Electroencephalographic Correlates of Temporal Bayesian Belief Updating and Surprise

Antonino Visalli (antonino.visalli@unipd.it)
Department of General Psychology, University of Padova, via Venezia 8
35131 Padova, Italy

Mariagrazia Capizzi (mgcapizzi@hotmail.com)
Department of Neuroscience, University of Padova, via Giustiniani 5
35128 Padova, Italy

Ettore Ambrosini (ettore.ambrosini@unipd.it)
Department of Neuroscience & Padova Neuroscience Center, University of Padova, via Giustiniani 5, Padova, 35128 Italy
Department of General Psychology, University of Padova, 35131 Padova, Italy

Bruno Kopp (Kopp.Bruno@mh-hannover.de)
Department of Neurology, Hannover Medical School, Carl Neuberg Strasse 1
30625 Hannover, Germany

Antonino Vallesi (antonino.vallesi@unipd.it)
Department of Neuroscience & Padova Neuroscience Center, University of Padova, via Giustiniani 5, Padova, 35128 Italy
Brain Imaging and Neural Dynamics Research Group, IRCCS San Camillo Hospital, Venice, 30126 Italy

Abstract

The brain predicts the timing of forthcoming events to optimize responses to them. Such predictions are driven by both prior expectations on the likely timing of stimulus occurrence and the information conveyed by the passage of time (hazard function). Events that violate expectations cause surprise and often induce updating of prior beliefs. Here we combined a Bayesian computational approach with electroencephalography (EEG) to investigate the neural dynamics associated with updating of temporal expectations in the human brain. Moreover, since belief updating is usually highly correlated with surprise, participants performed a temporal foreperiod task that was specifically designed for differentiating between these two processes. The results confirmed that updating and surprise can be functionally distinguished at the EEG level. We isolated two dissociable P3 subcomponents that specifically index the two processes, providing new insights on these event-related potential (ERP) components and their Bayesian interpretation. To the best of our knowledge, the present study delineates ERP correlates of belief updating and surprise about the timing of events for the first time.

Keywords: Temporal preparation; Bayesian brain; belief updating; surprise; P3 ERP

Introduction

The ability to predict the likely moment at which an event will occur is critical to optimize many cognitive processes that range from perception to action selection. Temporal predictions can be formalized in terms of the hazard function, which refers to the conditional probability that an event will occur given it has not yet occurred (Janssen & Shadlen, 2005; Nobre, Correa, & Coull, 2007). Temporal predictions thus rely on both prior expectations about the timing of events and the information inherent in the passage of time.

Previous reaction time (RT) studies - employing temporal foreperiods (FP) between warning signals and target stimuli - demonstrated that humans track the temporal hazard of target occurrence (Bueti et al., 2010; Herbst et al., 2018; Meindertsma et al., 2018). However, it is still unclear how prior temporal expectations are formed and revised by the brain. To fill this gap, we used a Bayesian computational approach to investigate EEG correlates of updating temporal expectations. Moreover, given that updating usually occurs in the presence of surprising events (O'Reilly et al., 2013), we also sought to disentangle EEG correlates of updating from those associated with surprise.

Following the EEG literature about Bayesian belief updating (Kopp, 2008; Mars et al., 2008, Kolossa, Kopp, & Fingscheidt, 2015), we predicted to differentiate updating and surprise at late, P3-like components on the present FP task.

Methods

Experimental Design

Twenty-six participants performed a FP task (Fig. 1) while their EEG activity was recorded from 64 electrodes.

Participants had to respond to target onsets that were separated from a neutral warning signal by variable FP. Most targets (80% of 790 trials; *predictable trials*) were predictable since they appeared after FP drawn from a Gaussian distribution with constant mean and standard deviation during a block of trials. However, FP means and standard deviations changed abruptly between blocks of trials. The 73 blocks were not temporally separated, and participants were explicitly instructed that a change in target color signaled the beginning of a new block. On few trials (20%; *uniform trials*), interspersed with the other trials, the target appeared after a FP drawn from a uniform distribution ranging from 250 to 2500 ms. Importantly, uniform targets were always white, signaling that the current trial was not igniting a new block (*update trial*). In sum, although update and uniform trials were both surprising, surprise was distinguishable from updating by target color.

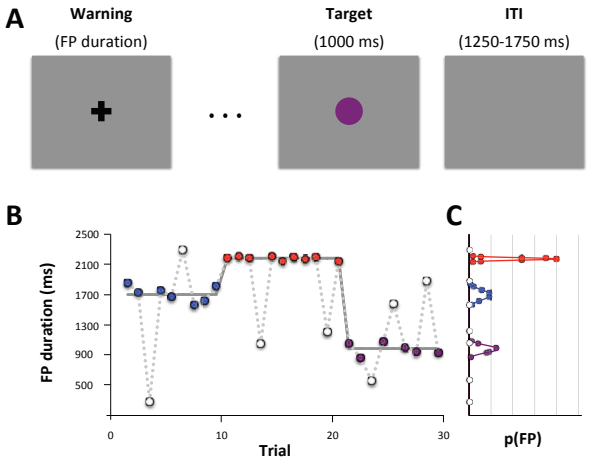

Figure 1: Paradigm and schematic depiction of the experimental approach. **(A)** Example of a trial. **(B)** Plot of FP duration over 30 trials. In most of the trials (i.e., 80%), target onsets were predictable since they occurred after FP (colored dots) drawn from a Gaussian distribution (panel C) whose mean and standard deviation were kept constant within a block. On 20% of trials, targets occurred after FP (white dots) extracted from a uniform distribution. **(C)** Distributions of FP duration in three exemplary blocks.

Ideal Bayesian Observer

To quantify updating and surprise, we developed a computational model (adapted from O'Reilly et al., 2013) that described beliefs about FP of an ideal Bayesian observer, and how such beliefs were trial-wise updated. After each observation, the model estimated the posterior probability of parameters μ and σ of the Gaussian distribution underlying normal FPs:

$$p(FP{\sim}\mathcal{N}(\mu,\sigma)|FP_{1:n}) \quad (1)$$

Updating was differently computed according to the trial type.

After *predictable* trials, the posterior (Eq. 1) was updated using Bayes' rule as:

$$\propto p\big(FP_n\big|FP{\sim}\mathcal{N}(\mu,\sigma)\big)p(FP{\sim}\mathcal{N}(\mu,\sigma)|FP_{1:n-1}) \quad (2)$$

After *update* trials, the change of color explicitly signaled the start of a new distribution. For this reason, the prior in update trials was blanked with a uniform distribution:

$$\propto p\big(FP_n\big|FP{\sim}\mathcal{N}(\mu,\sigma)\big)p\big(FP{\sim}\mathcal{U}(\mu,\sigma)\big) \quad (3)$$

According to the task instructions, no updating occurred after *uniform* trials, such that the posterior probability over parameter space at trial n was derived from the prior without modifications.

The model then translated the estimates of the parameters μ and σ into probability density functions over time. Specifically, the prior over time for a subsequent trial $n + 1$ was derived from the posterior over parameter space on trial n as follows:

$$\begin{aligned} p(FP_{n-1}|FP_{1:n}) &= p(\text{predictable}_{nb-1}) \\ &\sum_{\mu,\sigma}\big(FP_{n-1}\big|FP_{n-1}{\sim}\mathcal{N}(\mu_{n-1},\sigma_{n-1})\big) \\ &\times p(FP_{n-1}{\sim}\mathcal{N}(\mu_{n-1},\sigma_{n-1})|FP_{1:n}) \\ &= p(\text{uniform}_{nb-1} \cup \text{update}_{nb-1}) \end{aligned} \quad (4)$$

where $p(\text{predictable}_{nb+1})$ and $p(\text{uniform}_{nb+1} + \text{update}_{nb+1})$ represent the probability of incurring, respectively, in a predictable or in a uniform/update trial at the next trial of the current block (*nb* indicates the trial number within a block, which differs from *n* that indicates trial number referred to the whole task).

From the model (Fig. 2), we extracted trial-wise information-theoretic measures of updating and surprise (Baldi & Itti, 2010). Updating at trial n was formalized as the Kullback-Leibler divergence (D_{KL}; Fig. 2B) from prior to posterior beliefs:

$$D_{KL}(FP_n) = \sum_{FP} p(FP|\text{Prior}) \log \frac{p(FP|\text{Prior})}{p(FP|\text{Post})} \quad (5)$$

Since during the trial the prior probability of target onset changed as a function of the elapse of time (Janssen & Shadlen, 2005), we quantified surprise at trial n as the Shannon information (I_S; Fig. 2C) associated with the value of the hazard function at target onset:

$$I_S(FP_n) = = \log \frac{p(FP_n|\text{Prior})}{1 = F(FP_n|\text{Prior})} \quad (6)$$

where $F(FP_n|\text{Prior})$ was the cumulative probability $p(FP \leq FP_n|\text{Prior})$.

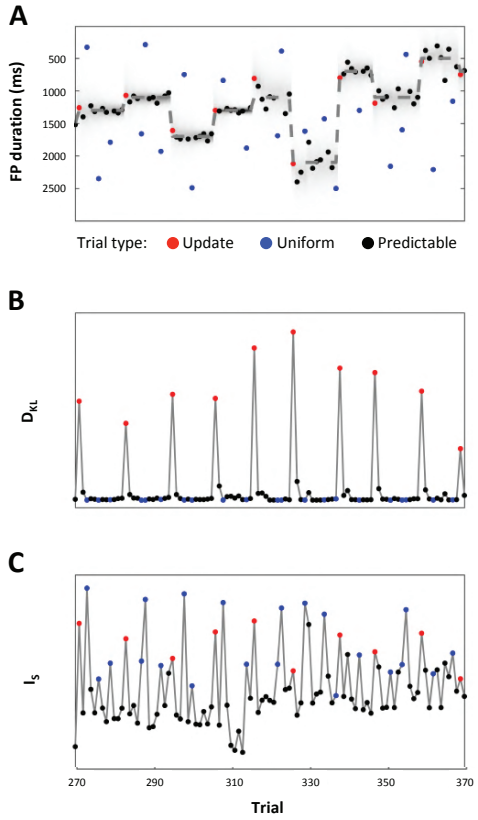

Figure 2: Ideal Bayesian observer and information-theoretic measures. All panels show the data from 100 trials. Dot colors indicate trial types as reported in the legend. **(A)** Plot of the state of the ideal Bayesian observer. The y axis shows FP duration. The dashed line indicates the mean of the generative Gaussian distribution from which update and predictable FPs were drawn. Dots indicate the true FP duration on each trial. Shading indicates the estimated probability of FP duration given the prior, p(FP|prior). **(B, C)** Model-based information-theoretic measures of updating (D_{KL}) and surprise (I_S).

Results

Behavioral analysis

Log-transformed RTs were analyzed by means of a linear mixed model (LMM) in which I_S and D_{KL} were used as explanatory variables along with the rank-order of a trial (TRIAL), and log-RT at the preceding trial (PRECEDING RT) to control for temporal trial-to-trial dependencies. Backward elimination of non-significant effects resulted in a model specified as the following Wilkinson-notation formula:

$$\log(RT) \sim \text{TRIAL} + \text{PRECEDING RT} + I_S * D_{KL} + (\text{TRIAL} + \text{PRECEDING RT} + I_S \mid \text{ID}) \quad (7)$$

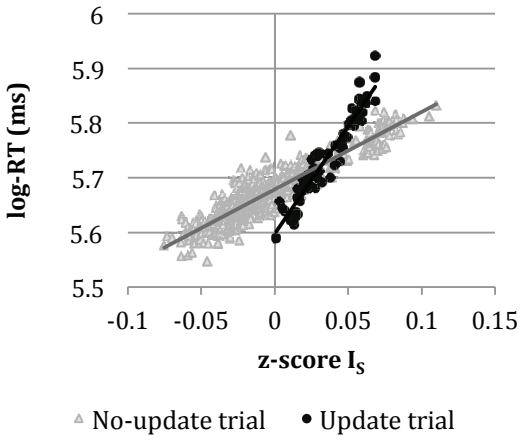

Figure 3: Interaction plot for log-transformed RTs. The plot shows the effect of surprise (I_S) for update and no-update (i.e., uniform and predictable) trials.

EEG analysis

First-level (subject-specific) analysis was performed using the *Unfold* toolbox (www.unfoldtoolbox.org) in MATLAB, which performs regression-based EEG analysis by integrating a mass-univariate approach with linear deconvolution. For the analysis we specified three events: cue onset, target onset and button press. Target onsets were, then, modeled according to the following Wilkinson-notation formula:

$$\text{EEG} \sim I_S + D_{KL} \quad (8)$$

Group-level analysis was performed using the ept-TFCE toolbox (https://github.com/Mensen/ept_TFCE-matlab) in MATLAB. Estimated D_{KL} and I_S parameters in the data space channels × epoch time points (0 - 1000 ms) were tested using threshold-free cluster enhancement (TFCE) one-sample *t*-test (number of permutations = 200000, alpha-level = .001). TFCE analysis (Fig. 4) showed that: (1) D_{KL} triggered a first series of early and fast deflections followed by a P3-like modulation; (2) I_S was associated with an early positive posterior modulation followed by a P3-like component, which emerged earlier and was less sustained compared to D_{KL}.

Conclusions

We identified EEG correlates elicited by updating of prior temporal expectations, and we showed that updating could be distinguished from surprise at the electrophysiological level. These findings are relevant for Bayesian modeling of temporal expectations. They are also of importance for the functional interpretation of the P3, one of the most widely used EEG indicators of information processing at the neural level.

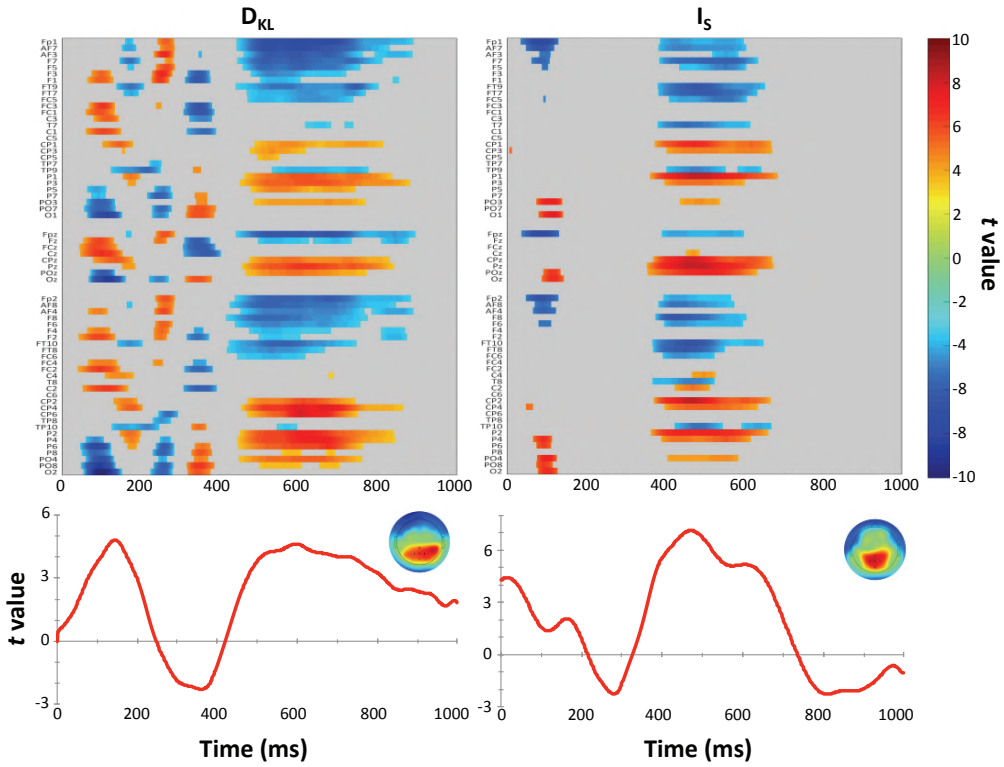

Figure 4: Mass univariate EEG results for Updating (D_{KL}; left panels) and Surprise (I_S; right panels). Raster diagrams show significant effects according to TFCE analysis. Rectangles in warm and cold colors indicate electrodes/time points significantly modulated. Trace plots depict the average *t* values pooled over the electrodes Cz, CP1, CPz, CP2, Pz. The topoplots show the *t* values averaged in the time windows from 550 to 650 ms for D_{KL}, and from 450 to 550 ms for I_S.

Acknowledgments

This work was partly funded by the European Research Council (ERC starting grant LEX-MEA, n° 313692, to A.Va). The first author acknowledges the support of the Boehringer Ingelheim Fonds in the form of a Travel Grant for his research period in the Lab of Prof. Kopp.

References

Baldi, P., & Itti, L. (2010). Of bits and wows: A Bayesian theory of surprise with applications to attention. *Neural Networks, 23*(5), 649-666.

Bueti, D., Bahrami, B., Walsh, V., & Rees, G. (2010). Encoding of temporal probabilities in the human brain. *Journal of Neuroscience, 30*(12), 4343-4352.

Herbst, S. K., Fiedler, L., & Obleser, J. (2018). Tracking Temporal Hazard in the Human Electroencephalogram Using a Forward Encoding Model. eNeuro, 5(2), ENEURO.0017-18.2018.

Janssen, P., & Shadlen, M. N. (2005). A representation of the hazard rate of elapsed time in macaque area LIP. *Nature Neuroscience, 8*(2), 234-241.

Kolossa, A., Kopp, B., & Fingscheidt, T. (2015). A computational analysis of the neural bases of Bayesian inference. *Neuroimage, 106*, 222-237.

Kopp, B. (2008). The P300 component of the event-related brain potential and Bayes' theorem. *Cognitive sciences at the leading edge*, 87-96.

Mars, R. B., Debener, S., Gladwin, T. E., Harrison, L. M., Haggard, P., Rothwell, J. C., & Bestmann, S. (2008). Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. *Journal of Neuroscience*, 28(47), 12539-12545.

Meindertsma, T., Kloosterman, N. A., Engel, A. K., Wagenmakers, E. J., Donner, T. H. (2018). Surprise about sensory event timing drives cortical transients in the beta frequency band. *Journal of Neuroscience*, 10.1523/JNEUROSCI.0307-18.2018

Nobre, A. C., Correa, A., & Coull, J. T. (2007). The hazards of time. *Current opinion in neurobiology, 17*(4), 465-470.

O'Reilly, J. X., Schuffelgen, U., Cuell, S. F., Behrens, T. E., Mars, R. B., & Rushworth, M. F. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences of the United Syayes of America, 110*(38), E3660-3669.