

LETTER TO THE EDITOR

Reply: Replicability and impact of statistics in the detection of neural responses of consciousness

Lionel Naccache,^{1,2,3,4,5} Jacobo Sitt,^{3,6} Jean-Rémi King,^{3,6} Benjamin Rohaut,^{1,3} Frédéric Faugeras,³ Srivas Chennu,⁷ Mélanie Strauss,⁶ Mélanie Valente,^{3,4} Denis Engemann,^{3,4} Federico Raimondo,^{3,4} Athena Demertzi,⁴ Tristan Bekinschtein⁸ and Stanislas Dehaene^{6,9}

1 AP-HP, Groupe hospitalier Pitié-Salpêtrière, Department of Neurology, Paris, France

2 AP-HP, Groupe hospitalier Pitié-Salpêtrière, Department of Neurophysiology, Paris, France

3 INSERM, U 1127, F-75013, Paris, France

4 Institut du Cerveau et de la Moelle épinière, ICM, PICNIC Lab, F-75013, Paris, France

5 Sorbonne Universités, UPMC Univ Paris 06, Faculté de Médecine Pitié-Salpêtrière, Paris, France

6 INSERM-CEA Cognitive Neuroimaging unit | CEA/SAC/DSV/DRM/Neurospin Center, Gif/Yvette cedex, France

7 Department of Clinical Neurosciences, University of Cambridge, UK

8 Department of Psychology, University of Cambridge, Cambridge, UK

9 Collège de France, 75005 Paris, France

Correspondence to: Lionel Naccache, AP-HP, Groupe hospitalier Pitié-Salpêtrière, Department of Neurology, Paris, France

E-mail: lionel.naccache@aphp.fr

Sir,

We read with interest the letter by Gabriel and colleagues (2016) addressing the major issue of replicability when probing conscious processing in non-communicating patients. This question—as well as the choice of the optimal statistical methodology—concerns the whole field of functional brain imaging in cognitive neuroscience (Kriegeskorte *et al.*, 2009), but its importance obviously culminates in single-subject analyses of non-communicating patients (see for instance the recent debate in Cruse *et al.*, 2011, 2013; Goldfine *et al.*, 2012). Gabriel *et al.* reacted to a recent discussion (Naccache *et al.*, 2015; Tzovara *et al.*, 2015a, b) following a report by Tzovara *et al.* (2015a), who adapted our auditory ‘local-global’ bedside EEG test (Bekinschtein *et al.*, 2009) to test comatose patients.

Briefly, in the local-global paradigm two levels of regularities are manipulated: local auditory irregularities correspond to a change of sound within a trial, whereas global irregularities correspond to a change of sound sequence across trials. When analysing data according to the local irregularities, one can typically extract a mismatch negativity response observable even in unconscious states. In sharp

contrast, when analysing event-related potentials (ERPs) to violations of global irregularities, we previously showed that a late global effect was present only in conscious or minimally conscious patients (Bekinschtein *et al.*, 2009; Faugeras *et al.*, 2011, 2012).

Two problems emerged from the study of Tzovara *et al.* (2015a), first, this ERP global effect was found positive in the vast majority of conscious controls we tested at two distinct sites using high-density EEG: 18/18 (100%) in Paris, France (with 256 electrodes), and 7 to 10/10 (70 to 100%) with the monaural and binaural versions of the task, respectively in Cambridge, UK (with 128 electrodes). In sharp contrast, Tzovara *et al.* (2015a) adopted a multivariate decoding approach and could detect it in only 36% (4/11) of conscious controls. Second and more surprisingly, Tzovara *et al.* reported a global effect in 42% (10/24) of post-anoxic comatose patients, some of whom presented very severe EEG patterns such as ‘non-reactive EEG’ or EEG under burst-suppression regime.

We previously exposed a probable explanation for these two problematic results (Naccache *et al.*, 2015): rather than identifying the genuine global effect, which

is a late (>250 ms) event associated with conscious access, Tzovara *et al.* (2015a) most probably captured a modulation of the early unconscious mismatch negativity response by statistical regularities differing across the different blocks. Note that we previously used a single-trial multivariate decoding approach, and reported this mismatch negativity response modulation, which is not to be confused with the late global effect (King *et al.*, 2013).

In reaction to this debate, Gabriel *et al.* (2016) presented a valuable and interesting original dataset of 27 conscious controls tested with a typical auditory mismatch negativity paradigm while being recorded with a high-density EEG machine. Given that one expects a reliable test to be able to identify such a mismatch negativity response in any control subject free of hearing impairments, they then analysed this dataset with six distinct statistical methods. They observed discrepancies in terms of statistical power across methods, and noted that when combining two methods they could identify a mismatch negativity in each of these conscious controls. This discussion calls for four types of remarks.

Cooperation and sharing are necessary

This short methodological study is a model for future studies. Indeed, when discussing the discrepancy between our results and those of Tzovara and colleagues, one has to be aware of the numerous differences distinguishing these datasets and their analyses: auditory stimuli were different (duration versus pitch mismatch negativity response paradigm), EEG devices were different (high-density EEG with 256 channels versus only 19 channels), analytic and statistical methods were different (visual inspection and univariate ERPs analysis versus multivariate analysis), and patients were different (early comatose patients under mild hypothermia versus awoken patients in the vegetative or minimally conscious states). These various differences make it even more difficult to interpret univocally any discrepancy between our respective results. Clearly, a massive effort in sharing data, paradigms and methods is necessary in order to converge towards robust and reliable tools. The very same datasets have to be analysed with different methods, as reported by Gabriel *et al.* (2016). Note that the multivariate method that we implemented is open-source and can be implemented following the steps described in https://github.com/mne-tools/mne-python/blob/maint/0.11/examples/decoding/plot_decoding_sensors.py. We are currently engaged in an international project ('Recovery of consciousness after severe brain injury Phase II' grant of the James S. McDonnell Foundation) and in a French multi-centric project ('ANR-CogniComa', ANR 14-CE15-0013-04), specifically designed for a large sharing stage

necessary to define relevant and standardized procedures. Broadening such data sharing initiatives to other research groups and data sources would advance us towards a scientifically robust consensus.

Sensitivity is a prerequisite of any test of consciousness

High sensitivity is a prerequisite of any functional brain imaging test of consciousness: it should be able to detect consciousness at the individual level in the vast majority of conscious controls. This is precisely why we were encouraged to use the global effect as a promising translational clinical tool. Tests or statistical methods that fail at this crucial stage should not be used in a clinical perspective. For instance, shortly after having designed the local-global test we tried to imagine another ERP test enabling us to probe conscious access to meanings of words. While this verbal semantic test gave interesting findings at the group level, it was not sensitive enough at the single-subject level in conscious controls (Rohaut *et al.*, 2015). So, we do not use it in the clinical practice in its current version. Note on the other hand that specificity of most functional tests, which usually require active cognitive participation of the subject, is usually limited (but see Casali *et al.*, 2013). For instance, we demonstrated in the local-global task that conscious subjects stop showing the global effect when their attention is distracted by a parallel visual task (Bekinschtein *et al.*, 2009).

A hierarchical approach of EEG-based measures

In the case of electrophysiology we typically use a hierarchical approach: for instance, a typical late (>250 ms), sustained and significant global effect has to be preceded by a mismatch negativity response (local effect), by a contingent negative variation (CNV) and by an early response to sounds (see such a hierarchy in Table 2 of Rohaut *et al.*, 2015). When all these conditions are met, and when significant results converge with experts' visual inspection (corresponding to Method 1 evaluated by Gabriel *et al.*, 2016), the value of a global effect is typically stronger than when this expected hierarchy of responses is violated. Such a hierarchical analysis completed by the visual and statistical inspection of ERPs is lacking from the report of Tzovara and colleagues (2015a) who focused their analysis on an automated multivariate method. In our opinion, the emergence of complex methods of EEG signal-processing, although extremely valuable, calls for an even more systematic inspection of the raw data and traditional ERP or spectral-power measures.

Expertise in consciousness assessment is a multivariate and metacognitive exercise

In the current absence of a unique and reliable test of consciousness, we build a clinical diagnosis by integrating the following multiple variables: (i) detailed and repeated neurological examination and behavioural scoring such as the one offered by the Coma Recovery Scale-Revised (CRS-R) (Giacino *et al.*, 2004); (ii) structural brain imaging (mostly MRI including several sequences such a diffusion tensor imaging) (Luyt *et al.*, 2012); (iii) EEG and functional MRI during resting conditions as well as active tasks or stimulations (including TMS-EEG measures used by some groups) (Casali *et al.*, 2013); and (iv) brain metabolism and perfusion measures such as PET of arterial spin labelling (Liu *et al.*, 2011; Stender *et al.*, 2014). When the combination of all these independent measures converges toward a common pattern, confidence in the diagnosis can be increased. Inversely, divergences and mismatches between these different tests and measurements call for caution, and should act as a metacognitive warning. In the future, this ability to correctly weigh the level of confidence in a diagnosis could perhaps be automated (Sitt *et al.*, 2014), but at present it lies at the core of human expertise in this field.

We close by reminding readers that a correct diagnosis conveys crucial prognosis information, not only about survival, but also about consciousness recovery and functional outcome (Luaute *et al.*, 2010; Sitt *et al.*, 2014; Faugeras *et al.*, 2016). This obvious point strengthens the claim of Gabriel *et al.* (2016): we must coordinate our creative and methodological efforts at the service of these patients and of their relatives.

Funding

L.N. is supported by the Fondation pour la Recherche Médicale ('Equipe FRM 2015' grant); 'ANR-CogniComa', ANR 14-CE15-0013-04; and the program 'Investissements d'avenir' ANR-10-IAIHU-06. L.N. and S.D. are supported by the 'Recovery of consciousness after severe brain injury Phase II' grant of the James S. McDonnell Foundation.

References

Bekinschtein TA, Dehaene S, Rohaut B, Tadel F, Cohen L, Naccache L. Neural signature of the conscious processing of auditory regularities. *Proc Natl Acad Sci USA* 2009; 106: 1672–7.

Casali AG, Gosseries O, Bruno M, Charland-Verville V, Vanhauzenhuyse A, Demertzi A, et al. A theoretically based index of consciousness independent of sensory processing and behavior. *Sci Trans Med* 2013; 5: 198ra105.

Cruse D, Chennu S, Chatelle C, Bekinschtein TA, Fernández-Espejo D, Pickard JD, et al. Bedside detection of awareness in the vegetative state: a cohort study. *Lancet* 2011; 378: 2088–94.

Cruse D, Chennu S, Chatelle C, Bekinschtein TA, Fernández-Espejo D, Pickard JD, et al. Reanalysis of "Bedside detection of awareness in the vegetative state: a cohort study" - Authors' reply. *Lancet* 2013; 381: 291–2.

Faugeras F, Rohaut B, Valente M, Sitt J, Demeret S, Bolgert F, et al. (2016). Survival and consciousness recovery are better in the minimally conscious state than in the vegetative state. Under review.

Faugeras F, Rohaut B, Weiss N, Bekinschtein T, Galanaud D, Puybasset L, et al. Event related potentials elicited by violations of auditory regularities in patients with impaired consciousness. *Neuropsychologia* 2012; 50: 403–18.

Faugeras F, Rohaut B, Weiss N, Bekinschtein TA, Galanaud D, Puybasset L, et al. Probing consciousness with event-related potentials in the vegetative state. *Neurology* 2011; 77: 264–8.

Gabriel D, Muzard E, Conte MM, Bardin JC, Schiff ND. Replicability and impact of statistics in the detection of neural responses of consciousness. *Brain* 2016; 139: e30.

Giacino JT, Kalmar K, Whyte J. The JFK Coma Recovery Scale-Revised: measurement characteristics and diagnostic utility. *Arch Phys Med Rehabil* 2004; 85: 2020–9.

Goldfine AM, Victor JD, et al. Bedside detection of awareness in the vegetative state. *Lancet* 2012; 379: 1701–2; author reply 1702.

King JR, Faugeras F, Gramfort A, Schurger A, El Karoui I, Sitt JD, et al. Single-trial decoding of auditory novelty responses facilitates the detection of residual consciousness. *Neuroimage* 2013; 83: 726–38.

Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci* 2009; 12: 535–40.

Liu AA, Voss HU, Dyke JP, Heier LA, Schiff ND. Arterial spin labeling and altered cerebral blood flow patterns in the minimally conscious state. *Neurology* 2011; 77: 1518–23.

Luaute J, Maucoort-Boulch D, Tell L, Quelard F, Sarraf T, Iwaz J, Boisson D, et al. Long-term outcomes of chronic minimally conscious and vegetative states. *Neurology* 2010; 75: 246–52.

Luyt CE, Galanaud D, Perlberg V, Vanhauzenhuyse A, Stevens RD, Gupta R, et al. Diffusion tensor imaging to predict long-term outcome after cardiac arrest: a bicentric pilot study. *Anesthesiology* 2012; 117: 1311–21.

Naccache L, King JR, Sitt J, Engemann D, El Karoui I, Rohaut B, Faugeras F, Chennu S, Strauss M, Bekinschtein T, Dehaene S. Neural detection of complex sound sequences or of statistical regularities in the absence of consciousness? *Brain* 2015 Dec; 138(Pt 12): e395. doi: 10.1093/brain/awv190. Epub 2015 Jul 1.

Rohaut B, Faugeras F, Chausson N, King JR, Karoui IE, Cohen L, et al. Probing ERP correlates of verbal semantic processing in patients with impaired consciousness. *Neuropsychologia* 2015; 66: 279–92.

Sitt JD, King JR, El Karoui I, Rohaut B, Faugeras F, Gramfort A, et al. Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain* 2014; 137: 2258–70.

Stender J, Gosseries O, Bruno M, Charland-Verville V, Vanhauzenhuyse A, Stevens RD, Gupta R, et al. Diagnostic precision of PET imaging and functional MRI in disorders of consciousness: a clinical validation study. *Lancet* 2014; 384: 514–22.

Tzovara A, Simonin A, Oddo M, Rossetti AO, De Lucia M. Neural detection of complex sound sequences in the absence of consciousness. *Brain* 2015; 138(Pt 5): 1160–6.

Tzovara A, Simonin A, Oddo M, Rossetti AO, De Lucia M. Reply: Neural detection of complex sound sequences or of statistical regularities in the absence of consciousness? *Brain* 2015; 138: e396.