

Not so primitive: context-sensitive meta-learning about unattended sound sequences

Juanita Todd,^{1,2,3} Alexander Provost,^{1,2} Lisa R. Whitson,^{1,2} Gavin Cooper,^{1,3} and Andrew Heathcote^{1,2}

¹School of Psychology, University of Newcastle, Callaghan, NSW, Australia; ²Priority Research Centre for Translational Neuroscience and Mental Health Research, University of Newcastle, Callaghan, NSW, Australia; and ³Schizophrenia Research Institute, Darlinghurst, NSW, Australia

Submitted 9 July 2012; accepted in final form 4 October 2012

Todd J, Provost A, Whitson LR, Cooper G, Heathcote A. Not so primitive: context-sensitive meta-learning about unattended sound sequences. *J Neurophysiol* 109: 99–105, 2013. First published October 17, 2012; doi:10.1152/jn.00581.2012.—Mismatch negativity (MMN), an evoked response potential elicited when a “deviant” sound violates a regularity in the auditory environment, is integral to auditory scene processing and has been used to demonstrate “primitive intelligence” in auditory short-term memory. Using a new multiple-context and -timescale protocol we show that MMN magnitude displays a context-sensitive modulation depending on changes in the probability of a deviant at multiple temporal scales. We demonstrate a primacy bias causing asymmetric evidence-based modulation of predictions about the environment, and we demonstrate that learning how to learn about deviant probability (meta-learning) induces context-sensitive variation in the accessibility of predictive long-term memory representations that underpin the MMN. The existence of the bias and meta-learning are consistent with automatic attributions of behavioral salience governing relevance-filtering processes operating outside of awareness.

mismatch negativity; perceptual inference; salience; learning; auditory evoked potential

HUMANS ARE ACCOMPLISHED at finding patterns in event sequences, an ability that is supported by automatic novelty detection mechanisms. In addition, automatic novelty detection is indexed by mismatch negativity (MMN), a fronto-central event-related potential (ERP) peaking 100–200 ms after a novel event. The MMN, which is primarily generated in the auditory cortex, was first described by Näätänen, Gaillard, and Mäntysalo (1978) in an auditory oddball paradigm (e.g., a series of standard longer tones containing an occasional shorter oddball or deviant tone) through the use of deviant-standard difference waveforms. MMN is elicited automatically and is usually measured while participants attend to another modality (e.g., while reading or watching a silent movie), as it does not require attention but can be masked by attention-related ERPs. MMN amplitude is proportional to the difference between deviant and standard and is inversely proportional to the probability of the deviant. Early interpretations of the MMN (e.g., Näätänen and Michie 1979) were in terms of a mismatch between low-level auditory sensory memory traces of the standard and the deviant. However, mounting evidence has implicated much more sophisticated processing, leading Näätänen, Tervaniemi, Sussman, Paavilainen, and Winkler

(2001) to characterize the MMN as a marker of “primitive intelligence” in the auditory cortex.

Primitive intelligence is revealed by phenomena usually associated with higher-order cognition ranging from prediction and simple concept formation to mnemonic characteristics more associated with long-term memory than short-term sensory memory. For example, MMNs indicative of a left hemisphere specialization in extracting abstract rules are associated with violations of contingencies embedded in sound sequences that are independent of low-level auditory features (e.g., “the higher the frequency the louder the intensity”; Paavilainen et al. 2001). Horváth, Czigler, Sussman, and Winkler (2001) found MMNs implicating simultaneous memory representations of more than one type of contingency (e.g., a global “every second tone is A and every other B” rule and a local “A follows B and vice versa” rule) that are compared in parallel with incoming sounds. These results and others (e.g., Tervaniemi et al. 1994) suggest that the auditory cortex automatically learns contingencies between the features of the successive events and makes predictions about forthcoming events (Winkler et al. 1996). The use of transition statistics (the statistical temporal dependencies linking stimuli) was formalized in a recent paper incorporating empirical data with computational modeling to explain a wide range of MMN findings. The authors provide additional support for the argument that low-level sensory effects of stimulation (e.g., habituation) are not sufficient to account for MMN results that instead conform to a more active process of cortical prediction (Wacongne et al. 2012).

This developing understanding of the role of MMN in the auditory modality complements Friston’s (2003, 2005, 2008) free-energy minimization framework for perceptual inference and learning, whereby sensory cortices are arranged hierarchically, with predictions over longer timescales made by representations in higher cortical levels modulating responses in lower levels occurring on faster timescales (see also Kiebel et al. 2008). The prefrontal cortex is a recognized contributor to the MMN (Alho et al. 1994). Escera, Yago, Corral, Corbera, and Nunez’s (2003) specific suggestion that the prefrontal cortex provides top-down modulation of mismatch detection in the temporal cortices was tested by Garrido, Kilner, Kiebel, and Friston (2009) in an auditory pitch oddball paradigm. Garrido et al. (2009) compared dynamic causal models varying in the involvement of generators in primary auditory cortex (A1), superior temporal gyrus (STG), and inferior frontal gyrus (IFG). Model selection supported the influence of adaptation in A1 and short-term plasticity of forward and backward connec-

Address for reprint requests and other correspondence: J. Todd, School of Psychology, Univ. of Newcastle, University Dr., Callaghan, NSW, Australia, 2308 (e-mail: juanita.todd@newcastle.edu.au).

tions across the auditory hierarchy in the generation of MMN (see also Schmidt et al. 2012 for a recent replication and extension). The model that best accounted for data specified a right IFG-STG-A1 hierarchy and a left STG-A1 hierarchy with extrinsic (feedforward and feedback) connections between generators within each hierarchy and intrinsic (lateral) connections within each A1 generator. These findings, and Friston's general multiple-timescale hierarchical framework, suggest that the known frontal involvement in the MMN might be related not only to proposed attention switching (Giard et al. 1990; Näätänen 1990, 1992) but also to modulating MMN magnitude based on predictive confidence over longer timescales.

We used a new technique, a multiple-context and -timescale MMN protocol, to explore the long-term memory characteristics of the context-dependent process that adjusts predictions about auditory regularity. The technique is a refined and expanded version of the protocol used by Todd, Provost, and Cooper (2011). They measured learning about the probability of a tone duration deviant in an oddball paradigm similar to that illustrated in the top row of Fig. 1. In each of a series of ~10-min sequences separated by several minutes of silence, either a short (e.g., 30 ms) or long (e.g., 60 ms) tone occurred every 300 ms. Over the entire sequence both durations occurred equiprobably, but in blocks within each sequence one duration, the standard, was more probable ($P = 0.875$), with the other duration being the MMN-eliciting deviant. The attribution of durations to deviant and standard roles alternated between blocks. Different sequences varied in block length, with Fig. 1 illustrating sequences with slow (2.4 min) and fast (0.8 min) block alternations. If MMN amplitude is dominated by the local probability within a block—consistent with an MMN developing on the scale of a few seconds in the oddball paradigm (i.e., after several standard repetitions)—it should not vary with alternation speed (block length). If, in contrast, the probability of a deviant is measured by a moving average over a larger temporal window, MMN amplitude should be larger in slow than fast block alternation sequences.

Surprisingly, Todd et al. (2011) found both patterns. For the deviant duration that occurred in the first block (which was the same for every sequence for a given participant) MMN amplitude was larger for the slow than the fast sequences. In contrast, for the duration that became the deviant in the second block, sequence speed had no effect on MMN magnitude. Todd et al. described this asymmetric finding as a “primacy bias.” They suggested that it might reflect latent inhibition (Lubow and Gewirtz 1995), a classical conditioning phenomenon whereby learning is attenuated to familiar stimuli that have previously been inconsequential. Irrespective of the cause, the

data imply a long-acting, order-driven limitation on how evidence affects perceptual inference.

The experiment reported here adds multiple contexts to the multiple temporal scales in the protocol of Todd et al. (2011) in order to investigate the cases and limits of the differential probability sensitivity indicated by the primacy bias. In the previous study, tone order was a between-subjects factor whereby half the participants always experienced the long-duration sounds as the standard in the first block of any sequence and half always experienced the short-duration sounds as the first standard. Furthermore, the primacy bias was assessed over a 50-min recording period including multiple block lengths. Here we presented participants with three pairs of sequences comprising only short and long block lengths (as illustrated in Fig. 1, *orders 1, 2, and 3*). Each sequence pair was separated by a 5-min break, and the duration that was used as the standard in the first block of each sequence changed between pairs (i.e., 30 ms in the 1st and 3rd pairs and 60 ms in the 2nd pair). These shorter sequences allow us to determine whether a reliable index of the bias can be extracted from a 20-min recording and how resistant the bias is to change (i.e., whether it reverses when tone order changes). Latent conditioning is well known to be context sensitive (e.g., Hall and Honey 1989), so if the 5-min breaks induce a sufficiently salient change of context the primacy bias should reverse between sequence pairs. Furthermore, the repeat of *order 1* in *order 3* allows us to examine whether the bias is always replicated with the same initial sequence structure or whether prior experience can alter the effect.

METHODS

Participants. Participants were 15 healthy adults (8 women, 7 men; 18–31 yr, mean = 25 yr, SD = 4 yr), community volunteers and first-year undergraduate Psychology students at the University of Newcastle. Participants were excluded if they were diagnosed with or being treated for mental illness, had a first-degree relative with schizophrenia, regularly used recreational drugs, or had history of neurological disorder, head injury or surgery, hearing impairments, or heavy alcohol use. Course credit was offered for participation to students and cash reimbursement to community volunteers. Written informed consent was obtained from all participants to complete the protocol as approved by the Human Research Ethics committee, University of Newcastle.

Stimuli and sequences. Sounds were 1,000-Hz pure tones presented binaurally over headphones at 75 dB SPL. Sounds were created with 5-ms rise/fall times and either a 20-ms or a 50-ms pedestal to produce 30-ms and 60-ms sounds, respectively. All sequences comprised 1,920 sounds presented at a regular 300-ms stimulus onset asynchrony (9.6 min per sequence). In short-standard blocks the 30-ms tone was more probable ($P = 0.875$) than the 60-ms tone; in long-standard

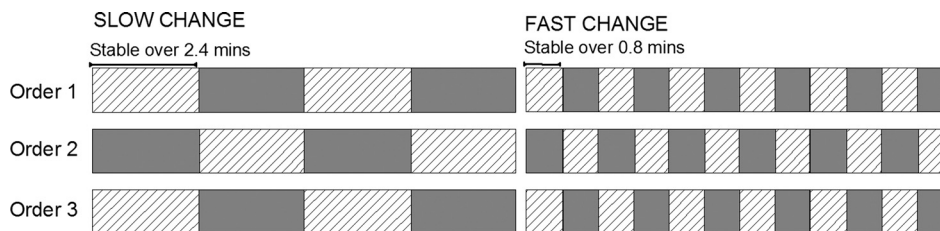


Fig. 1. Example structure of tone sequences used in the study of Todd et al. (2011; *top row only*) and the present study. Hatched rectangles represent blocks with a 30-ms standard and a 60-ms deviant tone, and gray rectangles represent blocks with reversed tone probabilities; in both $\text{Pr}(\text{standard}) = 0.875$ and $\text{Pr}(\text{deviant}) = 0.125$. Note that Todd et al. also used several intermediate speeds and found no difference between results when the different speeds occurred in different orders, and so only the slow then fast order was used in the present study.

blocks the probabilities were reversed. In the slow sequence, block type alternated after every 480 tones, creating a stable-standard period of 2.4 min (i.e., 2 repeats of each 2.4-min block). In the fast sequence, block type alternated every 160 tones, creating a stable-standard period of 0.8 min (i.e., 6 repeats of each 0.8-min block). The slow alternation sequence always preceded the fast alternation sequence. In *order 1* and in its repeat in *order 3*, the short-standard blocks were presented first. In *order 2*, the long-standard blocks were presented first. A 5-min break was enforced between order conditions, and shorter 1- to 2-min breaks occurred between sequences (total testing time \sim 1 h, 15 min).

Procedure. Participants completed a screening interview prior to testing to ensure that no exclusion criteria were present. Hearing thresholds (measured across 500–4,000 Hz) were assessed with a pure-tone audiometer to exclude those with hearing loss (thresholds $>$ 25 dB HL). Participants were fitted with a Neuroscan Quickcap with tin electrodes, which included nose and mastoid electrodes. Continuous EEG was recorded on a Synamps 2 Neuroscan system at a 1,000-Hz sampling rate (high pass 0.1 Hz, low pass 70 Hz, notch filter 50 Hz, and a fixed gain of 2,010). EEG data were recorded from 16 electrode locations (FZ, FCZ, CZ, PZ, F3, FC3, C3, F4, FC4, C4 in accordance with the 10-20 system plus left mastoid, right mastoid) referenced to the nose. We also measured vertical and horizontal electrooculograms. Impedances were reduced to below 5 k Ω before recording commenced. Sequences were presented over headphones while the participants viewed a silent DVD with subtitles and were instructed to ignore the sounds and focus attention on the movie.

Data processing. Continuous EEG was first examined off-line for major artifact before eyeblink artifact correction was completed off-line with Neuroscan Edit software. The method applies a regression analysis in combination with artifact averaging (Semlitsch et al. 1986). The average artifact response algorithm generated was assessed for adequacy ($>$ 30 sweeps in the average and $<$ 5% variance) and was applied to the continuous data files. The data were epoched from 50 ms before stimulus to 300 ms after stimulus. Epochs containing variations exceeding ± 70 μ V were excluded. The data were used to generate 12 ERPs to standard tones, 12 ERPs to deviant tones, and 12 difference waves per participant (a 30-ms and a 60-ms version for fast and slow sequences for each of the 3 orders). The first five standards in a block and the first standard after each deviant were excluded from averages.

ERPs were baseline corrected before stimulus. The standard and deviant ERPs were digitally filtered with a low pass of 30 Hz. Difference waveforms for 30-ms and 60-ms deviants were created for each condition by subtracting the ERP to that tone as a standard from that tone as a deviant. For example, the MMN to 30-ms deviants in fast change blocks was extracted from a difference waveform created by subtracting the ERP to the 30-ms standard in fast change blocks from the ERP to the 30-ms deviant tone in fast change blocks. This approach assists in reducing the contribution of exogenous effects in the computation of MMN (Jacobsen and Schröger 2003). The differ-

ence wave was then filtered with a low pass of 20 Hz (lower cutoff recommended for MMN; Kujala et al. 2007).

All ERPs were re-referenced to the averaged activity at the left and right mastoid sites. Individual data were then visually inspected to determine whether a MMN was present. One participant's data were rejected on this criterion, showing no evidence of a MMN to the 30-ms or 60-ms deviant for any condition. Three participants only completed *orders 1* and *2* of the study and were therefore excluded from statistical analyses and results display.

The within-subject variables of interest were order (*1, 2, 3*), speed of block alternation (slow, fast), and tone type (30 ms, 60 ms). Inspection of the data revealed that the speed and order effects on MMN amplitude were maximal at the front-central scalp site F4. MMN was quantified by identifying the peak latency in group-averaged data and extracting mean amplitude 10 ms either side of that peak. MMNs to the 30-ms tone peaked uniformly around 170 ms, and those to the 60-ms sound peaked uniformly around 150 ms (see Fig. 2 below). Mean amplitude was therefore extracted over 160 to 180 ms for 30-ms MMNs and from 140 to 160 ms for 60-ms MMNs. MMN amplitude was examined in an order \times speed \times tone repeated-measures ANOVA. Greenhouse-Geisser statistics are reported where appropriate.

RESULTS

The MMNs generated to the 30-ms and 60-ms tones as deviants are presented in Fig. 2 for site F4. The differential effect of tone order on the MMNs to 30-ms and 60-ms deviants is visibly apparent. In *order 1*, only the MMNs to 60-ms tones show evidence of the expected standard stability effect (slow-larger-than-fast alternation) on MMN size. In *order 2*, the pattern reverses entirely, where the slow-larger-than-fast alternation effect is only visible for the MMNs to the 30-ms tone as deviant. In *order 3*, however, the slow-larger-than-fast effect is clearly present for both the 30-ms and 60-ms MMNs.

Analysis of MMN amplitude exposed a main effect of speed [$F(1,10) = 8.00, P < 0.05$] modified by a significant three-way interaction between order, tone, and speed of block change [$\epsilon = 0.76, F(2,20) = 8.58, P < 0.005$]. In *order 1*, there was a tone \times speed interaction [$F(1,10) = 5.50, P < 0.05$], reflecting a significantly larger slow-change MMN than fast-change MMN for the 60-ms deviants only. In *order 2*, a significant tone \times speed interaction reflected the opposite pattern: a significantly larger slow- than fast-change MMN for the 30-ms deviant only [$F(1,10) = 26.31, P < 0.001$]. In *order 3*, only the speed of change main effect reached significance [$F(1,10) = 12.07, P < 0.01$], reflecting larger MMNs to deviants in the slow-change than the fast-change condition for both tone types. The full set of ANOVA results is presented in Table 1.

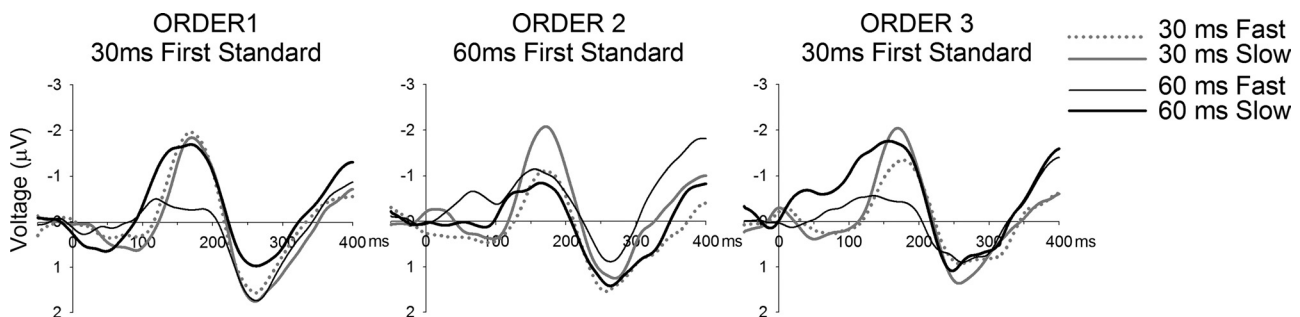


Fig. 2. Group-averaged mastoid re-referenced mismatch negativity (MMN) waveforms at F4 to 30-ms (gray) and 60-ms (black) deviant tones in the fast and slow change sequences for *orders 1–3*.

Table 1. Results for repeated-measures analysis of variance exploring main effects and interaction on mean MMN amplitude to 30-ms and 60-ms tones in fast- and slow-change conditions for sequence orders 1, 2, and 3

Effect	ϵ	F-Statistic	P Value	Mean Square	Error Mean Square
Overall					
Order	0.96	$F(2,20) = 0.72$	0.50	0.72	1.08
Speed	1.00	$F(1,10) = 8.00$	0.02	16.89	2.11
Tone	1.00	$F(1,10) = 1.76$	0.21	16.39	9.23
Order \times speed	0.86	$F(2,20) = 1.71$	0.21	1.24	0.73
Order \times tone	1.00	$F(2,20) = 1.02$	0.38	0.60	0.59
Speed \times tone	1.00	$F(1,10) = 0.61$	0.45	0.61	0.35
Order \times speed \times tone	0.76	$F(2,20) = 8.58$	0.01	9.91	1.16
Order 1					
Speed	1.00	$F(1,10) = 2.78$	0.13	4.73	1.70
Tone	1.00	$F(1,10) = 3.70$	0.08	10.45	2.82
Speed \times tone	1.00	$F(1,10) = 5.50$	0.04	6.80	1.24
Order 2					
Speed	1.00	$F(1,10) = 3.19$	0.10	2.10	0.66
Tone	1.00	$F(1,10) = 1.43$	0.26	3.56	2.50
Speed \times tone	1.00	$F(1,10) = 26.31$	0.01	7.57	0.29
Order 3					
Speed	1.00	$F(1,10) = 12.07$	0.01	12.12	1.01
Tone	1.00	$F(1,10) = 0.70$	0.42	3.59	5.09
Speed \times tone	1.00	$F(1,10) = 1.54$	0.24	0.89	0.58
30 ms					
Order	0.91	$F(1,10) = 1.18$	0.32	1.20	1.01
Speed	1.00	$F(1,10) = 8.66$	0.01	6.65	0.77
Speed \times order	0.65	$F(2,20) = 2.96$	0.10	4.21	0.92
60 ms					
Order	0.89	$F(1,10) = 0.37$	0.67	0.29	0.79
Tone	1.00	$F(1,10) = 6.18$	0.03	10.45	1.69
Speed \times order	0.91	$F(2,20) = 10.10$	0.01	6.42	0.58

MMN, mismatch negativity.

The group-averaged mean amplitudes of the MMN are presented in Fig. 3. The significant order effect on interactions between tone type and speed of alternation is very clear in Fig. 3C. The effect of order and speed of alternation on each tone type is presented separately in Fig. 3, A and B. A repeated-measures ANOVA within each tone type confirms a significant quadratic trend for the interaction between order and speed for both tones [30 ms $F(1,10) = 7.80$, $P < 0.05$; 60 ms $F(1,10) = 27.18$, $P < 0.001$], although the interaction only reaches significance for the 60-ms tone (see Table 1). This is visible in

Fig. 3, where the impact of speed on the difference in MMN amplitudes is maximal where the tone was the first encountered deviant in that order (*order 2* for the 30-ms tone and *orders 1* and 3 for the 60-ms tone). The modulations, in particular those for the 60-ms tone, show how order modulates MMN amplitude in both directions, consistent with a relative rather than absolute effect.

An examination of the ERPs to the repetitive sounds in each sequence revealed no significant impact of any of the within-subject variables supporting Todd et al.'s (2011) interpretation

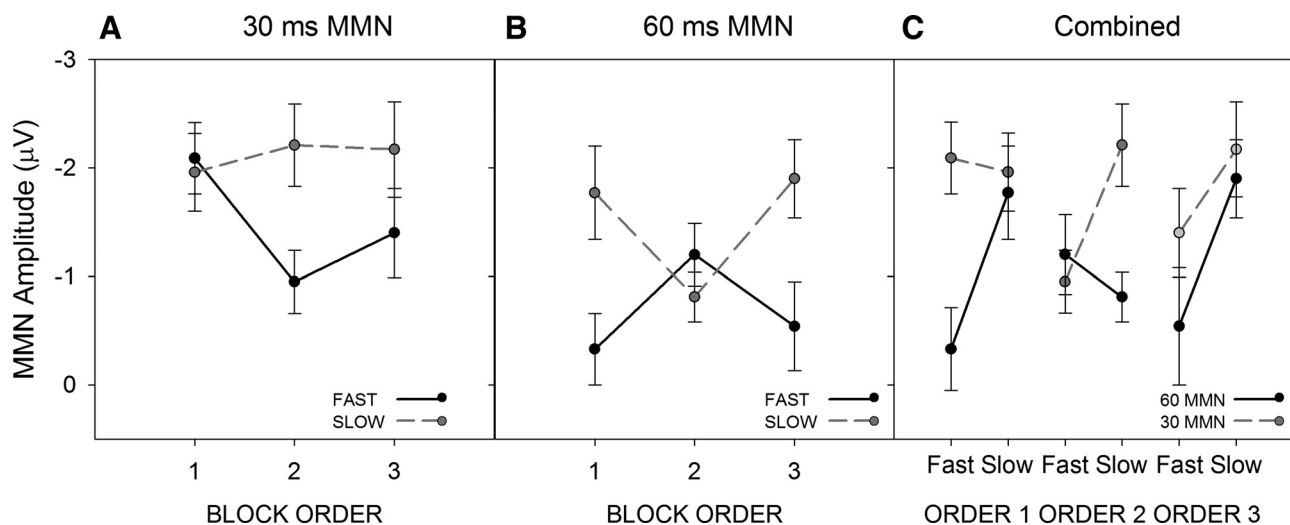


Fig. 3. Group-averaged mean amplitudes for MMN to 30-ms and 60-ms deviant sounds as a function of change speed and block order. A: fast and slow speed effects in the MMN to the 30-ms deviant across orders 1–3. B: fast and slow speed effects in the MMN to the 60-ms deviant across orders 1–3. C: interaction effects on MMN size between speed effects and tone type across orders 1–3. Error bars = Morey's (2008) corrected normalized within-subject standard errors.

that the origin of the effects, particularly the bias, is in response to the deviant tone.

DISCUSSION

Since its discovery by Näätänen et al. (1978), the MMN has not only found application in an increasing number of clinical and applied fields (Näätänen et al. 2012) but has also been central to revealing an increasingly sophisticated story about auditory processing by the brain (Näätänen et al. 2001). The early conception that the MMN reflects a simple mismatch between incoming sounds and a rapidly decaying trace of low-level auditory features has been replaced by the notion that it is integral to auditory scene analysis and reflects a learning process based on the success of multiple simultaneously active predictive models or “regularity representations” residing in long-term memory (Winkler and Cowan 2005).

In this study we have used a new multicontext, multitime-scale MMN paradigm revealing a bias in inferential processes underlying MMN. The results extend Todd et al.’s (2011) previous work by demonstrating 1) that a reliable index of the bias can be obtained in as little as 20 min; 2) within-subject evidence that the bias is anchored to the initial structure of the sequence and so reverses when tone order is reversed (results *order 1* vs. *order 2*); but 3) extraction of information about sequence structure over a much longer time course can abolish the bias (no bias when *order 1* is repeated in *order 3*). The data show that experience with sound can affect how subsequent evidence influences automatic perceptual inferences. Although lower-level processes like stimulus-specific adaptation have demonstrated sensitivity to event probability on multiple time-scales (Ulanovsky et al. 2003), we know of no mechanism by which it could account for the observed bias and, in particular, the disappearance of the bias in *order 3*. Given that ERP studies provide evidence that adaptation in subsets of neurons coding probability on multiple timescales can influence MMN size (e.g., Costa-Faidella et al. 2011), these factors must play a role in the phenomena we are measuring but seem inadequate to explain why the bias would be created, reverse, and then be overwritten over the three order conditions. A recent computational modeling study suggests that the process from which MMN derives reflects stored information about the conditional probability of observing a particular second stimulus at a certain latency after the first and that “MMN reflects, in a quantitative manner, the degree of violation of such transition probabilities” (Wacongne et al. 2012). The bias in the present data and that in Todd et al. (2011) indicate that such transition statistics are only part of the story and insufficient to account for these order-dependent phenomena.

Similar order-dependent biases observed in artificial grammar learning prompted the proposal that “adult learners have a prior probability, either innately or via early experience, that structures do not undergo rapid change without a strong contextual cue” (Gebhart et al. 2009). This prospect links well with recent conceptualizations of the MMN process and raises the possibility of a more top-down implementation of acquired knowledge. Winkler (2007) and Sussman (2007) discuss how mechanisms explaining the probability sensitivity of MMN in terms of the absolute strength of a memory trace for the standard, or its strength relative to a memory trace for the deviant, have been replaced by a regularity-violation interpre-

tation within which the MMN reflects learning about predictive confidence. Within this framework, our sequences can be conceptualized as setting up two competing models of the environment. *Model A* stipulates that the environment is best accounted for by the characteristics of the first standard (30 ms) tone. *Model B* reflects the competing expectation that the environment will match the characteristics of the second standard (60 ms) tone. Evidence for *models A* and *B* changes over time and at different rates in the fast and slow alternation sequences. The fact that deviations elicit larger MMN in slow than in fast change sequences for *model A* only implies an order-dependent differential impact of experience on predictive confidence. In other words, additional stability in the slow change sequence (and, conversely, instability in the fast change sequence) has an impact on MMN size for the violations of *model A* but not *model B*. It is as though the initial standard repetition in *order 1* (or *model A*) is accepted as a global structure and the dominant model. *Model B* becomes a local departure from this structure insensitive to modification by longer-term experience. In *order 2*, *model B* becomes the global structure/dominant model. The fact that the bias can be so readily reversed by a 5-min silence might then be explained by the silence preceding *order 2*, leading to the assumption that this sequence originates from an object different from that in *order 1*. By *order 3*, both models have played a role as global/dominant models and are recognized as equally likely (possibly as separate auditory objects), and therefore the bias is abolished. In this way, the bias creates a conservative preservation of stability in initial object perception, presumably until sufficient counterevidence is acquired.

A slightly different perspective emerges when considering the functional relevance of a prediction error signal. Model competition assumes that the bias occurs through preferential reevaluation of one prediction model (linked to the first standard). In contrast, an information value perspective assumes that the bias emerges because of the prediction error (linked to the first deviant). Prediction errors motivate learning by signaling when reality differs from inferences based on past experience. The goal of subsequent learning is to minimize the error (Friston 2005). This is achieved by enlisting resources that can provide more information on how to predict the event and/or on what the event predicts. The bias we observed is linked to the presentation order of tones. The first large prediction error signal is the MMN to the first encountered deviant (e.g., 60-ms tone in *order 1*, 30-ms tone in *order 2*). One perspective on the functional significance of MMN is that it signals that the environment departed significantly from the predicted state and this departure may be important. The best way to learn more about this event is to monitor its occurrence over a longer time frame. Over a longer sampling window, the 60-ms sound is less rare (or likewise the transition from 30-ms standard to 60-ms deviant is less rare) in fast change sequences than in slow change sequences, providing a probability-based explanation of why MMN amplitude to the 60-ms tone is modulated by speed of change. In contrast, the initial high repetition of the first encountered standard with no linked consequence may result in learned redundancy, failing to engage higher-order monitoring and, in turn, explaining why longer-term probability changes have no effect on MMN size.

Viewed from this information value perspective, the primacy bias is a failure to unlearn this redundancy, and so it

resembles latent inhibition attenuating learning about familiar inconsequential stimuli. If this is the case, it appears that the flexibility to be sensitive to variations in deviant probability at multiple temporal scales might be hampered by its implementation through a relatively simple learning mechanism. Our new finding that the primacy effect reverses after a 5-min break might also be consistent with this simple conditioning explanation, given that latent inhibition is known to be context sensitive. However, the complete disappearance of the primacy bias (i.e., the fact that speed modulates MMN for both tone durations) after a further break appears to be indicative of a more sophisticated meta-learning process. In particular, why would speed modulation for MMN to the 60-ms deviant, which occurs first in *order 1*, fail to occur when it subsequently occurs second in *order 2*, yet the speed modulation observed on the MMN to the 30-ms deviant that occurs first in *order 2* is also seen when it appears second in *order 3*? The disappearance of primacy bias suggests that by the time the third sequence pair occurs, higher-order learning promotes longer-term monitoring of all sounds to minimize prediction error in an environment with changing sound relevance (and/or multiple auditory objects).

Predictive confidence- and information value-based accounts have slightly different implications for learning. According to the former, the bias reflects how evidence is used to evaluate predictions about the environment. The latter implies that, even outside our awareness, the automatically determined information value of a sound will influence the level of engagement in monitoring its occurrence. In either case, it appears that with sufficient experience the MMN, and early auditory processing of unattended sequences, can reflect influences from brain processes with a hierarchy of temporal scales that enable quite sophisticated adaptation of learning processes to utilize higher-order patterning in predictions (Kiebel et al. 2008). At face value the bias appears to have methodological implications for studies that employ reversed-oddball control designs (e.g., Jacobsen and Schröger 2003). However, such designs generally hold standard probability stable for longer periods than that used here and do not alternate back and forth. Furthermore, it would appear from the outcomes of the present study that a period of silence between two opposing blocks is sufficient to “reset” or remove the former bias. At present we consider the implications minor unless a study runs the reverse-oddball sequences contiguously. The extent to which this is true depends on the outcomes of ongoing studies in our lab exploring the longevity of the effect in the face of countermanding evidence—that is, whether the bias holds for *model A* when it is followed by very long periods of stability in *model B*.

Our results suggest that the multicontext, multiscale MMN protocol provides a sensitive technique for probing the characteristics of perceptual learning about prediction at multiple temporal scales. For example, new studies could examine whether the primacy we demonstrated—one induced by an order-dependent consequential history in the present context—is also found with other types of prior bias (e.g., preexisting differences in stimulus salience). The way in which context change modulates learning also seems particularly suited to studying the role of long-term memory in the storage and retrieval of regularity representations in mismatch detection. Finally, all of these possibilities can be explored when

deviance is defined relative to recently acquired (e.g., Atienza and Cantero 2001) or long-term (e.g., Pulvermüller et al. 2001) knowledge, or potentially by higher-order relationships, such as various stimulus contingencies (e.g., Paavilainen et al. 2001; Tervaniemi et al. 1994), or when multiple simultaneous regularities are active (e.g., Horváth et al. 2001).

GRANTS

This research was supported by Project Grant 1002995 from the National Health and Medical Research Council of Australia.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: J.T., A.P., G.C., and A.H. conception and design of research; J.T. and L.W. analyzed data; J.T., A.P., and A.H. interpreted results of experiments; J.T. prepared figures; J.T. drafted manuscript; J.T., A.P., and A.H. edited and revised manuscript; J.T., A.P., L.W., G.C., and A.H. approved final version of manuscript; L.W. performed experiments.

REFERENCES

- Alho K, Woods DL, Algazi A, Knight RT, Näätänen R. Lesions of frontal cortex diminish the auditory mismatch negativity. *Electroencephalogr Clin Neurophysiol* 91: 353–362, 1994.
- Atienza M, Cantero JL. Complex sound processing during human REM sleep by recovering information from long-term memory as revealed by the mismatch negativity (MMN). *Brain Res* 901: 151–160, 2001.
- Costa-Faidella J, Grimm S, Slabu L, Diaz-Santaella F, Escera C. Multiple time scales of adaptation in the auditory system as revealed by human evoked potentials. *Psychophysiology* 48: 774–783, 2011.
- Escera C, Yago E, Corral MJ, Corbera S, Nunez MI. Attention capture by auditory significant stimuli: semantic analysis follows attention switching. *Eur J Neurosci* 18: 2408–2412, 2003.
- Friston K. Learning and inference in the brain. *Neural Netw* 16: 1325–1352, 2003.
- Friston K. A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* 360: 815–836, 2005.
- Friston K. Hierarchical models in the brain. *PLoS Comput Biol* 4: e1000211, 2008.
- Garrido MI, Kilner JM, Kiebel SJ, Friston KJ. Dynamic causal modeling of the response to frequency deviants. *J Neurophysiol* 101: 2620–2631, 2009.
- Gebhardt AL, Aslin RN, Newport EL. Changing structures in midstream: learning along the statistical garden path. *Cogn Sci* 33: 1087–1116, 2009.
- Giard MH, Perrin F, Pernier J, Bouchet P. Brain generators implicated in the processing of auditory stimulus deviance; a topographic event-related potential study. *Psychophysiology* 27: 627–640, 1990.
- Hall G, Honey RC. Contextual effects in conditioning, latent inhibition, and habituation: associative and retrieval functions of contextual cues. *J Exp Psychol Anim Behav Processes* 15: 232–241, 1989.
- Horváth J, Czigler I, Sussman E, Winkler I. Simultaneously active pre-attentive representations of local and global rules for sound sequences. *Cogn Brain Res* 12: 131–144, 2001.
- Jacobsen T, Schröger E. Measuring duration mismatch negativity. *Clin Neurophysiol* 114: 1133–1143, 2003.
- Kiebel SJ, Daunizeau J, Friston KJ. A hierarchy of time-scales and the brain. *PLoS Comput Biol* 4: e1000209, 2008.
- Kujala T, Tervaniemi M, Schröger E. The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. *Biol Psychol* 74: 1–19, 2007.
- Lubow RE, Gewirtz JC. Latent inhibition in humans: data, theory, and implications for schizophrenia. *Psychol Bull* 117: 87–103, 1995.
- Morey RD. Confidence intervals from normalized data: a correction to Cousineau (2005). *Tutor Quant Methods Psychol* 4: 61–64, 2008.
- Näätänen R. The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behav Brain Sci* 13: 201–288, 1990.

- Näätänen R.** *Attention and Brain Function*. Hillsdale, NJ: Erlbaum, 1992.
- Näätänen R, Gaillard AW, Mäntysalo S.** Early selective attention effect on evoked potential reinterpreted. *Acta Psychol (Amst)* 42: 313–329, 1978.
- Näätänen R, Kujala T, Escera C, Baldeweg T, Kreegipuu K, Carlsen S, Ponton C.** The mismatch negativity (MMN)—a unique window to disturbed central auditory processing in ageing and different clinical conditions. *Clin Neurophysiol* 123: 424–458, 2012.
- Näätänen R, Michie PT.** Early selective attention effects on the evoked potential: a critical review and reinterpretation. *Biol Psychol* 8: 81–136, 1979.
- Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I.** “Primitive intelligence” in the auditory cortex. *Trends Neurosci* 24: 283–288, 2001.
- Paavilainen P, Simola J, Jaramillo M, Näätänen R, Winkler I.** Preattentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity (MMN). *Psychophysiology* 38: 359–365, 2001.
- Pulvermuller F, Kujala T, Shtyrov Y, Simola J, Tiitinen H, Alku P, Alho K, Martinkauppi S, Ilmoniemi RJ, Näätänen R.** Memory traces for words as revealed by the mismatch negativity. *Neuroimage* 14: 607–616, 2001.
- Schmidt A, Diaconescu AO, Kommer M, Friston KJ, Stephan KE, Vollenweider FX.** Modeling ketamine effects on synaptic plasticity during the mismatch negativity. *Cereb Cortex* (August 8, 2012). doi:10.1093/cercor/bhs238.
- Semlitsch HV, Anderer P, Schuster P, Presslich O.** A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology* 23: 695–703, 1986.
- Sussman ES.** A new view on the MMN and attention debate: the role of context in processing auditory events. *J Psychophysiol* 21: 164–175, 2007.
- Tervaniemi M, Maury S, Näätänen R.** Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *Neuroreport* 5: 844–846, 1994.
- Todd J, Provost A, Cooper G.** Lasting first impressions: a conservative bias in automatic filters of the acoustic environment. *Neuropsychologia* 49: 3399–3405, 2011.
- Ulanovsky N, Las L, Nelken I.** Processing of low-probability sounds by cortical neurons. *Nat Neurosci* 6: 391–398, 2003.
- Wacongne C, Changeux JP, Dehaene S.** A neural model of predictive coding accounting for the mismatch negativity. *J Neurosci* 32: 3665–3678, 2012.
- Winkler I.** Interpreting the mismatch negativity. *J Psychophysiol* 21: 147–163, 2007.
- Winkler I, Cowan N.** From sensory to long-term memory: evidence from auditory memory reactivation studies. *Exp Psychol* 52: 3–20, 2005.
- Winkler I, Karmos G, Näätänen R.** Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Res* 742: 239–252, 1996.

