# Subliminal words durably affect neuronal activity

Raphaël Gaillard[a,b], Laurent Cohen[a,b], Claude Adam[a], Stéphane Clemenceau[a], Dominique Hasboun[a], Michel Baulac[a], Jean-Claude Willer[a], Stanislas Dehaene[b] and Lionel Naccache[a,b]

[a]AP-HP, Pitié-Salpêtrière Hospital & University Pierre et Marie Curie, Paris and [b]INSERM, U562 Cognitive Neuroimaging Unit, Orsay, France

Correspondence to Dr Lionel Naccache, MD, PhD, Assistance Publique Hôpitaux de Paris and INSERM, 47-83 Bd Hôpital, Paris, France
Tel: + 33 1 40 77 97 99; fax: + 33 1 40 77 97 89; e-mail: lionel.naccache@normalesup.org

Received 16 July 2007; accepted 26 July 2007

Unconscious mental representations elicited by subliminal stimuli are marked by their fleeting lifetimes, usually below 1 s. Can such evanescent subliminal stimuli, nevertheless, lead to long-lasting learning? To date, evidence suggesting a long-term influence of briefly perceived stimuli on behaviour or brain activity is scarce and questionable. In this study, we used intracranial recordings to provide the first direct demonstration that unconsciously perceived subliminal words could exert long-lasting effects on neuronal signals. When repeating subliminal words over long inter-stimulus intervals, we observed electrophysiological repetition effects. These unconscious repetition effects suggest that the single presentation of a masked word can durably affect neural architecture. *NeuroReport* 18:1527–1531 © 2007 Wolters Kluwer Health | Lippincott Williams & Wilkins.

**Keywords**: cognition, epilepsy, human, intracranial recordings, subliminal stimuli, unconscious mental representations

## Introduction

Unconscious active mental representations are marked by their fleeting lifetimes, usually below 1 s. This core limitation of unconscious cognitive processing has been convincingly reported in a large set of behavioural and electrophysiological studies across various classes of stimuli and tasks (see Ref. [1] for a synthetic review). Under conditions of subliminal perception during which a visual stimulus is presented without the participants' concious knowledge, unconscious representations of the stimulus exert a behavioural priming influence on the processing of a subsequent stimulus for just a few hundred milliseconds [2]. Using human intracranial recordings, we recently extended this behavioural finding by showing that masked words elicit abstract forms of active mental representations that correlate with fleeting neural effects within various cortical regions such as the left fusiform gyrus [3] and the amygdala [4]. The neural correlates of the masked words had totally vanished less than 1 s after the stimulus presentation.

Can such evanescent subliminal stimuli, nevertheless, lead to long-lasting learning? The 'global neuronal workspace model' of consciousness postulates that beyond active representations, an unconsciously perceived stimulus can elicit modifications in the neural architecture such as in the synaptic weights of local networks [1]. This 'structural unconscious' can theoretically lead to long-lasting traces observable indirectly in terms of either behavioural effects or neural activity. In this work, we aimed at exploring this theoretical prediction.

To date, very few studies have reported any evidence supporting this proposal. The mere exposure effect shows that an arbitrary visual stimulus presented subliminally can affect preference judgement when presented several minutes, hours or days later [5]. Most behavioural studies are, however, weakened by a severe methodological caveat: the absence of conscious perception of the stimuli was not assessed through reliable measures such as forced-choice tasks and objective discriminability parameters (e.g. d′ index from signal detection theory).

A single brain-imaging PET study has been performed on the mere exposure effect, reporting smaller activations for abstract visual stimuli that had previously been presented subliminally than for novel stimuli. The neural structures modulated by stimulus repetition included the left fusiform gyrus, right hippocampus and left thalamus (see Table 1 in Ref. [6]). Note, however, that participants were scanned when they were consciously perceiving novel or repeated stimuli and were performing either an explicit memory task or a preference judgement task. Therefore, differences between brain activation patterns observed in this study probably included the correlates of multiple processes, the genuine durable forms of unconscious memory not only subtending the mere exposure effect, but also the additional conscious metacognitive effects [7].

The second source of evidence of unconscious long-term learning effects deals with visual motion direction learning. Watanabe *et al.* [8] repeatedly presented participants with an irrelevant background motion signal so weak that its direction was not visible when they were performing a letter detection task. Despite being below the threshold of visibility and being irrelevant to the central task, the repetitive exposure improved performance specifically for the direction of the exposed motion when tested in a subsequent suprathreshold test. This elegant study is purely behavioural and concerns a low-level visual parameter.

In this study, we provide the first evidence of the long-term neural effects of subliminal words. Seven patients suffering from refractory epilepsy were implanted with

0959-4965 © Wolters Kluwer Health | Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

**Table 1** Unconscious repetition effects

| Patient number | Electrode Talairach coordinates | | | Median ISI (minutes) | Repetition effect |
|---|---|---|---|---|---|
| | *X* | *Y* | *Z* | | |
| 1 | 42 | −75 | 20 | 20.7 | 0.047 |
| 1 | 42 | −75 | 20 | 20.7 | < 0.01 |
| 1 | −33 | −2 | −28 | 20.7 | 0.058 |
| 1 | −34 | −30 | −10 | 20.7 | 0.041 |
| 2 | 39 | −51 | −4 | 24.3 | 0.049 |
| 2 | 40 | −60 | 3 | 24.3 | 0.010 |
| 2 | −30 | −51 | 0 | 24.3 | < 0.01 |
| 2 | −30 | −60 | 6 | 24.3 | < 0.01 |
| 2 | −30 | −60 | 6 | 24.3 | < 0.01 |
| 2 | −30 | −70 | 10 | 24.3 | 0.029 |
| 3 | −42 | −33 | −8 | 16.7 | 0.064 |
| 3 | −64 | −44 | 10 | 16.7 | < 0.01 |
| 3 | −44 | −45 | −22 | 16.7 | 0.028 |
| 4 | −9 | −88 | −8 | 19.2 | 0.092 |
| 4 | −30 | −93 | −8 | 19.2 | 0.086 |
| 4 | −30 | −93 | −8 | 19.2 | 0.047 |
| 5 | 30 | 3 | −36 | 11.6 | 0.029 |
| 5 | 40 | −10 | −39 | 11.6 | 0.088 |
| 6 | 14 | −94 | −12 | 15.7 | < 0.01 |
| 6 | 14 | −94 | −12 | 15.7 | 0.010 |
| 6 | 33 | −92 | 0 | 15.7 | 0.044 |
| 7 | 27 | −61 | −9 | 11.7 | < 0.01 |
| 7 | 32 | −62 | −9 | 11.7 | 0.046 |

For each repetition effect we provide the electrode coordinates in Talairach anatomical space, the median ISI value and statistical significance (false discovery rate-corrected *P* value of Monte Carlo estimation).
ISI, interstimulus interval.

intracranial electrodes recording local field potentials (LFP). We report that in the absence of objective measures of consciousness, subliminal words can indeed have long-lasting effects on LFPs, over a time range exceeding 20 min.

## Materials and methods

### Patients

Seven patients (four men) suffering from drug refractory epilepsy were stereotactically implanted with depth electrodes as part of a presurgical evaluation. Their ages ranged from 26 to 47 years. Six patients were right-handed; one woman was ambidextrous. Neuropsychological assessment revealed normal or mild impairment in general cognitive functioning: verbal IQ ranged from 87 to 97 and performance IQ from 76 to 120. Experiments were approved by the Ethical Committee for Biomedical Research of Pitié-Salpêtrière Hospital in Paris (agreement #99–04 issued on 15 December 2004), and participants gave informed consent.

### Experimental protocol

Patients were presented with 552 randomized trials corresponding to three occurrences each of 92 masked words and 92 unmasked words. On each trial, a single word was presented briefly (29 ms), which was preceded and followed by masks consisting of strings of characters (71 ms and 400 ms, respectively; see Fig. 1). To enhance attention and semantic processing, masked trials were randomly intermixed with visible trials in which the poststimulus mask had been removed. Furthermore, participants were engaged in a forced-choice task of categorizing each word as threatening or nonthreatening by pressing response buttons, even on the masked trials. To prevent automatic stimulus–response learning, we used two distinct sets of 92 French words each for the masked trials and the visible trials, so that the masked words were never seen consciously. In each list, half of the words were threatening (e.g. 'danger', 'kill'), with variable frequencies, lengths (3–8 letters) and lexical categories (verbs and nouns). The other half included nonthreatening emotionally neutral words (e.g. 'cousin', 'see'), matched for frequency, length and category. For each patient, the hands assigned to the threatening and nonthreatening responses were inverted halfway through the experiment. Each word was repeated with random interstimulus intervals (ISI) ranging from 5 s to 47 min.

### Local field potential recording

Our approach is similar to the one we previously used in two earlier publications [3,4]. Patients were implanted with depth electrodes (Ad-Tech Medical Instruments, Racine, Wisconsin, USA). Electrodes were 2.3-mm long, 1-mm diameter cylinders, with the interelectrode distance being 10 mm. The structures to be explored were defined on the basis of ictal manifestations, electroencephalography (EEG) and neuroimaging studies. For each electrode, the cartesian coordinates $(x,y,z)$ were calculated after normalization of the anatomical three-dimensional SPGR anatomical cerebral MRI into Talairach space using SPM99 (Matlab, London, UK).

### Local field potential processing

The local field potential was digitized at 400 Hz from intracerebral electrodes referenced to the vertex (Nicolet-BMSI, Madison, Wisconsin, USA). Epochs were then extracted (−500 ms plus 1000 ms from word onset), submitted to automatic artifact rejection (±300-μV threshold), visually inspected, notch filtered (50 Hz) and then low-pass filtered (20 Hz) using EEGLAB software (Matlab, San Diego, California, USA) [9]. Baseline correction (from −500 to 0 ms before word onset) was applied, and potentials were averaged.

### Statistics

#### *Behavior*

Response times (RTs) of less than 250 ms were discarded. Median RTs were calculated for each patient and for each condition (first, second and third occurrences). RTs were compared across conditions using analysis of variance (ANOVA) *F*-tests.

For each patient we also computed objective discriminability d′ separately for each of the three word occurrences on the 92 experimental trials. We then assessed better-than-chance performance on d′ using both individual and group statistical criteria. First, individual $\chi^2$ statistics were calculated for each word occurrence: proportions of correct and incorrect responses to emotional and neutral words were compared with those in the expected random distribution. Second, group analysis was performed using a distribution statistic to estimate d′ performance compared with chance-level performance, which is defined by a null d′ value (Z test *P* values comparing the observed distribution of d′ values to a zero-centred Gaussian).

#### *Local field potentials*

We used a four-step strategy to assess the statistical significance of our results.

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

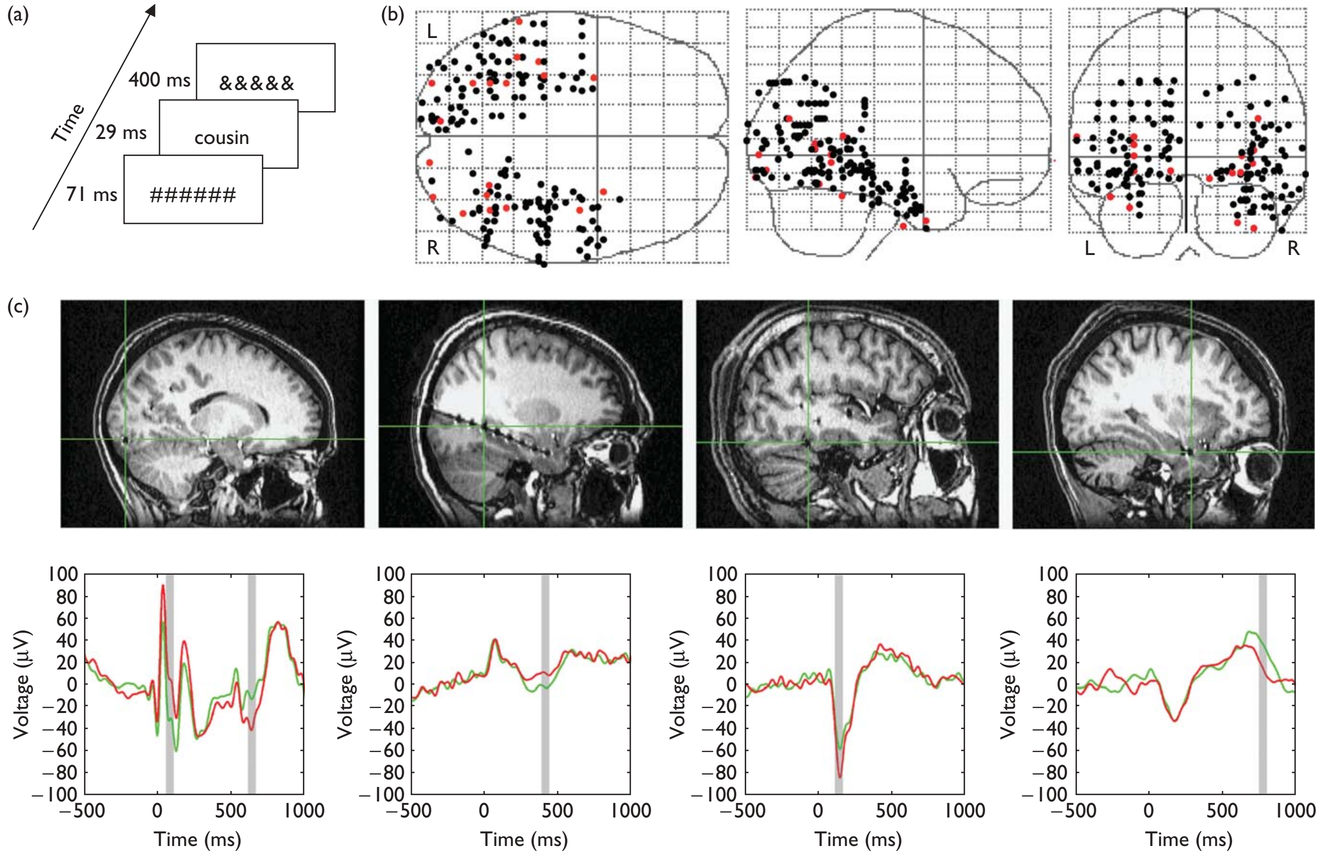

**Fig. 1** Long-lasting unconscious word-repetition effects. (a) Experimental paradigm used to present masked words. (b) Axial (left), sagittal (middle) and coronal (right) normalized glass brain with electrodes showing (red) or not showing (black) unconscious repetition effects. (c) Local field potentials (LFPs) of four electrodes with averaged responses to first (green) and second (red) word occurrences. Shadowed areas indicate significant repetition effects ($P < 0.01$ on 15 successive samples).

First, the averaged LFPs of first word-occurrence trials were compared with those of second word-occurrence trials by using sample-by-sample *t*-tests, with a criterion of significance being set at $P < 0.01$ for a minimum of 15 consecutive samples. Second, we further checked the statistical significance of the observed effects (number of consecutive samples with $P < 0.01$ on *t*-test) through Monte Carlo permutations. This method provides an estimation of type I error rate (false positives) by using resampling procedures. Precisely, for each patient and for each electrode showing a significant effect, we computed 1000 random permutations of the observed trials in two surrogate conditions: trials were randomly assigned to one of the two groups, and then for each permutation we counted the number of surrogate effects satisfying the observed effect anywhere within a time window of 0–1000 ms after stimulus onset.

Third, for each repetition effect satisfying criterion 1 (a minimum of 15 consecutive samples with $P < 0.01$ on *t*-test), we then averaged the LFP on each trial across a 40-ms temporal window centred on the most significant sample estimated in step 1 (sample-by-sample *t*-tests). We then subjected these time-window average values to an ANOVA with Time (first/second half of experiment) and ISI (inferior/superior to median ISI) factors. We also calculated the repetition effect restricted to long ISI trials.

Fourth, we compared the time-window average LFP values of the 55 earliest second occurrences of words with the LFPs of the 55 latest first occurrences of words. This allowed us to invert median and average times: second occurrences corresponded with the earlier times more than first occurrences did.

Statistical values obtained in the Monte Carlo procedure were corrected for multiple comparisons across electrodes using the False Discovery Rate procedure. The false discovery rate of a test is defined as the expected proportion of false positives among the declared significant results.

Finally, for each repetition effect, we averaged the LFPs across a 40-ms window, and ran an ANOVA analysis distinguishing both short and long ISIs (ISI inferior/superior to individual participant median ISI), and presentations happening early or late in the experiment (before/after midexperiment time).

## Results

### Behavioral data

RTs were similar across the three word occurrences ($P > 0.6$ in F test performed on median RTs; mean RT, 1586 ms), and comparable with previous observations in the same paradigm tested on normal controls [4]. For each patient, d′ was calculated separately for each of the three word occurrences on the 92 experimental trials. All d′ values ranged from

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

$-0.22$ to $+0.41$. We then assessed better-than-chance performance on d′ measures using both individual and group statistical criteria. First, the individual $\chi^2$ statistic was calculated for each word occurrence. For each of the seven participants, performance was at chance level for first and second word occurrences (all *P* values $>0.2$), whereas a slight nonsignificant trend was observed on the third word occurrence for three participants (*P* values $>0.1$). Distribution statistics confirmed chance-level performance for the first and second word occurrences (Z test *P* values $>0.2$); whereas, again the third word occurrence was better than at chance level ($P=0.04$). Taken together these data convincingly demonstrate the absence of conscious word perception for the first and second word occurrences. Therefore, we now focus our interest exclusively on first and second word-occurrence trials.

### Electrophysiological data

Out of a total of 177 intracranial electrodes (see Fig. 1b and c), 19 (10.7%) presented significant unconscious word-repetition effects. These electrodes were located in almost all the implanted cortical structures, from the occipital pole to the temporal and amygdala cortices. Further analyses confirmed that these effects held over long time intervals and were not due to temporal confounds. In 13 (7.3%) electrodes, repetition effects were still observed on the long ISI trials, with no impact of the time and ISI factors. Furthermore, for 10 (5.6%) of these electrodes across five patients, these repetition effects continued to be significant when contrasting the repeated words that occurred in the first half of the experiment with the novel words that occurred in the second half of the experiment, indicating an absence of confound with time.

## Discussion

In this study, we could demonstrate that unconsciously perceived words might have a long-lasting impact on neuronal activity in spite of the fleeting lifetimes of their own representations. We used objective measures of stimulus discriminability to confidently claim the absence of conscious perception of the masked words. Note that the experimental trials that contributed to these consciousness assessment measures were exactly the same ones during which LFPs were recorded. We probed the existence of long-lasting effects of unconsciously perceived words by comparing LFP signals elicited by the first and second occurrences of each word, respectively. Contrary to the earlier reports of long-lasting repetition effects of briefly presented stimuli [6–8], neither the first nor the second stimulus was consciously perceived in this study. This point makes our experimental design more immune to metacognitive effects than earlier studies because, when processing the second word unconsciously and unknowingly, the participants could not have introspected on any perceptual ease.

The long-lasting neural effects of unconscious word repetition were observed in almost all the implanted regions across the seven epileptic patients and included the bilateral occipital regions, fusiform gyri, other temporal regions and amygdale structures. Our study, however, did not sample all the cortical territories. None of these patients was implanted within the frontal lobes, the cingulate cortices, superior temporal regions or superior parietal areas. Compared with the Eliott and Dolan study [6], we observe a close replication of several anatomical loci of unconscious repetition effects. Among the five regions reported by Eliott and Dolan, only two corresponded to the anatomical loci of implantation in our study: the hippocampus gyrus and the left fusiform gyrus. The left fusiform effect reported in the Eliott and Dolan study ($-30$, $-68$, 6) was replicated in this study less than 8 mm away ($-30$, $-60$, 6). Similarly, we observed a left hippocampal effect less than 12 mm away from the symmetrical coordinates of the effect reported by Elliot and Dolan within the right hippocampus (26, $-38$, $-12$ versus $-34$, $-30$, $-10$). The right/left asymmetry of this effect might originate in the verbal (words) versus abstract visual (Japanese ideograms presented to English speakers) nature of the visual stimuli used.

We should note that these unconscious repetition effects are not necessarily related to the abstract processing of the masked words, but that they can originate from very different levels of processing, from low perceptual vision up to semantic stages. Both the spatial and temporal diversities of these effects suggest that they might be heterogeneous. The earliest of the most posterior effects might be related to a retinotopic stage of processing, whereas the latest might correspond to the more abstract forms of processing. Indeed, using the same experimental paradigm we previously demonstrated that masked words were processed up to the semantic and emotional level. Electrodes located within or near the amygdala revealed a modulation of LFPs according to the emotional valence of words [4]. Notably, LFP data from the three patients explored in this study correspond to a subset of the dataset analyzed in this study.

Although previous reports demonstrated that some repetition effects might rely on conscious response learning processes or on metacognitive effects [10], our results reveal that durable unconscious repetition effects do exist. According to the evanescence that characterizes most, if not all, unconscious mental representations, we hypothesize that these durable unconscious effects are not the correlates of unconscious representatives processes, but rather that they correspond to structural changes – such as synaptic weight modifications – induced by the unconscious presentations of words. Further experiments are necessary to demonstrate the validity of our proposal, and to precisely describe the temporal dynamics of these postulated structural changes.

## Acknowledgements

The authors thank Dr Edward Hubbard for his helpful comments. RG was supported by the Fédération Française pour la Recherche sur l'Epilepsie.

## References

1. Dehaene S, Naccache L. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 2001; **79**:1–37.
2. Greenwald AG. Three cognitive markers of unconscious semantic activation. *Science* 1996; **273**:1699–1702.
3. Gaillard R, Naccache L, Pinel P, Clémenceau S, Volle E, Hasboun D, *et al*. Direct intracranial, FMRI, and lesion evidence for the causal role of left inferotemporal cortex in reading. *Neuron* 2006; **50**:191–204.
4. Naccache L, Gaillard R, Adam C, Hasboun D, Clémenceau S, Baulac M, *et al*. A direct intracranial record of emotions evoked by subliminal words. *Proc Natl Acad Sci USA* 2005; **102**:7713–7717.

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

5. Kunst-Wilson WR, Zajonc RB. Affective discrimination of stimuli that cannot be recognized. *Science* 1980; **207**:557–558.
6. Elliott R, Dolan RJ. Neural response during preference and memory judgments for subliminally presented stimuli: a functional neuroimaging study. *J Neurosci* 1998; **18**:4697–4704.
7. Bornstein RF, D'Agostino PR. The attribution and discounting of perceptual fluency: preliminary tests of a perceptual fluency/ attributional model of the mere exposure effect. *Soc Cogn* 1994; **12**:103–128.
8. Watanabe T, Nanez JE, Sasaki Y. Perceptual learning without perception. *Nature* 2001; **413**:844–848.
9. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 2004; **15**:9–21.
10. Dobbins IG, Schnyer DM, Verfaellie M, Schacter DL. Cortical activity reductions during repetition priming can result from rapid response learning. *Nature* 2004; **428**:316–319.

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.