PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/122315

Please be advised that this information was generated on 2019-03-20 and may be subject to change.
Trial-by-Trial Fluctuations in the Event-Related Electroencephalogram Reflect Dynamic Changes in the Degree of Surprise

Rogier B. Mars,1,3,4 Stefan Debener,5 Thomas E. Gladwin,6,7 Lee M. Harrison,2 Patrick Haggard,3,8 John C. Rothwell,1 and Sven Bestmann1,2,3

1Sobell Department of Motor Neuroscience and Movement Disorders and 2Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London WC1N 3BG, United Kingdom, 3Institute of Cognitive Neuroscience, University College London, 4Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, United Kingdom, 5Medical Research Council Institute of Hearing Research, Royal South Hants Hospital, Southampton SO14 0YG, United Kingdom, 6Rudolf Magnus Institute of Neuroscience, Department of Psychiatry, University Medical Center Utrecht, Utrecht 3508 GA, The Netherlands, 7Department of Psychiatry, Stuivenberg Hospital, Antwerpen B-2-60, Belgium, and 8Department of Psychology, University College London, London WC1E 6BT, United Kingdom

The P300 component of the human event-related brain potential has often been linked to the processing of rare, surprising events. However, the formal computational processes underlying the generation of the P300 are not well known. Here, we formulate a simple model of trial-by-trial learning of stimulus probabilities based on Information Theory. Specifically, we modeled the surprise associated with the occurrence of a visual stimulus to provide a formal quantification of the “subjective probability” associated with an event. Subjects performed a choice reaction time task, while we recorded their brain responses using electroencephalography (EEG). In each of 12 blocks, the probabilities of stimulus occurrence were changed, thereby creating sequences of trials with low, medium, and high predictability. Trial-by-trial variations in the P300 component were best explained by a model of stimulus-bound surprise. This model accounted for the data better than a categorical model that parametrically encoded the stimulus identity, or an alternative model of predictability. Trial-by-trial fluctuations in the human event-related P300 can be explained by predictions made by an ideal observer keeping track of the probabilities of possible events. This provides evidence for theories proposing a direct link between the P300 component and the processing of surprising events. Furthermore, this study demonstrates how model-based analyses can be used to explain significant proportions of the trial-by-trial changes in human event-related EEG responses.

Key words: P300; single-trial EEG; information theory; surprise; attention; independent component analysis

Introduction

Late positive components of the human event-related brain potential (ERP), in particular the P300, have traditionally been associated with the processing of unexpected events (Sutton et al., 1965) (for review, see Nieuwenhuis et al., 2005). The amplitude of the P300 appears to be determined at least partly by the probability and relevance of an event (Duncan-Johnson and Donchin, 1977). Functionally, the P300 has commonly been linked to the revision of a participant’s expectation about the current task context (Donchin, 1981; Donchin and Coles, 1988; Barcelo et al., 2006), as well as the updating of task-relevant information in anticipation of subsequent events (Barcelo et al., 2008). The P300 has widely been suggested to be modulated at least in part by the anticipation of subsequent events (Barcelo et al., 2008). The P300 has been associated with the processing of rare, surprising events (Donchin, 1981) and some authors have used a terminology related to information theory to describe processes underlying generation of the P300 (Ruchkin and Sutton, 1978; Johnson, 1986; Barcelo et al., 2008).

However, we are not aware of any study that has quantified fluctuations in surprise on a trial-by-trial basis to study its impact on the P300. A number of recent computational models have been proposed that formally quantify the surprise conveyed by sensory stimuli. In these models, the surprise associated with an event relates to its improbability, given a prediction of the occurrence of all possible events (Strange et al., 2005). Computationally, it might be an efficient strategy to focus processing resources on such surprising events, because these provide the most information to an observer (Baldi, 2005). One apparent advantage of using a model-based approach to quantify the intuitive notion of surprising events is that competing models about the cognitive processes underlying observed neural data can be formally tested (Corrado and Doya, 2007). Using this approach, recent neuroimaging studies in humans have shown that activity in a wide-spread parietal-premotor network is associated with the surprise associated with the presentation of a visual stimulus (Strange et al., 2005).
Here, we asked whether trial-by-trial variations in the P300 can be explained by such a formal model of surprise and whether this provides a more parsimonious description of the data than alternative models. Healthy participants performed a choice reaction-time (RT) task while their brain activity was measured using electroencephalography (EEG). We then quantified the surprise associated with the unique stimulus sequence given to every participant and investigated whether these quantifications could explain variations in P300 on a single-trial basis. Our findings show that trial-by-trial variabilities in the P300 component are not random noise. A substantial proportion of this variability can be explained by formal quantifications of surprise, providing a direct confirmation of previous heuristics about the computations underlying the P300 component.

**Materials and Methods**

**Participants, experimental design, and data acquisition.** Twelve healthy participants (eight women, age range 18–29 years), all with normal or corrected-to-normal visual acuity participated in the experiment. All were recruited via the participants’ database of the Department of Psychology of University College London. Experimental procedures were approved by the local ethics committee and in accordance with the Declaration of Helsinki. Participants received £15 compensation for their time and travel.

Before the experiment, participants learned by trial-and-error the associations between four arbitrary visual stimuli (equated for surface area and brightness) and four button responses (using the index and middle fingers of both hands) for 60 trials. During this training, all stimuli were presented an equal number of times in random order. If participants did not perform the task without errors on the last 15 trials of the training block, it was repeated. During the main experiment participants performed 12 blocks of 60 trials of a choice reaction time task without feedback (see Fig. 1a). Visual stimuli were presented for 200 ms each, with a stimulus onset asynchrony of 2 s. Participants were required to respond to each stimulus with the previously associated button as quickly as possible, but not at the expense of accuracy. The probability of the occurrence of each event was manipulated between blocks such that the relative probabilities of events were either 0.25 for each event (low predictability), [0.4, 0.4, 0.1, 0.1] (medium predictability), or [0.7, 0.1, 0.1, 0.1] (high predictability). Participants were not informed about these probabilities. They were simply instructed to respond as quickly as possible to each presented stimulus and that the four different stimuli were randomly distributed across blocks. All stimuli occurred equally often over the course of the experiment and all stimuli had an equal behavioral relevance. Participants were given a break between blocks; they were free to initiate the subsequent block at their own pace.

The experiment was realized using the Cogent 2000 toolbox (University College London, http://www.vislab.ucl.ac.uk/Cogent2000/index.html) for Matlab (The Mathworks). EEG was recorded (bandpass filter: 0.05–100 Hz, 500 Hz sampling rate) using a Synamps2 amplifier (Neuroscan) from the following electrode positions, using Ag/AgCl electrodes mounted in an elastic electrode cap: AF3, AF4, F7, F3, Fz, F4, P8, FC5, FC1, FC2, FC6, T7, T3, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O3, O4, Oz, and left and right mastoids. Horizontal and vertical eye movements were recorded using electrodes placed lateral to both eyes and above and below the left eye. Electrode AFz served as reference during recording and the electrode common was placed on the participants’ chin. Electrode impedances were kept at <10 kΩ.

**Electrophysiological analyses.** EEG data were analyzed using EEGLAB (Delorme and Makeig, 2004), implemented in Matlab 7.1. Each participant’s EEG data were bandpass filtered (0.3–30 Hz), down-sampled to 250 Hz, and re-referenced to average reference. Subsequently, epochs of −600 to 1400 ms around the presentation of the visual stimuli were extracted from each trial and linearly detrended. During the first step of artifact rejection, epochs containing unique, nonstereotyped artifacts (swallowing, head movements, etc) were rejected. In a second step, repeatedly occurring, stereotyped artifacts were removed using independent component analysis (ICA) (Jung et al., 2000a), which has been used in a number of recent studies on P300 (Debener et al., 2005a; Eichele et al., 2005; Jongma et al., 2006). This method assumes that the EEG data recorded at the electrode level is a linear mixture of underlying brain signals and artificial signals such as eye blinks, muscle activity, cardiac signals, and line noise. The ICA algorithm (extended infomax ICA) (Makeig et al., 1996) finds an “unmixing” square matrix of the size of the number of channels, which is then matrix-multiplied with the raw data to reveal maximally temporally independent components. Each independent component can then be characterized by a time course and a scalp topography. All individual independent components whose signal and scalp topography resembled known artifacts were removed from the dataset (Jung et al., 2000a,b). The remaining components were back-projected to the scalp to reveal EEG data without the contributions of the artifacts. Epochs were baseline corrected using the interval −400–0 ms before stimulus presentation as the baseline.

From these data, single-trial P300s were estimated at electrode Pz, where this ERP component is traditionally reported to be maximal (Duncan-Johnson and Donchin, 1977; Debener et al., 2005a; Jongma et al., 2006). ERPs were created as trial averages for each participant and for each a priori stimulus category. To estimate single-trial amplitudes, for each participant, the time point at which the averaged P300s were modulated maximally by relative stimulus frequency was determined. Single-trial P300 estimates were then extracted over a window of ±60 ms around this time point of maximal modulation (cf. Jongma et al., 2006; Barcelo et al., 2008). This method was chosen over simple peak detection (Bénar et al., 2007) to capture the condition effects and improve the reliability of single-trial amplitude measures, similar to previous studies (Debener et al., 2005b).

**Ideal observers.** We modeled participants learning of the task by assuming they acted as ideal observers who learn the probability of selecting each of the four responses after presentation of the stimuli. Following previous studies (Strange et al., 2005; Harrison et al., 2006; Bestmann et al., 2008), we assume that participants start each block assuming that all events are equally likely and update their estimate of the probability of each event type on each trial, based on the events they previously observed. The same procedure was repeated for each block, i.e., the maximum number of observations was the number of trials in a block. This amounts to assuming that each participant starts each block “anew,” without memory of the previous blocks. Although future work may focus more directly on modeling different types of information transfers between blocks, previous work has shown the suitability of this assumption (Strange et al., 2005; Harrison et al., 2006; Bestmann et al., 2008).

Formally, we can consider a discrete variable, x, that can take values from 1 to K, where in our case K = 4, i.e., each trial contained one of four possible events, corresponding to the four visual stimuli and their respective responses. This distribution is parameterized by the random vector \( P(x) = [p_1, \ldots, p_K] \) (which we abbreviate using \( P(x) = p \)), whose elements \( p_k \) denote the probability of the kth event as \( P(x = k) = p_k \). This is a multinomial distribution, where \( p_k \) is the probability of the kth trial type occurring. We will refer to this as the generative distribution, as it was from this that a sequence of events were sampled. A simple example is a coin toss where \( K = 2 \). The probability of “heads” and “tails” is then given by \( P(x = heads) = p_1 \) and \( P(x = tails) = p_2 \), respectively, which sum to one.

The aim of the observer, i.e., the participant, is to estimate the distribution of event probabilities, using the information conveyed by the encountered train of stimuli. In other words, the observer tries to estimate parameters, i.e., probabilities, contained in the vector p. Given a sample of j events, denoted by \( X = [x_1, \ldots, x_j] \), there are a number of ways to estimate these. An issue with using the maximum likelihood estimate is that the observer’s estimate of \( p_k \) will be zero if event k has not been observed. For example, if only three tosses of a coin are sampled with the outcome of three heads, then the estimate of the probability of heads is equal to one. A prediction based on this small sample would be that a tail could never occur, which is contrary to intuition. This can be resolved by giving the observer prior knowledge, as done in the Bayesian paradigm. We can assure that the observer has a greater than zero expectation of all stimuli occurring by giving it a uniform prior, i.e., by having it
Timeframe and examples of surprise models. a, Example of the task. Participants were trained on stimulus–response associations (top right) and then performed a simple choice–RT task. Visual stimuli were presented every 2 s for 200 ms each, requiring the participants to respond with the previously associated button. b, Raster plots of events sampled from two distributions where the probability of event 1 was either greater than all other events (high predictability environment) or the same (low-predictability environment). c–e, For each unique stimulus sequence models of blockwise stimulus-bound surprise (c), I_p stimulus-bound surprise with no forgetting between blocks (d), I_e and KL surprise (e).

Quantifying surprise. Following Strange et al. (2005), we can quantify the surprise, I, on each trial as follows (cf. Shannon, 1948):

\[ I(x^i = k) = -\log_2 p\tilde{k}^i. \]  

This states that the surprise of observing event type \( k \) at the \( j \)th trial is equal to the negative log of its predicted probability given all preceding trials. Accordingly, the amount of surprise conveyed by the occurrence of an event is high when an infrequent stimulus occurs in a stimulus sequence with high predictability. For example, in highly predictable blocks \( \{0.7, 0.1, 0.1, 0.1\} \), the probability of one particular event is high, whereas the other three events occur only rarely. Given repeated samples of this distribution, these low frequency events are more surprising. An event is more surprising when occurring with 0.10 probability, compared with an event with a 0.70 probability of occurring (Fig. 1c). Note that in this experiment the generative distribution did not include dependencies between consecutive events. That is, the event at one time did not depend on earlier events. This is the same as in the study by Strange et al. (2005) and different to that investigated by Harrison et al. (2006), where the current event depended on the previous. Given the assumption that participants start each block anew, we refer to this model as blockwise surprise, \( I_p \).

Alternative models. We compared the model of the previous section with a number of alternatives. The ideal observer described above assumed the generative model being stationary, i.e., unchanging within a block. This assumption is ideal in that it matches the true distribution used to generate trial types in the experiment. Furthermore, the model described above assumes participants start each block anew with the expectation that all events occur equally often, i.e., with a uniform (i.e., flat, uninformative) prior. Alternatively, one might expect that participants view each block merely as a continuation of the previous block, such that the experiment can be seen as one long session. We therefore also created a model based on an observer with no forgetting, here referred to as experiment-wise surprise, \( I_e \) (Fig. 1d). This is suboptimal because contingencies did change from block to block.

An alternative formulation of surprise has been suggested by Baldi et al. (Baldi, 2002; Itti and Baldi, 2006), based on the Kullback–Leibler (KL) divergence (Kullback, 1959; Clover and Thomas, 1999). The KL divergence is a scalar quantity that summarizes the difference between two probability distributions. In our case, it is used to measure the change in
belief about the stimulus probabilities, P(x), after an event (i.e., visual stimulus). If this change is large then the event has a high degree of "surprise", compared with one that has little or no effect. For the current experimental setting, the Kullback–Leibler divergence (KL surprise) at trial j is a function of the current, prior, and posterior distributions, D' – 1 and D' (cf. Baldi, 2002):

$$KL_{surprise} = KL(D' – 1, D')$$

In words, the (blockwise) KL surprise is the "distance" between the distributions before and after observing the jth trial. Intuitively this means that events can be quantified in terms of how much they change posterior beliefs.

The difference between the KL divergence measure of surprise (see Fig. 1e) and surprise as defined in Equation 4 is that the former is an average quantity, i.e., summed over all probabilities in the distribution. In contrast, the latter is a function of the predicted probability of an observed event, i.e., trial type presented to the subject. In other words, the KL divergence is a distance measure between the current, prior, and posterior distributions, whereas I_p is a function of the predicted probability of the observed event, i.e., just one event and not an average over all possible events. The KL surprise measure relates to those proposed by Ruchkin et al. (1988) and Kopp (2007) to account for variations in P300.

Last, we included a conventional explanation using a categorical model of events parametrically modulated by the probability of occurrence. In this model, each trial had one of the values [0.10, 0.25, 0.40, 0.70]. This regressor models variance related to stimulus probability within a block and does not take into account any learning; hence it is similar to the traditional method of averaging ERP data over a priori probability of a stimulus within a block) showed that participants responded slower to stimuli with a lower (0.10; 573 ± 21 ms; RT ± SEM) probability of occurrence, than stimuli with a higher probability of occurrence (0.70; 427 ± 18 ms): F(3,9) = 107.904, p < 0.001. Participants responded incorrectly on 4.2% (SEM ± 0.69) of trials, making more errors in response to less frequent stimuli (F(3,9) = 15.01, p = 0.001).

Results

Behavioral results

Inspection of average reaction times on correct trials showed that participants’ reaction times were affected by changes in probabilistic context. A repeated-measures ANOVA with factor “probability” (4 levels: 0.1, 0.25, 0.40, and 0.70, indicating the overall a priori probability of a stimulus within a block) showed that participants responded slower to stimuli with a lower (0.10; 573 ± 21 ms; RT ± SEM) probability of occurrence, than stimuli with a higher probability of occurrence (0.70; 427 ± 18 ms): F(3,9) = 107.904, p < 0.001. Participants responded incorrectly on 4.2% (SEM ± 0.69) of trials, making more errors in response to less frequent stimuli (F(3,9) = 15.01, p = 0.001).

Event-related potentials: trial-averaged results

Figure 2 shows the scalp topographies and grand average ERP over all trials, showing the traditional distribution of the P300. Our statistical analyses focused on the single-trial estimates of P300. Central latency of the time window used for single-trial P300 estimates was on average 531 (SEM ± 24, range 392–660 ms) after stimulus onset. To verify that our averaged single-trial P300 estimates showed the same scalp topography and ordering by stimulus probability as commonly reported for average P300s, averaged single-trial estimates were entered into a repeated measures ANOVA with factors electrode (4 levels: Fz, Cz, Pz, and Oz) and probability (4 levels: 0.10, 0.25, 0.40, and 0.70). This analysis showed that average single-trial estimates differed reliably between electrodes (F(3,9) = 19.484, p < 0.001) and probability (F(3,9) = 14.936, p < 0.001). The difference in average P300 for each probability was most pronounced at electrode Pz (electrode × probability interaction: F(9,33) = 7.659, p < 0.001), as is well established for the P300 (Duncan-Johnson and Donchin, 1977; Debener et al., 2005a) (Fig. 2c). This ordering of single-trial estimates is not due simply to a potential confounding relationship between trials rejected by the artifact correction and a priori stimulus probability, as there as no systematic relationship between the two (F(3,33) = 0.008, not significant).

Event-related potentials: model-based single-trial analyses

Having replicated the traditional P300 effects in choice reaction time tasks, we subsequently focused on the model-based analyses of the single-trial P300 estimates, following the procedure advocated by MacKay (1992). First, each model was fitted to the data using the procedure described above. Second, the model evidence
estimates for each participant. Error bars indicate STD. bar indicates SEM.

For the P300, we found that averaged P300s over central-parietal electrode sites interacted with the relative probability of event occurrence. Importantly, a model of the surprise within a block of trials provided a more parsimonious explanation of single-trial P300 changes than alternative models, including a categorical model of stimulus frequency, an alternative model of surprise based on the KL divergence, and surprise without forgetting. This novel model-based approach applied to single-trial EEG data allows for a formal quantification of the psychological variable “surprise” and its relationship to the psychophysiological marker P300.

Previous studies on the P300 have introduced the term “subjective probability” to denote that it is participants’ estimation of the environment that is crucial in predicting modulations in P300 (Donchin and Coles, 1988). This has led to the suggestion that P300 reflects the updating of information in anticipation of subsequent information processing (Sutton et al., 1965; Nieuwenhuis et al., 2005; Verleger et al., 2005; Barcelo et al., 2008). The P300 has previously been linked with information theoretic concepts (Ruchkin and Sutton, 1978; Johnson, 1986; Barcelo et al., 2006, 2008; Barcelo and Knight, 2007) or Bayes’ theory (Kopp, 2007). Here, we draw on information theoretic concepts to investigate the trial-by-trial influence of stimulus-bound surprise on P300 variation. Characterizing the subjective estimate of task probabilities has only recently become a major focus of research in cognitive and neurosciences (Oaksford and Chater, 2007). In the present case, we used a model of how the “subjective probability” is represented and updated over time, rather than how it changes on average.

To achieve this, the present approach combines two novel methodologies that, to our knowledge, have not been combined earlier in studies of event-related potentials. First, the model-based approach provides models about the trial-by-trial variability of the surprise conveyed by events experienced by participants. Behavioral data indicated that on average participants responded slower to less frequently occurring, i.e., more surprising, events. Consistent with earlier reports on the P300, probability has been linked with information theoretic concepts (Ruchkin and Sutton, 1978; Johnson, 1986; Barcelo et al., 2006, 2008; Barcelo and Knight, 2007) or Bayes’ theory (Kopp, 2007). Here, we draw on information theoretic concepts to investigate the trial-by-trial influence of stimulus-bound surprise on P300 variation. Characterizing the subjective estimate of task probabilities has only recently become a major focus of research in cognitive and neurosciences (Oaksford and Chater, 2007). In the present case, we used a model of how the “subjective probability” is represented and updated over time, rather than how it changes on average.

To achieve this, the present approach combines two novel methodologies that, to our knowledge, have not been combined earlier in studies of event-related potentials. First, the model-based approach provides models about the trial-by-trial variability of the surprise conveyed by events experienced by participants. Behavioral data indicated that on average participants responded slower to less frequently occurring, i.e., more surprising, events. Consistent with earlier reports on the P300, probability has been linked with information theoretic concepts (Ruchkin and Sutton, 1978; Johnson, 1986; Barcelo et al., 2006, 2008; Barcelo and Knight, 2007) or Bayes’ theory (Kopp, 2007). Here, we draw on information theoretic concepts to investigate the trial-by-trial influence of stimulus-bound surprise on P300 variation. Characterizing the subjective estimate of task probabilities has only recently become a major focus of research in cognitive and neurosciences (Oaksford and Chater, 2007). In the present case, we used a model of how the “subjective probability” is represented and updated over time, rather than how it changes on average.

To achieve this, the present approach combines two novel methodologies that, to our knowledge, have not been combined earlier in studies of event-related potentials. First, the model-based approach provides models about the trial-by-trial variability of the surprise conveyed by events experienced by participants. Behavioral data indicated that on average participants responded slower to less frequently occurring, i.e., more surprising, events. Consistent with earlier reports on the P300, probability has been linked with information theoretic concepts (Ruchkin and Sutton, 1978; Johnson, 1986; Barcelo et al., 2006, 2008; Barcelo and Knight, 2007) or Bayes’ theory (Kopp, 2007). Here, we draw on information theoretic concepts to investigate the trial-by-trial influence of stimulus-bound surprise on P300 variation. Characterizing the subjective estimate of task probabilities has only recently become a major focus of research in cognitive and neurosciences (Oaksford and Chater, 2007). In the present case, we used a model of how the “subjective probability” is represented and updated over time, rather than how it changes on average.

To achieve this, the present approach combines two novel methodologies that, to our knowledge, have not been combined earlier in studies of event-related potentials. First, the model-based approach provides models about the trial-by-trial variability of the surprise conveyed by events experienced by participants. Behavioral data indicated that on average participants responded slower to less frequently occurring, i.e., more surprising, events. Consistent with earlier reports on the P300, probability has been linked with information theoretic concepts (Ruchkin and Sutton, 1978; Johnson, 1986; Barcelo et al., 2006, 2008; Barcelo and Knight, 2007) or Bayes’ theory (Kopp, 2007). Here, we draw on information theoretic concepts to investigate the trial-by-trial influence of stimulus-bound surprise on P300 variation. Characterizing the subjective estimate of task probabilities has only recently become a major focus of research in cognitive and neurosciences (Oaksford and Chater, 2007). In the present case, we used a model of how the “subjective probability” is represented and updated over time, rather than how it changes on average.
combining previous knowledge with a current event. Second, although previous studies have compared predictions from computational models qualitatively with the results from averaged evoked potentials (Nieuwenhuis et al., 2002; Cohen and Ranganath, 2007), recent advances in EEG data processing, such as ICA (Eichele et al., 2005; Debener et al., 2006; Jongsma et al., 2006), now allow for trial-by-trial analyses. We here combine this model-based approach and the single-trial data analysis by formally testing the predictions of the model to the data. Moreover, this combined approach allows for comparing the evidence of different models, given the observed ERP data. The present approach differs from that used by Duncan-Johnson and Donchin (1977). These authors used regression analysis to fit single-trial P300 amplitudes to a model of a priori stimulus probability. Their approach thus focused on the overall true probabilities that were a priori known to the experimenter, but not the participant. In contrast, we here used a formal model of how participants’ learned these probabilities over the course of the experiment. In addition, we scrutinized our model against several alternative models.

We have modeled surprise $I_s$ here according to measures described by information theory (Shannon, 1948; Clover and Thomas, 1999), consistent with previous studies showing that surprise is associated with activity in an extended corticothalamic network (Strange et al., 2005; Harrison et al., 2006) and changes in corticospinal excitability (Bestmann et al., 2008). Here, we assumed that events were stationary andunchanging within a block, matching the true generative distribution from which events were sampled. Therefore, all previous blocks and events were forgotten in an optimal way and trials within the current block were weighted equally. Note that this assumption is ideal in relation to the actual experimental paradigm but assumes participants were privy to different blocks of events being sampled from different distributions. We therefore included an alternative model in which our ideal observers had suboptimal (i.e., no) forgetting with respect to the actual experimental paradigm. In the present experiment, a model of an ideal observer beginning each block with flat priors, was superior to a model without forgetting.

Moreover, we also compared our model to an alternative measure of surprise based on the Kullback–Leibler divergence. This latter measure can be taken as a formal description of “equivocation” that has been suggested to underlay the generation of the P300 amplitudes to a model of a priori stimulus probability. Their approach thus focused on the overall true probabilities that were a priori known to the experimenter, but not the participant. In contrast, we here used a formal model of how participants’ learned these probabilities over the course of the experiment. In addition, we scrutinized our model against several alternative models.

To conclude, model-based single-trial analyses can be used for testing hypotheses of event-related EEG fluctuations. This approach provides a bridge between cognitive theories and more formal neurophysiological models of the P300 ERP. The focus on single-trial EEG data provides a more direct link to behavior and neural processing than averaged EEG activity (Debener et al., 2005). This is supported by our observation that P300 trial-by-trial amplitude fluctuations are not random noise, and can be explained by a formal model of surprise experienced in the context of a behavioral task. Our findings provide direct evidence for theories linking the P300 component and the processing of surprising events.

References
Barcelo F, Escera C, Corral MJ, Perañález JA (2006) Task switching and nov-