

Evidence for a hierarchy of predictions and prediction errors in human cortex

Catherine Wacongnea, Etienne Labyta, Virginie van Wassenhovea, Tristan Bekinschteind, Lionel Naccachee, Stanislas Dehaene

▶ **To cite this version:**

Catherine Wacongnea, Etienne Labyta, Virginie van Wassenhovea, Tristan Bekinschteind, Lionel Naccachee, et al.. Evidence for a hierarchy of predictions and prediction errors in human cortex. Proceedings of the National Academy of Sciences of the United States of America, 2011, 108 (51), pp.20754-20759. 10.1073/pnas.1117807108 . cea-00843043

HAL Id: cea-00843043

https://cea.hal.science/cea-00843043

Submitted on 6 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evidence for a hierarchy of predictions and prediction errors in human cortex

Catherine Wacongne[a,b,c,1,2], Etienne Labyt[a,b,c,1], Virginie van Wassenhove[a,b,c], Tristan Bekinschtein[d], Lionel Naccache[e,f], and Stanislas Dehaene[a,b,c,g,2]

[a]Cognitive Neuroimaging Unit, Institut National de la Santé et de la Recherche Médicale, U992, F-91191 Gif/Yvette, France; [b]NeuroSpin Center, Institute of BioImaging Commissariat à l'Energie Atomique, F-91191 Gif/Yvette, France; [c]Université Paris 11, Orsay, France; [d]Cognition and Brain Sciences Unit, Medical Research Council, Cambridge, United Kingdom; [e]Institut du Cerveau et de la Moelle Épinière Research Center, Institut National de la Santé et de la Recherche Médicale, U975 Paris, France; [f]Faculté de Médecine Pitié-Salpêtrière, Université Paris 6, Paris, France; and [g]Collège de France, F-75005 Paris, France

Contributed by Stanislas Dehaene, October 30, 2011 (sent for review June 24, 2011)

According to hierarchical predictive coding models, the cortex constantly generates predictions of incoming stimuli at multiple levels of processing. Responses to auditory mismatches and omissions are interpreted as reflecting the prediction error when these predictions are violated. An alternative interpretation, however, is that neurons passively adapt to repeated stimuli. We separated these alternative interpretations by designing a hierarchical auditory novelty paradigm and recording human EEG and magnetoencephalographic (MEG) responses to mismatching or omitted stimuli. In the crucial condition, participants listened to frequent series of four identical tones followed by a fifth different tone, which generates a mismatch response. Because this response itself is frequent and expected, the hierarchical predictive coding hypothesis suggests that it should be cancelled out by a higher-order prediction. Three consequences ensue. First, the mismatch response should be larger when it is unexpected than when it is expected. Second, a perfectly monotonic sequence of five identical tones should now elicit a higher-order novelty response. Third, omitting the fifth tone should reveal the brain's hierarchical predictions. The rationale here is that, when a deviant tone is expected, its omission represents a violation of two expectations: a local prediction of a tone plus a hierarchically higher expectation of its deviancy. Thus, such an omission should induce a greater prediction error than when a standard tone is expected. Simultaneous EEE- magnetoencephalographic recordings verify those predictions and thus strongly support the predictive coding hypothesis. Higher-order predictions appear to be generated in multiple areas of frontal and associative cortices.

mismatch negativity | P300 component

According to the predictive coding hypothesis, the architecture of the cortex implements a top-down prediction algorithm that constantly anticipates incoming sensory stimuli. Each cortical area houses an internal model of the environment, which is generated by compiling the statistical regularities that govern past inputs. This model is used to generate top-down predictions that are compared with novel incoming inputs. Only the difference, called the "prediction error," is transmitted to higher cortical stages, where it can be used to adjust the internal model. Importantly, this process can be organized hierarchically (1–4), so that the prediction error arising from a given area in turn serves as the input to the next area. The outcome is an active system that constantly updates models of its environment at multiple hierarchically organized levels.

Although considerable evidence supporting predictive coding has been provided at the perceptual level (e.g., 5–10), here we specifically set out to test the notion of a hierarchy of predictions, using a variant of the classical auditory violation paradigm (11).

When a rare sound is introduced within a sequence of repeated frequent sounds, it elicits a novelty response in the event-related potential, which has been termed the "mismatch negativity" (MMN) (12). This response is interpreted, within the predictive coding framework, as reflecting the violation of a prediction: The MMN would directly reflect the cortical prediction error signal (4–6, 13). This interpretation is supported by sophisticated modeling studies which suggest that the MMN can be accounted for only by postulating a top-down predictive contribution (5–8). However, an alternative interpretation exists, whereby the MMN would reflect solely a passive, bottom-up process of adaptation to the repeated stimuli (14, 15). According to this interpretation, the repeated stimulus, by constantly stimulating the same afferent pathways, leads to synaptic adaptation and therefore to reduced activation. The rare stimulus, by contrast, activates fresh afferents that have not been adapted, resulting in a distinctly larger response. Thus, mismatch responses could reflect adaptation rather than predictive coding.

How could adaptation and predictive models be distinguished? An interesting variant of the mismatch paradigms consists of omitting the expected stimulus, rather than replacing it with another stimulus. It is a rather remarkable fact that the auditory cortex generates extensive responses locked to the absence of a predictable sound. This omission response can be detected by a variety of methods, including event-related potentials (ERPs) (16), magnetoencephalography (MEG) (17), and intracranial recordings (18). Omission responses fit quite naturally within the predictive coding framework: If stimulus-evoked brain activity indexes the difference between a sensory signal and its top-down prediction, then, when the sensory signal is omitted, the evoked activity should reflect the pure prediction signal within the same cortical area (8, 19, 20). Omission responses seem more difficult to explain within the adaptation framework. They might reflect the automatic rebound of a cortical oscillator entrained by the rhythm of the past stimuli (14), but this hypothesis meets difficulties in explaining why omission responses are still present in nonrhythmic paradigms, for instance when the second tone of a pair is omitted (8, 18–20).

Omission responses therefore might constitute a critical test of the predictive coding framework. Here, we capitalized on omission responses, combined with a hierarchical violation-of-expectation paradigm, to demonstrate that auditory signals are indeed submitted to multiple, hierarchically organized stages of top-down prediction. We used a recently introduced auditory paradigm that can dissociate two types of predictions, based on local probabilities versus global rules (11). In a given block, a frequent sequence of five tones is presented (in 75% of trials), interspersed with rare violations (in 15%) in which the frequency of the fifth tone deviates from the expected, and with rare

Author contributions: E.L., T.B., L.N., and S.D. designed research; E.L. performed research; V.v.W. contributed new reagents/analytic tools; C.W., E.L., and S.D. analyzed data; and C.W., L.N., and S.D. wrote the paper.

The authors declare no conflict of interest.

[1]C.W. and E.L. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: stanislas.dehaene@cea.fr or catherine.wacongne@gmail.com.

Downloaded at CEA/GRENOBLE on July 6, 2020

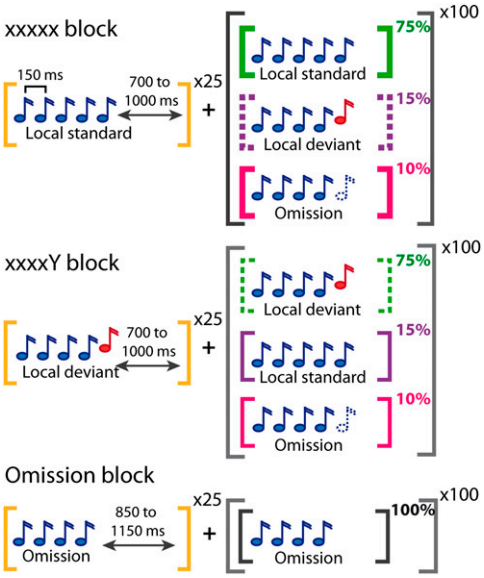


Fig. 1. Experimental design. Three auditory stimuli could be presented: local standards (a series of five identical tones, denoted xxxxx), local deviants (four identical tones followed by a different tone; denoted xxxxY), and omissions (four identical tones; denoted xxxx).These stimuli were presented in three types of blocks in which one series was presented with a high frequency (initially 100%, then 75%) and the other series were rare. This design thus separated the local deviancy of the fifth sound from the global deviance of the entire sequence and also allowed us to probe whether the omission effect differed when a standard or a deviant tone was expected.

omissions (in 10%) in which the fifth tone is simply omitted (Fig. 1). Crucially, on some blocks the frequent sequence is of the "xxxxY" type, i.e., four identical tones followed by a distinct one. ERP recordings reveal that the fifth, locally deviant tone, although fully expected, still elicits an MMN. However, only the rare violation sequence, which contains the five identical tones "xxxxx", elicits a distinct and later novelty response, the "late positive complex" or P3b wave (11). In the context of hierarchical predictive coding models, this observation can be interpreted as reflecting a "violation of a violation": The monotonous "xxxxx" sequence is surprising because it fails to contain the fifth deviant tone, which normally generates an MMN. Hierarchical predictive coding models thus hypothesize two levels of predictions in this situation: A first low-level expectation, based on local transition probabilities, incorrectly predicts a fifth "x" tone after the first four "xxxx," thus generating an MMN, whereas a second, higher-level expectation, based on the knowledge of the overall "xxxxY" rule, cancels the surprise elicited by the first level.

A simple prediction ensues. When we omit the fifth sound, thus presenting a series of four identical tones "xxxx," the observed omission response should vary according to the expectation induced by the overall context. On "xxxxY" blocks, where two successive predictions are generated, we should observe a large event-related response to omission, composed of superimposed waves corresponding to the predictions of the "x" tone and of the MMN. On "xxxxx" blocks, however, only the first of these predictions should exist, and therefore the event-related response to omission should be significantly smaller. We tested this hypothesis by recording simultaneous EEG and MEG signals evoked by these stimuli, relative to a low-level control in which the omission of the fifth tone was entirely expected.

Results

We first examined our data for the presence of a local mismatch response evoked by the deviance of the fifth tone. Cluster analysis, implemented in FieldTrip software, was used to identify clusters of neighboring sensors where a significant difference between local standards and local deviants was seen over several consecutive time points (*Methods*). This analysis was performed separately for EEG sensors, MEG magnetometers (MEGm), and MEG gradiometers (MEGg), using the root mean square of the two orthogonal gradiometers available at each position (Fig. 2). The results revealed an early effect of local deviance, peaking at around 120 ms after the onset of the fifth deviant tone, and which reached corrected significance for each sensor type (EEG: range of first significant window 85–150 ms, $P = 0.002$; MEGg, 80–170 ms, $P < 0.0001$; MEGm, 82–150 ms, $P = 0.02$).

Fig. 2 shows the corresponding topography and time course of a relevant sensor for each block type. In both the xxxxx and xxxxY blocks, the response to local deviance has the topography of a classical mismatch field, with bilateral responses over the left and right temporal regions. The mismatch response was significant and with the same sign in each block (block xxxxx: EEG, 85–180 ms, $P = 0.002$; MEGg, 84–150 ms, $P < 0.0001$; MEGm, 76–130 ms, $P < 0.0001$; block xxxxY: MEGg, 80–140 ms, $P < 0.0001$). This response thus indexes a first local level of auditory novelty detection which is blind to global context. Indeed, in xxxxY blocks, although the deviant "Y" sound could be fully expected, the mismatch response to the final Y tone remained. Nevertheless the MMN amplitude was reduced on xxxxY compared with xxxxx blocks (Fig. 2*C*; EEG: 134–190 ms, $P = 0.014$; MEGg, 103–700 ms, $P < 0.00001$; MEGm, 95–210 ms, $P = 0.006$).

We then examined the presence of a second-level novelty response, dependent on the frequency of the overall sequence rather than of individual tones. On all sensor types, rare sequences differed from frequent sequences on a later time window than the MMN (EEG, 327–540 ms, $P < 0.00001$; MEGg, 103–600 ms, $P < 0.00001$; MEGm, 275–600 ms, $P < 0.00001$). Note that on xxxxY blocks this higher-order novelty response was elicited by the monotonic but unexpected xxxxx stimulus relative to the frequent xxxxY stimulus, leading to a complete inversion of the classical mismatch response (Fig. 2*B*). On such trials, a sequence of two successive novelty events, hereafter termed "local" and "global" effects, was thus revealed. In EEG, as previously described, the second, global effect has the classical topography and latency of the P3b component, which differs strongly from the MMN (Fig. 2 *A* and *C*). Surprisingly, in MEG, these two events have very similar topographies: Both are dominated by bilateral responses over temporal cortices.

The next step was to examine omission responses. The omission effect was computed by recording the brain responses to rare omissions (presentation of only four identical tones instead of five) separately within xxxxx and xxxxY blocks and comparing theses responses with the responses to a block where only sequences of four identical tones were presented (expected omissions). The results showed an early effect of unexpected omission peaking around 100 ms after the onset of the omitted tone (i.e., 250 ms after the onset of the fourth tone), with a topography similar to the MMN topography for all sensor types (Fig. 2 *D* and E). The early latency of this peak response to an absent stimulus is consistent with the hypothesis that this response corresponds to an unfulfilled prediction. The omission effect was significant in both block types (xxxxx blocks: significant only for MEGg, 76–200 ms, $P < 0.0001$; xxxxY blocks: EEG, 104–160 ms, $P = 0.022$; MEGg, 150–200 ms, $P < 0.0001$; MEGm, 150–200 ms, $P = 0.002$). In both blocks, the difference between rare versus expected omissions also was significant in a later time window with a topography similar to the above global effect (xxxxx blocks: EEG, 425–440 ms, $P = 0.032$; MEGg, 327–500 ms, $P < 0.0001$; MEGm, 234–500 ms, $P < 0.0001$; xxxxY blocks: MEGg, 134–500 ms, $P < 0.0001$; MEGm, 272–500 ms, $P < 0.0001$). Thus, the omission effect consists of a sequence of early and late responses, the latter coinciding with the P3b-like global

Downloaded at CEA/GRENOBLE on July 6, 2020

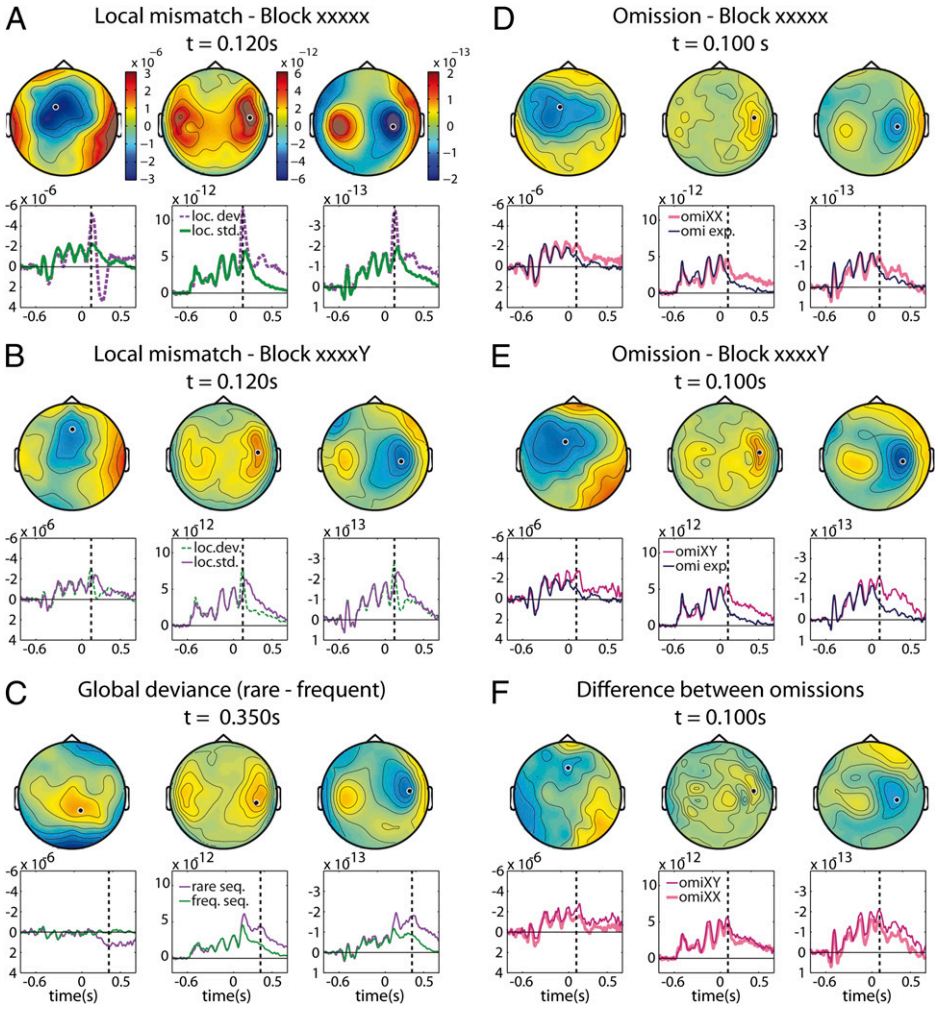


Fig. 2. Sensor-level topography and time course of the brain responses to distinct forms of novelty. In each panel, the three topographies show the spatial distribution, on a top view of the scalp, of EEG signals (*Left*), MEG gradiometers (norm; *Middle*), and MEG magnetometers (*Right*) at the time indicated by the vertical dotted line in the graphs. Graphs show the time course of these signals as recorded from an individual sensor (marked by a dot on the corresponding topographical map). (*A* and *B*) Effects of local mismatch: Bilateral auditory areas show a rapid response to the fifth deviant tone, whether it is rare (xxxxx blocks, in *A*) or frequent (xxxxY blocks, in *B*). (*C*) Effect of global deviance: A temporally and spatially extended response, corresponding to the P3b in ERPs, is seen in response to rare sequences. (*D–F*) Responses to omission of the fifth tone. The brain responds to omission by emitting a sharp response whose amplitude is smaller when a standard is expected (*D*) than when a deviant is expected (*E*), resulting in a significant difference (*F*). omiXX, rare omissions in the xxxxx block; omi XY, rare omissions in the xxxxy block; omi exp, expected omissions in the control omission block.

effect observed in all rare conditions (rare sequences and rare omissions). This finding is consistent with the hypothesis that this global effect is a correlate of detection of any deviance from the rule currently entertained in working memory.

Finally, we compared the amplitude of the omission effect between the xxxxx and xxxxY blocks, testing the prediction, unique to hierarchical predictive coding models, that the early omission effect should be bigger on xxxxY blocks when a deviant stimulus is expected. The difference between omissions is plotted in Fig. 2*F*. The early omission response was significantly higher in amplitude for xxxxY blocks than for xxxxx blocks (EEG: 109–130 ms, $P = 0.03$, and MEGg, 68–80 ms, $P = 0.042$). Fig. 2 *D* and *E* show the topography of the difference between omissions, indicating a slightly more anterior source for the MEG omission effect on xxxxY than on xxxxx blocks.

We used a minimal norm estimation method to reconstruct distributed cortical sources based on MEG data (similar results were obtained when combining EEG and MEG data). The results dissociated regions sensitive to local and global regularities (Fig. 3). Maximal responses to incoming tones arose from bilateral superior temporal cortices, in the vicinity of Heschl's gyrus and the underlying segment of the right superior temporal gyrus. These regions also showed the maximal response to local deviants (maximum *z*-score, in Talairach coordinates: right hemisphere, x = 45 mm, $y = -19$ mm, $z = 13$ mm; left hemisphere, $x = -48$ mm, $y = -16$ mm, $z = 13$ mm) and to omissions (same sources) (Fig. 3*C*). Another set of regions did not respond strongly to incoming tones but responded in a categorical manner to global deviance. The activated sites were highly distributed in bilateral anterior and posterior superior temporal gyri, supramarginal gyri, dorsolateral, inferior, polar, and ventromedial prefrontal cortices, anterior cingulate, and the superior parts of the precentral and postcentral gyri (Fig. 3*D*). As shown in Fig. 3*B*, their activity was minimal for frequent sequences but converged toward a higher and temporally sustained level of activity whenever a rare sequence or omission was presented. Finally, we examined the cortical origins of the difference between the omission effects on xxxxY versus xxxxx blocks. Consistent with the trend seen on sensor-level topographies, the maximal difference between omissions originated from a more anterior

Downloaded at CEA/GRENOBLE on July 6, 2020

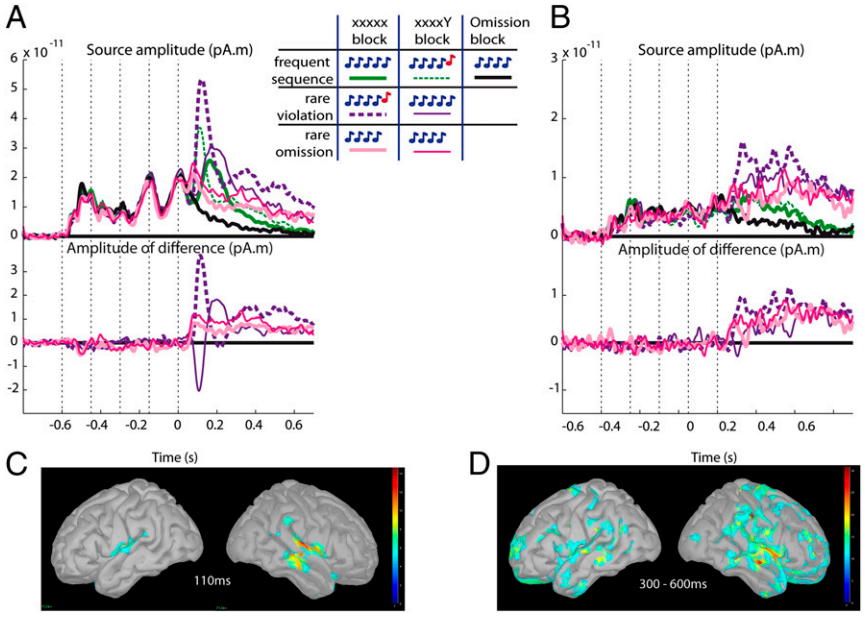


Fig. 3. Source modeling of the effects. The reconstructed signals from the right auditory cortex (*A*) and right precentral cortex (*B*) are shown. Upper panels show the signals in each of the seven experimental conditions, and the lower panels show subtractions of each rare sequence (identified by the same color as in the upper plot) minus the frequent sequence of the same block, and of omissions in each block type (same color as in the upper plot) minus the expected omission from the control block. (*C* and *D*) The *z*-score–corrected source reconstruction of the local effect (*C*) and effect of deviance from global rule (*D*) on inflated cortex. Local effect (*C*) is the average effect of deviance from local regularity over both types of blocks at time = 110 ms. Global effect (*D*) is the average of contrasts between rare and frequent trials, at constant stimulus (rare violation – frequent sequences and rare omissions – expected omissions), averaged over the late period of the trials (300–600 ms). The source in auditory cortex shows a sharp and rapid response to local deviance, followed by a late and sustained response to global deviance. The source in precental cortex shows only the late sustained response. Both effects are present on omission trials (red/pink curves), with greater initial response to omissions in xxxxY blocks (thin red curve) than in xxxxx blocks (thick pink curve).

temporal region than either the MMN or the basic omission effect. This region was lateralized to the right hemisphere (x = 53 mm, y = −2 mm, z = 4 mm).

Discussion

By recording ERPs and magnetic fields while manipulating and violating the participants' auditory expectations at two distinct levels, we obtained direct evidence that an active, predictive, and hierarchical system underlies the brain's response to auditory stimuli.

First, we replicated the earlier finding of a double dissociation between the early MMN and a later, temporally extended and distributed response (P3b) (11). The MMN was sensitive to local violations of transition probabilities and was essentially blind to higher-order regularities, because it continued to be evoked, at a reduced level, by a fifth deviant tone that could be expected (in the xxxxY blocks). Contrariwise, we observed a late (~300 ms) divergence which reflected solely the deviance of the overall sequence rather than of its individual component tones.

Although MEG and EEG revealed functionally and temporally similar responses, their spatial pattern diverged. The EEG topographies differed strongly for the MMN and P3b stages, but in MEG these two stages showed similar topographies involving mainly temporal sources. This difference in sensitivity to sources between MEG and EEG stresses the interest of combining the two methods (21). Overall, the results suggest an initial stage confined to temporal cortex and a later stage at which this activity is amplified and expands into distributed additional regions, particularly in prefrontal and parietal cortices (11). The weak influence of the latter sources on MEG topography might result from their multiplicity and dispersion.

It is important to note that the previous results by Bekinschtein et al. (11) in a similar paradigm were observed in the context of a counting task where participants counted the rare stimuli. Thus, the P3b response that they observed with rare compared with frequent stimuli could have arisen from the counting process, which occurred with global deviant stimuli but not with frequent stimuli. By contrast, in the present trials the participants were instructed only to attend to the stimuli, not to count them. Thus, our results show that the counting task is not necessary and that the late P3b response reflects, at least in part, the response of a higher-order novelty-sensitive system.

Our findings refine earlier results by showing that the local and global effects are not fully independent (11) but interact in an early time window. Specifically, the local mismatch response was significantly smaller in xxxxY blocks than in xxxxx blocks. There are at least two interpretations of this effect. First, the effect could be caused solely by a difference in transition probabilities. Indeed, MMN amplitude decreases when the probability of the deviant increases; in the blocks where the xxxxY sequence is frequent, the transition probability x→Y is necessarily higher relative to xxxxx blocks. However, this effect also is fully consistent with the hierarchical predictive coding hypothesis, which predicts that on xxxxY blocks a second-level prediction can be used to cancel out partially the first-order error novelty response to the expected deviant sound Y.

Although the theoretical implications of this early modulation of the initial mismatch response are ambiguous, the complete inversion of the mismatch signals observed in a later time window argues strongly for a hierarchical process. Indeed, on xxxxY blocks, the xxxxx stimulus becomes the rare stimulus and elicits a P3b-like brain response. The fact that a stimulus that consists solely of a repetition of five identical tones (xxxxx) can elicit a novelty signal if the participants expected a different sequence is in itself highly suggestive that the brain operates as a multilevel predictive system sensitive to prediction errors.

Having established the existence of a hierarchy of at least two novelty systems, we used sound omissions to provide a stronger

Downloaded at CEA/GRENOBLE on July 6, 2020

test of the hypothesis that these novelty responses arise from active prediction systems rather than from passive neural adaptation (4–6, 13, 19). Our results confirm earlier findings that the omission of an expected tone leads to a time-locked brain response, which is easily detectable by MEG and EEG and has a topography similar to the original evoked response (16–18). Furthermore, our design tested the prediction, unique to the hierarchical predictive coding framework, that the omission response should vary with the context. Specifically, this framework supposes that evoked responses reflect a series of prediction errors indexing the difference between the incoming signal and its prediction at successive hierarchical levels. Accordingly, when the incoming signal is omitted, brain responses should reflect solely the predictive signals and how they vary depending on the current context (19). In agreement with this notion, we observed that the brain response to an omitted signal following a strictly identical series of four tones varied depending on whether the participants expected the fifth tone to be identical to or different from the preceding ones (xxxxx versus xxxxY blocks). A significantly larger omission response was observed on xxxxY blocks. This difference between the two omissions effects is exactly as predicted by the hierarchical view: On xxxxY blocks, an additional higher-order predictive signal is needed to cancel out the predictable MMN inevitably arising from the novelty of the fifth tone.

Passive adaptation models of mismatch responses attempt to account for omission responses in terms of an oscillatory or rebound response, the result of an entrainment of brain oscillators by the rhythm of the preceding stimuli (14). This hypothesis, however, cannot explain our observation of a larger omission response on xxxxY than on xxxxx blocks. Under the adaptation interpretation, we would have expected either a constant entrainment by the four preceding tones, and hence a constant omission response, or, if anything, a larger entrainment on the regular xxxxx blocks than on the xxxxY blocks, where the fifth item interrupts the rhythm of the first four—exactly the contrary of what was observed. Our results therefore are very difficult to explain with the adaptation hypothesis alone. They do not rule out the possibility that sensory adaptation may exist but only prove that it cannot be the only mechanism at work, as also argued by others (10). Note also that the size of the omission effect goes in the direction opposite that of the MMN: As described above, the MMN is larger on xxxxx than on xxxxY blocks, but the omission effect is larger on xxxxY than on xxxxx blocks. This inverse relation between the magnitude of the MMN and of the omission response is exactly as expected from a hierarchy of predictive systems but cannot be accommodated easily by a single process of novelty detection. In particular, it rules out the possibility that the observed modulations are caused by one of the blocks being intrinsically more interesting, motivating, or attention-grabbing.

The latencies of the observed novelty responses also are indicative of a predictive system. First, the timing of the omission response arises too early to correspond to a rebound of a putative oscillation induced by preceding stimuli. As shown in Fig. 3 (*Left*, pink curves), the cortical response on omitted xxxx trials does not consist of a series of five equally spaced peaks, as would be predicted by the oscillatory adaptation/rebound model. Rather, omission responses arise earlier than the MMN, which itself arises earlier than the response to an expected tone. This temporal order is the opposite of what would be expected from an ascending feedforward system, where the stimulus first must be processed bottom-up before its departure from the familiar can be detected. It is, however, in full agreement with a hierarchical predictive system in which first the presence and then the precise identity of the incoming tones are predicted successively in advance of the actual stimulus.

The fact that the omission effect is equally early on xxxxx and xxxxY blocks may seem counterintuitive: According to a hierarchical model, one might expect a sequence of two successive omissions effects. In reality, however, although two predictions are indeed assumed, both must come quite early if they are to act as predictors that cancel out the effects of the incoming signals. Predictive coding models thus predict that, during the xxxxx block, the omission effect must arise simultaneously with the earliest activation evoked by the fifth stimulus. Furthermore, during the xxxxY blocks, an additional second-order omission effect must arise before or simultaneously with the MMN to act as a predictor of it. The timing of the observed effects is compatible with these hypotheses. Furthermore, their topography suggests that the second-order omission is generated at a distinct cortical site about 2 cm more anterior in temporal cortex.

In summary, in agreement with recent theoretical models of cortical architecture (1–4), our findings suggest a hierarchical organization consisting of several successive prediction and novelty-detection systems. The present paradigm, combined with MEG, EEG, or intracranial recordings (11), dissociates at least two levels of prediction: The MMN responds to local auditory predictions, and the later P3b responds to more global and integrative violations of expectations. In that respect, our observations add to a growing number of dissociations of these two systems. Bekinschtein et al. (11) demonstrated that the early MMN resists visual distraction and remains present in subjects who are not conscious of the rule linking the five successive tones as well as in patients with in coma and vegetative state (22), whereas none of these properties hold for the global P3b, which therefore seems to index a conscious process. Prior ERP and functional MRI evidence confirms that the superior temporal region can respond to novel stimuli that are subliminal and fail to be detected (23, 24), whereas a much broader frontoparietal network, indexed by the P3b, underlies conscious detection (24–26). Pegado et al. (27) observed that, when the delay between tones is prolonged up to several seconds, the MMN is drastically reduced, but the P3b remains constant in size, although slightly delayed, in correspondence with the participants' preserved capacity to detect the violations. Ritter et al. (28) found, as we did, that the MMN remains but the P3b vanishes in a context where the local auditory deviance is fully predictable (in their case, because it is systematically preceded by a visual cue presented 600 ms earlier).

We conclude that detection of auditory novelty appears to be organized in several stages (29). The MMN reflects the operation of a temporally and conceptually limited prediction system that uses the recent past to predict the present, based solely on a compilation of the probabilities of the stimuli and their transitions. The auditory prediction underlying the MMN may rely on several recent stimuli (30), but it uses only a limited time window (27, 31) and is blind to the global overall rule or pattern followed by the stimuli (11). The extraction of such rules and the detection of their violations involve a later, more distributed predictive system (11, 24, 28, 29). Frequently, the operation of both these systems is undetectable, because their sole effect is to reduce or cancel the responses evoked by predictable sensory stimuli. The omission paradigm, by unveiling these responses, provides a flexible method to dissect the brain's multiple top-down expectation systems.

Methods

Subjects. Ten healthy subjects (mean age 26 ± 4.5 y; five females) with no known neurological or psychiatric pathology were studied. All subjects gave written informed consent to participate to this study, which was approved by the local Ethics Committee.

Auditory Stimuli. Two tones composed of three superimposed sine waves (350, 700, and 1,400 Hz, tone A; or 500 Hz, 1,000 Hz, and 2,000 Hz, tone B)

Downloaded at CEA/GRENOBLE on July 6, 2020

were synthesized. The tones were 50 ms long, with 7-ms rise and fall times. Series of four or five such tones were presented via headphones with an intensity of 70 dB and a 150-ms stimulus onset asynchrony. The series could comprise five identical tones (local standard, denoted xxxxx), four identical tones and a fifth different tone (local deviant, denoted xxxxY), or only four identical tones (omission, denoted xxxx). Series were presented in semi-randomized blocks of ~3-min duration, separated by silences of variable duration (700–1,000 ms), during which one series was designated as frequent and the other as rare (Fig. 1). Each block started with 25 frequent series of sounds to establish the global regularity (global rule). Of the next 100 occurrences, 75% were the frequent series, 15% the rare series, and 10% the omission series. A separate block contained 125 presentations of the omission sequence (expected omissions). Each participant received a total of 14 blocks of 125 trials each (three replications of the four rules xxxxY and xxxxx with either x = A and Y = B, or x = B and Y = A, plus two xxxx omission blocks with x = A for one and x = B for the other). All stimuli were presented using E prime v1.2 (Psychology Software Tools).

Simultaneous EEE-MEG Recordings. Measurements were carried out with the Elekta Neuromag NeuroSpin system (Elekta Neuromag Oy), which comprises 204 planar gradiometers and 102 magnetometers in a helmet-shaped array. The built-in EEG system (64 electrodes) was used to record EEG and MEG simultaneously. An electrode on the tip of the nose was used as EEG reference. ECG and electrooculogram (EOG) (horizontal and vertical) were recorded simultaneously as auxiliary channels. MEG, EEG, and auxiliary channels were low-pass filtered at 330 Hz, high-pass filtered at 0.1 Hz, and sampled at 1 KHz. The head position with respect to the sensor array was determined by four head-position indicator coils attached to the scalp. The locations of the coils and EEG electrode positions were digitized with respect to three anatomical landmarks (nasion and preauricular points) with a 3D digitizer (Polhemus Isotrak system). Then, head position with respect to the device origin was acquired before each block. Subjects were asked to keep their eyes open, to avoid eyes movements by fixating a cross, and were constantly reminded to pay attention to the auditory stimuli. At the end of the recording, a questionnaire assessed which regularities and violation types had been detected. All subjects reported detecting both rare sound series and omissions.

Data Analysis. Signal space separation correction, head movement compensation, and bad channels correction were applied using the MaxFilter Software (Elekta Neuromag). Principal components analysis was used to remove EKG and EOG artifacts. With Fieldtrip software (http://fieldtrip.fcdonders.nl/), trials were epoched from 200 ms before to 1,300 ms after the onset of the first sound, low-pass filtered at 40 Hz, and baseline corrected using the first 200 ms of the epoch. After visual rejection of artifacts, the trials were averaged per condition and per subject. The latitudinal and longitudinal gradiometers were combined by computing the mean square root of the signals at each sensor position.

Cluster-based statistics were performed using Fieldtrip software. Statistics were computed between 50 and 250 ms for mismatch and omission effects, between 50 and 700 ms for the global effect, and between 50 and 500 ms after the onset of the omitted sound for late omission effect. The threshold was fixed to $P = 0.05$, corrected for the size of the search space (time and sensors). We report only the most significant clusters for each sensor type.

Source Reconstruction. Anatomical T1-weighted MRIs were obtained for each participant after the MEG experiment with a 3-T Siemens MRI scanner, with a resolution of $1 \times 1 \times 1.1$ mm. Head-position indicators and the digitized head shape were used for the coregistration of the anatomical images with the MEG signals. Gray and white matter then was segmented using BrainVisa/Anatomist software package (http://brainvisa.info/). The scalp and cortical surfaces were reconstructed for each subject using BrainStorm software (http://neuroimage.usc.edu/brainstorm/). Models of the cortex and of the head were used to estimate the current-source density distribution over the cortical surface. The forward model was computed using an overlapping-spheres analytical model. The inverse model was constrained to a minimum-norm solution (weighted minimum-norm current estimate, wMNE). Sources were reconstructed at each time point. For each subject, the sources were then projected to a standard anatomical template (MNI). Contrasts between conditions were normalized using *z*-score normalization.

ACKNOWLEDGMENTS. We thank Ghislaine Dehaene-Lambertz, Lucie Hertz-Pannier, Caroline Huron, Antoinette Jobert, Andreas Kleinschmidt, and the NeuroSpin infrastructure groups for their help in subject recruitment and testing; Marco Buiatti, Lucie Charles, Sebastien Marti, and François Tadel for help in data analysis; and Jean-Pierre Changeux, Ghislaine Dehaene-Lambertz, Alain Destexhe, and Karim Benchenane for useful discussions. This project was supported by a senior grant of the European Research Council (NeuroConsc Program) to S.D. and by the Fondation pour la Recherche Médicale Equipe FRM 2010 grant to L.N. The NeuroSpin MEG facility was sponsored by grants from the Institut National de la Santé et de la Recherche Médicale, the Commissariat à l'Energie Atomique, the Fondation pour la Recherche Médicale, the Bettencourt-Schueller Foundation, and the Région île-de-France and is part of a general research program on functional neuroimaging of the human brain headed by Denis Le Bihan.

1. Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2:79–87.
2. Kiebel SJ, von Kriegstein K, Daunizeau J, Friston KJ (2009) Recognizing sequences of sequences. *PLOS Comput Biol* 5:e1000464.
3. Kiebel SJ, Daunizeau J, Friston KJ (2008) A hierarchy of time-scales and the brain. *PLOS Comput Biol* 4:e1000209.
4. Friston K (2005) A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* 360: 815–836.
5. Garrido MI, Kilner JM, Kiebel SJ, Friston KJ (2009) Dynamic causal modeling of the response to frequency deviants. *J Neurophysiol* 101:2620–2631.
6. Garrido MI, Kilner JM, Kiebel SJ, Friston KJ (2007) Evoked brain responses are generated by feedback loops. *Proc Natl Acad Sci USA* 104:20961–20966.
7. Parmentier FBR, Elsley JV, Andrés P, Barceló F (2011) Why are auditory novels distracting? Contrasting the roles of novelty, violation of expectation and stimulus change. *Cognition* 119:374–380.
8. Todorovic A, van Ede F, Maris E, de Lange FP (2011) Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: An MEG study. *J Neurosci* 31: 9118–9123.
9. Summerfield C, Trittschuh EH, Monti JM, Mesulam M-M, Egner T (2008) Neural repetition suppression reflects fulfilled perceptual expectations. *Nat Neurosci* 11: 1004–1006.
10. Garrido MI, et al. (2008) The functional anatomy of the MMN: A DCM study of the roving paradigm. *Neuroimage* 42:936–944.
11. Bekinschtein TA, et al. (2009) Neural signature of the conscious processing of auditory regularities. *Proc Natl Acad Sci USA* 106:1672–1677.
12. Näätänen R, Gaillard AW, Mäntysalo S (1978) Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol (Amst)* 42:313–329.
13. Winkler I (2007) Interpreting the mismatch negativity (MMN). *J Psychophysiol* 21: 147–163.
14. May PJC, Tiitinen H (2010) Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology* 47:66–122.
15. Garrido MI, Kilner JM, Stephan KE, Friston KJ (2009) The mismatch negativity: A review of underlying mechanisms. *Clin Neurophysiol* 120:453–463.
16. Yabe H, Tervaniemi M, Reinikainen K, Näätänen R (1997) Temporal window of integration revealed by MMN to sound omission. *Neuroreport* 8:1971–1974.
17. Raij T, McEvoy L, Mäkelä JP, Hari R (1997) Human auditory cortex is activated by omissions of auditory stimuli. *Brain Res* 745:134–143.
18. Hughes HC, et al. (2001) Responses of human auditory association cortex to the omission of an expected acoustic event. *Neuroimage* 13:1073–1089.
19. Bendixen A, Schröger E, Winkler I (2009) I heard that coming: Event-related potential evidence for stimulus-driven prediction in the auditory system. *J Neurosci* 29: 8447–8451.
20. den Ouden HEM, Friston KJ, Daw ND, McIntosh AR, Stephan KE (2009) A dual role for prediction error in associative learning. *Cereb Cortex* 19:1175–1185.
21. Cohen D, Cuffin BN (1983) Demonstration of useful differences between magnetoencephalogram and electroencephalogram. *Electroencephalogr Clin Neurophysiol* 56:38–51.
22. Faugeras F, et al. (2011) Probing consciousness with event-related potentials in patients who meet clinical criteria for vegetative state. *Neurology* 77:264–268.
23. Allen J, Kraus N, Bradlow A (2000) Neural representation of consciously imperceptible speech sound differences. *Percept Psychophys* 62:1383–1393.
24. Diekhof EK, Biedermann F, Ruebsamen R, Gruber O (2009) Top-down and bottom-up modulation of brain structures involved in auditory discrimination. *Brain Res* 1297: 118–123.
25. Del Cul A, Baillet S, Dehaene S (2007) Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biol* 5:e260.
26. Sergent C, Baillet S, Dehaene S (2005) Timing of the brain events underlying access to consciousness during the attentional blink. *Nat Neurosci* 8:1391–1400.
27. Pegado F, et al. (2010) Probing the lifetimes of auditory novelty detection processes. *Neuropsychologia* 48:3145–3154.
28. Ritter W, Sussman E, Deacon D, Cowan N, Vaughan HG, Jr. (1999) Two cognitive systems simultaneously prepared for opposite events. *Psychophysiology* 36:835–838.
29. Winkler I, Takegata R, Sussman E (2005) Event-related brain potentials reveal multiple stages in the perceptual organization of sound. *Brain Res Cogn Brain Res* 25:291–299.
30. Horváth J, Czigler I, Sussman E, Winkler I (2001) Simultaneously active pre-attentive representations of local and global rules for sound sequences in the human brain. *Brain Res Cogn Brain Res* 12:131–144.
31. Sabri M, Campbell KB (2001) Effects of sequential and temporal probability of deviant occurrence on mismatch negativity. *Brain Res Cogn Brain Res* 12:171–180.

Downloaded at CEA/GRENOBLE on July 6, 2020