deviants in the effect of a change in block length. Deviant responses were compared using family-wise-error corrected paired t tests to test for significant (p < .001) differences in these responses at each sampling point within the epoch at each scalp site. This method implemented within SPM software enables identification of locations where event-related response amplitude differs between two conditions at a given time point reliably across participants (see Litvak et al., 2011, for further details of the analysis methods). A cluster-level analysis was applied also to identify local maxima with spatiotemporal voxels forming separable clusters, based on the random field theory (Kilner, Kiebel, & Friston, 2005). This included a direct comparison of the two deviant tone types and a comparison before and after change in block length for each deviant and the deviants combined.

DCM

DCM was used to estimate population output and connectivity parameters associated with the two deviant types (first block deviant and alternate block deviant) before and after change in block length in order to test hypotheses that (1) the precision associated with deviants will be higher in the first than alternate block type, (2) the error signal generated to deviants in the alternate block type will trigger more change in network connectivity reflecting higher levels of new learning (model updating), and (3) that the precision associated with first and alternate block deviants will be associated with hierarchical differences in precision modulation.

DCM allows for a mapping from data measured at the sensor level to source-level activity, in a sparse network of interconnected sources, each consisting of a set of neural populations based on a canonical microcircuit architecture (Bastos et al., 2012). The activity in each source evolves as described using coupled differential equations that model the dynamics of postsynaptic voltage and current in each neural population. These populations (spiny stellate cells, superficial and deep pyramidal cells, and inhibitory interneurons) have distinct connectivity profiles of ascending and descending projections linking different sources (extrinsic connectivity; A parameters) and coupling neural populations within each source (intrinsic connectivity; G
parameters). In addition, the superficial pyramidal cells can be subject to activity-dependent gain modulation, modeling short-term plasticity (M parameters; Auksztulewicz et al., 2018). Both extrinsic connections and gain parameters were assumed to undergo condition-specific changes, giving rise to differences between responses (i.e., main and interaction effects described below). These condition-specific changes were modeled as B parameters (extrinsic connectivity changes) and N parameters (pyramidal gain changes). DCM based on canonical microcircuits has been used in several other studies of mismatch responses (e.g., Auksztulewicz & Friston, 2015) and validated using invasive recordings in humans.

The DCM adopted a standard electromagnetic forward model based on the boundary elements model in Montreal Neurological Institute space as the default SPM 12 template (“Statistical Parametric Mapping”; Functional Imaging Laboratory, 2016). Lead-fields specified by the forward model were used to reconstruct responses at all electrodes and latencies (0–300 msec) from six cortical sources considered for inclusion in the DCM: bilateral primary auditory cortex (A1), bilateral superior temporal gyrus (STG) and bilateral inferior frontal gyrus (IFG), using the following Montreal Neurological Institute coordinates (Garrido, Kilner, Kiebel, & Friston, 2009): left A1 (−42, −22, 7), right A1 (46, −14, 8), left STG (−61, −32, 8), right STG (59, −25, 8), left IFG (−46, 20, 8), right IFG (46, 20, 8). The time window included the MMN and extended into the P3a, following previous literature. Source selection was based on recent literature on the procedures for Bayesian magnetoencephalography/EEG source reconstruction (Litvak et al., 2011; Kiebel, Daunizeau, Phillips & Friston, 2008) and previous DCM studies of MMN paradigms informed by both neuroimaging data and neuroanatomical knowledge of connectivity (Garrido, Kilner, Kiebel, & Friston, 2009) with these separate nodes within the hierarchy suggested to service different stages in the analysis of sound relevance (Schönwiesner et al., 2007). The free-energy approximation to model evidence was used as a metric of model fit to the data, penalized by model complexity.

The DCM analysis modeled changes in the deviant response only given that order effects on MMN amplitude modulations are driven almost entirely by the deviant auditory ERPs (Fitzgerald & Todd, 2018) and focused on differential changes in connectivity associated with the first and alternate deviant tone and before versus after a change in block length. We conducted the analysis in a hierarchical manner—first, optimizing the model structure at the individual participants’ level and then inferring the significant parameters at the group level.

At the first level, DCMs were fitted to single participants’ data over the two factors of deviant type (first vs. alternate) and change in block length (before vs. after) as well as their interaction testing for differences in the directionality of effects. We considered a large model space of alternative models, whereby each model allowed for a different subset of parameters to contribute to the observed data. Specifically, the chosen model space examined each combination of changes in ascending connections, descending connections, and gain parameters for each factor, resulting in $8 \times 8 \times 8$ factorially designed models (eight models for each factor, corresponding to the modulation of ascending, descending, ascending/descending, and no extrinsic connections, each with and without modulatory gain changes). Rather than fitting each model separately, Bayesian model reduction was used to identify the subset of parameters whose changes explained the observed auditory ERP data. Bayesian model reduction uses inversion of the “full” model incorporating changes in all identified parameters to estimate model evidence for a range of “reduced” models where some parameters are not permitted to vary (Friston & Penny, 2011). The free-energy approximation to log-model evidence was used to calculate each model’s posterior probability and select the winning model for each participant. Because in all participants the full model was the winning model (see Results section), the individual participants’ full models were taken into further analysis.

Model inversion in DCM is susceptible to local maxima issues because of the inherently nonlinear nature. To overcome this potential issue, at the second level, we implemented parametric empirical Bayes, an iterative hierarchical implementation of the empirical Bayesian inversion method (Friston, Zeidman, & Litvak, 2015) where group-level effects are inferred by (re)fitted the same model to each participant’s data under group constraints (e.g., the assumption that model parameters are normally distributed in the participant sample) updating the posterior distribution of the individual parameters. This process was applied using the built-in SPM 12 function sdm_dcm_peb_fit.m. Specifically, the winning models were entered into parametric empirical Bayes to hierarchically estimate the variation in parameters that explained systematic changes in response to each factor (deviant type, block length change, and their interaction). This resulted in obtaining Bayesian confidence intervals for each parameter as a measure of uncertainty in the estimates. Parameters with 99.95% confidence intervals falling either side of zero were considered statistically significant (i.e., corresponding to $p < .005$).

All data and code associated with the above analyses are not currently made openly available as sufficient permissions are not clearly made in the ethical clearance for the study. However, permission can be sought to make code available on specific request to the corresponding author.

RESULTS

Sensor Space Results

A traditional analysis of the deviant response amplitudes produced by participants in this study has been published previously (Fitzgerald & Todd, 2018), and corresponding waveforms are presented in Figure 3 averaged according to the key analyses featured in this paper (Figure 3A) and
their sensitivity before and after the change in block length (Figure 3B and C). The results of the analysis of differences in the common average-referenced deviant responses are presented in the bottom of Figure 3 in which the points of difference are depicted in 2-D space and over time (see caption for detailed description). A comparison of the response to the two deviant tones yielded widespread differences over frontal–central scalp sites that correspond to larger negative values for the first deviant at cluster level at 175 msec and at various time points throughout ~80–175 msec at peak level as evident in shaded areas in Figure 3A, bottom. Cluster-level analyses revealed that the response to deviants was significantly more negative before versus after the change in block length at 126 msec for the first deviant. As is evident in Figure 3B, the response to the first deviant is visibly smaller after the change in regular block length with the difference being dominant over the right frontal scalp regions. No such differences were evident for the alternate deviant (Figure 3C). When averaged over both deviant types (not shown), a marginally significant main effect of block change on the deviant ERPs emerged frontally at 126 msec clearly driven by the effects on the first deviant. In short, the overall response was larger to the first deviant and the effect of a block length change only apparent for the first deviant.

**Connectivity Effects**

The six-source cortical network within which connectivity changes were modeled is depicted in Figure 4 and composed of bilateral sources A1, STG, and IFG in accordance with the source locations chosen in previous DCM analyses of auditory MMN paradigms (Garrido, Kilner, Kiebel, & Friston, 2009; Garrido, Kilner, Kiebel, Stephan, et al., 2009). The full model permitted changes in ascending, descending, and intrinsic coupling between sources (model FBi in Figure 4), and was compared with a set of reduced models consisting of changes in each parameter alone, each combination, and a null model permitting no changes (see Figure 4 for representation of full model space).

Bayesian model reduction was performed on all models. The overall winning model was the most complex model permitting changes in ascending, descending, and intrinsic connections (i.e., model FBi; see Figure 3A). This model was favored in 100% of individual participants, with a posterior probability exceeding 0.99 in all cases indicating an excellent fit.

**Effects of First and Second Deviant and Block Length Change**

Planned contrasts between responses to the alternate and first deviant tones, the two sequence components (before and after a change in block length), and their interaction were conducted to test for significant differences in connectivity associated with an order-driven modulation based on deviant type and block length pattern violation. This analysis revealed significant ($p < .005$) main effects of deviant type, block length change, and interactions between deviant type and block length.

Figure 3. Sensor-level results for first versus alternate deviant comparison (A), first deviant before and after the block length change (B), and alternate deviant before and after the block length change (C). The group-averaged deviant responses are presented in the top for the site F4 where responses and differences were maximal. The lower panels represent the sensor space flattened to a 2-D representation and enables identification of locations (based on $x$, $y$ space) and time points (~50 to 300 msec as displayed for $x$ and $y$ spaces) where response amplitude differed between two conditions (F test, $p < .001$ FWE-corrected). The upper left of these in each case represents the topographical location of significant differences based on $x$ and $y$ coordinates (averaged over time). The bottom and right represent the significant differences over time in a 1-D spatial map averaged along the $x$ and $y$ axes, respectively.
change throughout the network. The main effects and their interaction are depicted in Figure 5. A full summary of parameter averages is provided in Tables 1 and 2.

The differences in connectivity dynamics between the first and alternate deviant responses and the effect of a change in block length are generally in alignment with the three hypotheses. First, relative to the alternate deviant, the first deviant is generally associated with higher precision in the prediction error (i.e., lower inhibition of the superficial pyramidal cells). This is evident at right A1 and bilaterally at the STG (blue curved arrows Figure 4 left) with the opposite pattern at left A1 (see Discussion section). The first deviant is also associated with a lower level of forward coupling between nodes on the right, and this is consistent with these errors being recognized as comparatively weaker learning signals prompting less model updating than the alternate deviant. There is also a significantly higher coupling in the backward connections for the first relative to the alternate deviant. Backward connections are thought to access modulatory voltage-dependent postsynaptic effects with long time constants making them ideal for mediating context-dependent effects. They are thought to convey the predictions that should suppress activity in lower levels that encode the prediction errors (Chen, Henson, Stephan, Kilner, & Friston, 2009). The results suggest that the tendency to suppress error signals is comparatively stronger for alternate deviant overall between IFG and STG, but stronger for the first than alternate deviant overall between STG and A1.

Blocks presented after a change in block length were associated with a lower precision of prediction error (higher inhibition of superficial pyramidal cells) bilaterally at A1 (red curved arrows Figure 5 center) relative...
to blocks before the change in block length. The data after
the change in block length were also associated
with lower forward coupling from bilateral A1 to STG, and
greater forward coupling from bilateral STG to IFG.
Meanwhile, a lower level of backward coupling from IFG
to STG is seen bilaterally for the data after relative to before
the change in block length suggests a relatively lower
top–down suppression of error signals and increased for-
ward error signaling after the change in block length.

The connectivity changes in response to the two devi-
ant types were hypothesized to be differentially affected
by the change in block length. This was supported by the
significant interactions between deviant tone and change
in block length effect on coupling parameters where col-
ored arrows indicate where the directionality of change
differed for the two deviant types. The results in
Figure 5 right indicate that, relative to the alternate devi-
ant, the first deviant exhibited the hypothesized drop in
precision bilaterally at A1 and at left IFG but was charac-
terized by higher precision at STG after the block length
change. The forward message passing associated with the
first deviant decreases at lower levels (between A1 and

Table 1. Estimated Changes in Forward, Backward, and Intrinsic Connections within/between Sources

<table>
<thead>
<tr>
<th></th>
<th>First vs. Alternate Deviant</th>
<th>After vs. Before Block Length Change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Forward</td>
<td>Backward</td>
</tr>
<tr>
<td>IA1-lSTG</td>
<td>0.057</td>
<td></td>
</tr>
<tr>
<td>rA1-rSTG</td>
<td>-0.470*</td>
<td></td>
</tr>
<tr>
<td>lSTG-lIFG</td>
<td>-0.033</td>
<td></td>
</tr>
<tr>
<td>rSTG-rIFG</td>
<td>-1.11*</td>
<td></td>
</tr>
<tr>
<td>lSTG-lA1</td>
<td></td>
<td>-0.010</td>
</tr>
<tr>
<td>rSTG-rA1</td>
<td></td>
<td>0.216*</td>
</tr>
<tr>
<td>lIFG-lSTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rIFG-rSTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IA1</td>
<td>0.686*</td>
<td></td>
</tr>
<tr>
<td>rA1</td>
<td>-0.144*</td>
<td></td>
</tr>
</tbody>
</table>

* Significant changes assessed as 99.95% of the estimated probability distribution demonstrating a change greater or less than zero.
**DISCUSSION**

In this study, DCM was used to explore the causes of order effects on the responses to sounds that violate regular patterns. These order effects suggest that the statistical learning that governs auditory perceptual inference is not local, but it is instead potentially influenced by learning about regularities on multiple timescales and timescales longer than originally thought. The DCM results reveal differences in the precision-weighting associated with the response to pattern violations within an initial context versus a later context in which two sounds invert periods of being a predictable regular tone and rare pattern deviation. This precision-weighting is proposed to reflect the quality of evidence upon which predictions about the regular pattern are based. The finding of differential precision-weighting is important because the quality of evidence on a shorter timescale (minutes) is identical for both the initial and later context—the difference is in the order in which they were encountered and therefore the longer-term history with the sounds. The second DCM result of note is that these precision-weightings are differentially impacted by a higher-level pattern violation. The local context patterns in this study were nested in a higher-order longer timescale regularity (i.e., the length of periods or blocks of alternating tendencies/probabilities in sound). When the regular higher-order pattern was violated, the precision-weightings on the local pattern violations change, with the DCMs indicating different directions of change at different levels of the underlying network for the two deviant types. These DCM results are discussed here with reference to what they might reveal about how order effects and violations of patterns on different timescales impact perceptual inference.

The MMN that occurs in the response to pattern violating deviant sounds has been described as an index of surprise (Friston, 2010), shown by animal research to be generated by the response of superficial pyramidal cells in the receptive regions of cortex (Lee et al., 2017; Javitt, Steinschneider, Schroeder, & Arezzo, 1996). Using DCM, the precision that should theoretically modulate MMN amplitudes is modeled by the level of gain on inhibitory interneurons that synapse with the superficial pyramidal cells (Aukstulewicz & Friston, 2015). Therefore, where gain on these inhibitory cells is lower, the precision in the error signaling from superficial pyramidal cells is higher as their activity is less suppressed by these inhibitory connections. In the present data, the precision-weighting was significantly higher to the first deviant tones in general within primary and secondary auditory areas. We know of no reason why the left A1 region was an exception to this pattern, although in general the deviant response generated to a change in simple tone properties is often right dominant (Paavilainen, Alho, Reinkainen, Sams, & Näätänen, 1991), which is consistent with the data here in the sense that many coupling changes were more evident in right hemisphere nodes. The overall larger response to first deviants than the alternate deviants in scalp-recorded potentials is therefore consistent with higher output from superficial pyramidal cells within these nodes.

The deviant type associated with the higher precision errors (first deviant) was however associated with less pronounced change in forward coupling between nodes on the right. With respect to the scalp-recorded potentials, this is consistent with less change in the deviant response amplitude over the course of a block of sound (see Todd, Petherbridge, et al., 2018; Todd, Heathcote, et al., 2014). This dissociation between error signaling precision and changes in message passing in these data is also consistent with the notion that deviants in the first block are experienced as highly precise errors where rapid learning enables these sounds to be heard as clear outliers within an environment well described by a current model. In contrast, the errors occurring in the alternate

---

**Table 2.** Estimated Changes in Forward, Backward, and Intrinsic Connections within/between Sources Associated with Alternate Compared to First Deviant After versus Before the Change in Block Length

<table>
<thead>
<tr>
<th></th>
<th>Forward</th>
<th>Backward</th>
<th>Intrinsic</th>
</tr>
</thead>
<tbody>
<tr>
<td>IA1-ISTG</td>
<td>0.018</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rA1-rSTG</td>
<td>-0.486*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STG-IFG</td>
<td>0.368*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rSTG-rIFG</td>
<td>0.225*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STG-IA1</td>
<td>-0.437*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rSTG-rA1</td>
<td>-0.244*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lIFG-ISTG</td>
<td>-0.316*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rIFG-rSTG</td>
<td>0.039</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IA1</td>
<td>0.702*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rA1</td>
<td>0.228*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STG</td>
<td>-0.186*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rSTG</td>
<td>-0.337*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lIFG</td>
<td>0.305*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rIFG</td>
<td>-0.017</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Significant changes assessed as 99.95% of the estimated probability distribution demonstrating a change greater or less than zero.
within the paradigm—once in the transition from faster to slower alternating blocks and once in the transition from slower to faster alternating blocks. In each sequence, there is therefore only one opportunity to learn that block lengths can change and it is likely to have been a surprising event. At this point in the sequence, there would be sufficient learning that there are both regular tone tendencies (shortest timescale prediction), and that tone tendencies change regularly (middle timescale prediction). Therefore, the surprise is linked to a sudden change in the regularity with which tone tendencies have changes throughout the sequence. Surprise should reduce precision-weighting and trigger relearning, and the DCM results provide evidence of this. The precision-weighting on deviants was differentially impacted for the two deviant types with the first relative to alternate deviant showing the predicted drop in error precision at the primary auditory cortices and left IFG, which occurs with a decline in forward message passing from A1 to STG. In contrast, there is a comparative increase in precision-weighting of errors at the STG coupled with higher message passing from the STG to IFG. These hierarchical changes in the weighting of errors throughout the network after the change in block length could be taken as evidence in a shift in priority for learning from errors to relearn longer timescale regularities. The associated decreased coupling in backward connections could be taken as evidence that all errors become more important as learning signals (therefore lead to less suppression of future errors) after the violation of longer timescale patterns.

With respect to the scalp-recorded response, the MMN elicited by first deviants drops in amplitude significantly after the block length changes, whereas the alternate deviant response is unchanged. When considered as a precision-weighted error signal, the change in first deviant responses appears to reflect the weighting at the lowest level (A1) and/or left IFG but not the STG. This observation drives home the point that DCM models can uncover complex relationships between the inferred activity of multiple brain regions contributing to auditory ERPs observed at the scalp level. Within the scalp-recorded response, we see only the resultant time-locked response to the deviant and the differences in amplitude occurring at different stages of the sequence. Within the DCM, we can model how changes in coupling within and between different brain regions might generate these observations.

The results are consistent with the assertion that perceptual inference engages a hierarchical network architecture where more rostral projections such as those to the pFC should be responsible for generating and updating beliefs about longer-term patterns like changes in volatility, whereas more caudal regions are sensitive to short-timescale change like local deviance (Kiebel, Daunizeau, & Friston, 2008). Despite the encouraging pattern of results, the interpretation should be balanced against a number of study limitations. First, one might ideally like to see replication of these results in a fully counterbalanced design (e.g., with additional data sets in which the 30-msec tone is the first deviant and where the order of sequence presentations is balanced). These forms of counterbalancing do exist in prior studies auditory ERPs (reviewed in the introduction and presented in Figure 1) and replicate the order effects observed here in MMN amplitude modulations. These earlier studies were acquired with a reduced electrode montages and are not suitable for the application of DCM. However, given that DCM was applied to model changes in deviant amplitudes, and these changes are established to follow tone order and not tone properties (Figure 1), there is no reason to suppose the DCM results are specific to hearing the longer tone as a first deviant.

Another limitation is in the DCM model selection where relative evidence is calculated only for the model space which is predefined. This entails the possibility that the data could be more accurately explained by an alternative model that is not captured in the model space. The chosen model space represents that supported by the literature pertaining to auditory change detection and MMN generation (Garrido, Kilner, Kiebel, & Friston, 2009; Garrido et al., 2008). This prior literature adds confidence that it represents a plausible model for generation of the observed responses based on current knowledge and also how connectivity changes within a commonly accepted inferential network structure can account for our data.

While this study is also limited in its focus on applying DCM to auditory ERPs (i.e., event-related EEG signals averaged over many trials), recent trial-by-trial analyses have supported Bayesian learning across multiple levels of volatility during a multi-feature visual roving standard paradigm (Stefanics, Kremláček, & Czigler, 2014), and similar models of hierarchical Bayesian belief updating have also been successfully applied in DCM studies of visuospatial attention (Vossel et al., 2014). The current findings could therefore be interpreted as supporting hierarchical Bayesian inference as a general framework for inference and learning in the brain, rather than merely a specific feature of auditory processing.

Fitzgerald et al. 1559
Conclusions
The application of DCM in this study offers insights into patterns of modulation in auditory ERPs that do not make sense based on local timescale statistical learning. The estimates of precision-weighting on responses to pattern violations add to a body of work that significantly challenges the notion that the MMN component of the auditory ERP is principally influenced by referencing the content of auditory sensory memory. We propose that the DCM results are compatible with the idea that auditory inference might involve learning over multiple timescales, where surprise generated by violations of different timescale patterning may be expressed differently within an underlying hierarchical network with violations of longer timescale patterns leading to remodeling of predictions at higher levels. Finally, elucidating the neurophysiological basis for perceptual inference and learning in the healthy brain has important implications for understanding the etiology of disorders such as schizophrenia where these inference processes differ significantly (e.g., Powers, Mathys, & Corlett, 2017; Friston, Brown, Siemerkus, & Stephan, 2016) and also the processing of aging (Cheng, Hsu, & Lin, 2013). An expanded understanding of the reference periods for determining surprise, and the influence of surprise at different timescales and levels of inferential networks, could stimulate more elegant and targeted experimental designs to elucidate the causes of group differences evident in perceptual inference.

Reprint requests should be sent to Kaitlin Fitzgerald, School of Psychology, University of Newcastle, University Drive, Callaghan NSW 2308, Australia, or via e-mail: kaitlin.fitzgerald@uon.edu.au.

Author Contributions
Kaitlin Fitzgerald: Conceptualization; Methodology; Visualization; Writing—Original draft; Writing—Review & editing. Ryszard Aukstulewicz: Methodology; Software; Validation; Writing—Review & editing. Alexander Provost: Software; Formal analysis. Bryan Paton: Software; Validation; Writing—Review & editing. Zachary Howard: Software; Formal analysis; Validation; Writing—Review & editing. Juanita Todd: Conceptualization; Supervision; Writing—Review & editing.

Funding Information
K. F. acknowledges receipt of Australian Postgraduate Award scholarships. This research was supported by funds provided by the National Health and Medical Research Council of Australia, grant number: APP1002995.

Diversity in Citation Practices
A retrospective analysis of the citations in every article published in this journal from 2010 to 2020 has revealed a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the Journal of Cognitive Neuroscience (JoCN) during this period were M(an)/M = .408, W(oman)/M = .355, M/W = .108, and W/W = .149, the comparable proportions for the articles that these authorship teams cited were M/M = .579, W/M = .243, M/W = .102, and W/W = .076 (Fulvio et al., JoCN, 33:1, pp. 3–7). Consequently, JoCN encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article’s gender citation balance.

REFERENCES


