UNIVERSITÀ DEGLI STUDI DI PADOVA

CORSO DI DOTTORATO IN PSYCHOLOGICAL SCIENCES

XXXI CICLO

BAYESIAN MODELING OF TEMPORAL EXPECTATIONS IN THE HUMAN BRAIN

Supervisore

Prof. Antonino Vallesi

Dottorando

Antonino Visalli

To my family, Laila, and Lex-Mea mates, thank you.

ABSTRACT **3**

GENERAL INTRODUCTION **9**

1.1 TEMPORAL PREPARATION 10

1.2 BAYESIAN BELIEF UPDATING 13

1.3 PROJECT OVERVIEW 20

TEMPORAL BELIEF UPDATING AND SURPRISE MODULATE COGNITIVE CONTROL NETWORKS **23**

2.1 INTRODUCTION 23

2.2 METHODS 26

2.2.1 Participants *26*

2.2.2 Task and procedure *27*

2.2.3 Normative Bayesian learner *30*

2.2.4 Model-based measures of updating and surprise *31*

2.2.5 Behavioral data analysis *34*

2.2.6 fMRI data analysis *34*

2.3 RESULTS 38

2.3.1 Behavioral results *38*

2.3.2 Whole-brain fMRI results *39*

2.4 DISCUSSION 43

ELECTROPHYSIOLOGICAL CORRELATES OF TEMPORAL BELIEF UPDATING AND SURPRISE **49**

3.1 INTRODUCTION 49

3.2 METHODS 52

3.2.1 Participants *52*

3.2.2 Task and procedure *53*

3.2.3 EEG data acquisition *54*

3.2.4 Normative Bayesian learner and regressors *55*

3.2.5 EEG data analysis *55*

3.3 RESULTS 57

3.3.1 Behavioral results *57*

3.3.2 Electrophysiological results *58*

3.4 DISCUSSION 64

BEYOND EXPLICIT INFERENCE: EEG CORRELATES OF IMPLICIT UPDATING OF TEMPORAL EXPECTATIONS 69

4.1 INTRODUCTION 69

4.2 METHODS 71

4.2.1 Participants 71

4.2.2 Task and procedure 72

4.2.3 Normative Bayesian learner and regressors 72

4.2.4 EEG data analysis 77

4.3 RESULTS 78

4.3.1 Electrophysiological results 78

4.3.2 EEG results, between-study comparison 78

4.4 DISCUSSION 87

GENERAL DISCUSSION 93

ACKNOWLEDGEMENTS 101

REFERENCES 103

Abstract

The ability to predict when a relevant event might occur is critical to survive in our dynamic and uncertain environment. This cognitive ability, usually referred to as *temporal preparation*, allows us to prepare temporally optimized responses to forthcoming stimuli by anticipating their timing: from safely crossing a busy road during rush hours, to timing turn taking in a conversation, to catching something in mid-air, are all examples of how important and ubiquitous temporal preparation is in our everyday life (e.g., Correa, 2010; Coull & Nobre, 2008; Nobre, Correa, & Coull, 2007).

In laboratory settings, temporal preparation has been traditionally investigated, in its implicit form, through the "variable foreperiod paradigm" (see Coull, 2009; Niemi & Näätänen, 1981, for a review). In such a paradigm, the foreperiod is a time interval of variable duration that separates a warning stimulus and a target stimulus requiring a response. What is usually observed with this paradigm is that response times (RTs) reflect the temporal probability of stimulus onset: RTs decrease with increasing probability. This implies that participants learn to use the information implicitly afforded by the passage of time and that related to the temporal probability of the onset of the target stimulus (i.e., hazard rate; Janssen & Shadlen, 2005). In other words, it seems that they are able to use predictive internal models of event timing in order to optimize behaviour.

Despite previous studies have started to investigate which brain areas encode temporal probabilities (i.e., predictive models) to anticipate event onset (e.g., Bueti, Bahrami, Walsh, & Rees, 2010; Cui, Stetson, Montague, & Eagleman, 2009; also see Vallesi et al., 2007), to our knowledge, there is no evidence on *how* the brain does

form and update such predictive models. Based on such premises, the **overarching goal** of the present PhD project was to pinpoint the neural mechanisms by which predictive models of event timing are dynamically updated. Moreover, given that in real life updating usually occurs in the presence of surprising events (i.e. low probable events under a predictive model), it is challenging to disentangle between updating and surprise (O'Reilly et al, 2013). Therefore, our second and interrelated research goal was to understand whether, and to which extent, it is possible to dissociate between the neural mechanisms specifically involved in updating and those dealing with surprising events that do not require an update of internal models. To accomplish our research goals, we capitalized on both state-of-the-art methodologies [i.e., functional magnetic resonance imaging (fMRI) and electrophysiology (EEG)] and computational modelling. Specifically, we considered the brain like a Bayesian observer. Indeed, Bayesian frameworks are gaining increasing popularity to explain cognitive brain functions (Friston, 2012). In a nutshell, the construction of computational Bayesian models allows us to quantitatively describe temporal expectations in terms of probability distributions and to capture updating using Bayes' rule.

In order to accomplish our goals, the present PhD project is composed of three studies. In the first two studies we implemented a version of the foreperiod paradigm in which participants could predict target onsets by estimating their underlying temporal probability distributions. During the task, these distributions changed, hence requiring participants to update their temporal expectations. Furthermore, a simple manipulation of the colors in which the target were presented (cf., O'Reilly et al., 2013) allowed us to independently vary updating and surprise across trials. Then, we constructed a *normative Bayesian learner* (a computational model adapted from

O'Reilly et al., 2013) in order to obtain an estimate of a participant's temporal expectations on a trial-by-trial basis. In **Study 1**, trial-by-trial fMRI data acquired during our foreperiod paradigm were correlated with two information theoretical parameters calculated with reference to our Bayesian model: the Kullbach-Leibler divergence (D_{KL}) and the Shannon's information (I_S). These two measures have been previously used to formally describe belief updating and surprise associated with events under a predictive model, respectively (e.g., Baldi & Itti, 2010; Kolossa, Kopp, & Fingscheidt, 2015; O'Reilly et al., 2013; Strange et al., 2005). Our results showed that the fronto-parietal network and the cingulo-opercular network were differentially involved in the updating of temporal expectations and in dealing with surprising events, respectively.

Having successfully validated the use of Bayesian models in our first fMRI study and dissociated between updating and surprise, the next step was to investigate the temporal dynamics of these two processes. Do updating and surprise act on similar or distinct processing stage(s)? What is the time course associated with the two? To address these questions, in **Study 2** participants performed our adapted foreperiod task (same task as in Study 1) while their EEG activity was recorded. In this study, we relied on the literature on the P3 (a specific ERP component related to information processing) and the Bayesian brain (e.g., Kopp, 2008; Kopp et al., 2016; Mars et al., 2008; Seer, Lange, Boos, Dengler, & Kopp, 2016). Importantly, however, we also took advantage from the combination of a mass-univariate approach with novel deconvolution methods to explore the entire spatio-temporal pattern of EEG data. This enabled us to extend our analyses beyond the P3 component. Results from study 2 confirmed that surprise and updating can be differentiated also at the electrophysiological level and that updating elicited a more complex pattern than

surprise. As regards the P3 in relation to the literature on the Bayesian brain (Kolossa, Fingscheidt, Wessel, & Kopp, 2013; Kolossa et al., 2015; Mars et al., 2008), our findings corroborated the idea that such a component is selectively modulated by surprise and updating.

While in Studies 1 and 2, participants were explicitly encouraged to form and update temporal expectations using the target color, in **Study 3** we wanted to make a step further by asking whether the use of a more implicit task structure might influence the construction of the predictive internal model. To that aim, during the foreperiod task designed for the third study, participants were not explicitly informed about the presence of the underlying temporal probability distributions from which target onsets were drawn. In this way, we aimed to investigate behavioural and EEG differences in the way participants learnt to form and updated temporal expectations when changes in the underlying distributions were not explicitly signalled. Critically, we again found that surprise and updating could be differentiated. Moreover, coupled with the results from study 2, we isolated two EEG signatures of the inferential process underlying updating of prior temporal expectations, which responded to both explicit and implicit contextual changes.

Overall, we believe that the results of the present PhD project will further our understanding of the cognitive processes and neural mechanisms that allow us to optimize our temporal preparation abilities.

Chapter 1

General introduction

To get access to my supervisor's office, I usually take the elevator. On one occasion I was waiting for it together with other people. While I was still entering the elevator, the door "unexpectedly" started to close hitting my shoulder. This embarrassing situation happened again a few times more before I learnt that the waiting interval between door opening and closing was very short and to update my incorrect temporal expectation in such a way to safely enter the elevator in future occasions!

The example above illustrates the importance of the ability to accurately predict the likely moment at which an event might occur in everyday situations, an ability usually labeled "temporal preparation" (Correa, 2010; Nobre, Correa, & Coull, 2007). While previous studies have started to unveil the neural mechanisms by which temporal expectations are updated, a direct modeling of how our brain faces this key task is still poorly estimated. To address this issue, in the present thesis we combined a temporal preparation task with Bayesian modeling during either functional magnetic resonance imaging (fMRI) or electrophysiological (EEG) studies. The reasons why we believe that a Bayesian approach would be particularly suited to investigate temporal preparation are outlined below. Specifically, in order to fully appreciate the rationale behind the work presented here, we will begin by briefly

reviewing the literature on both temporal preparation and Bayesian belief updating. In the last part of the Introduction, we will then present an overview of the project in which we make a link between the literature on temporal preparation and Bayesian models.

1.1 Temporal preparation

In laboratory settings, temporal preparation has been usually studied through the "foreperiod paradigm" (see Coull, 2009; Niemi & Näätänen, 1981, for reviews). The foreperiod is the time interval that separates a warning stimulus from a subsequent target that calls for a fast and accurate response. When the foreperiod duration is kept constant throughout a block of trials (e.g., in one block the target always appears after 500 ms, whereas in another block it does so after 1000 ms), participants' reaction times (RTs) are usually faster for the short rather than the long block of trials, a phenomenon known as the "fixed foreperiod effect" (e.g., Bausenhart, Rolke, & Ulrich, 2008; Mattes & Ulrich, 1997; Vallesi, McIntosh, Shallice, & Stuss, 2009). However, if short and long foreperiod durations are randomly intermixed across the trials and each one has the same a priori probability of being presented, participants will be faster for targets appearing after long than short foreperiods, i.e., "the variable foreperiod effect" (Niemi & Näätänen, 1981; Woodrow, 1914).

The different findings associated with fixed and variable foreperiod paradigms have been traditionally explained by two mechanisms: time estimation and monitoring the conditional probability of target occurrence, respectively. Since in the fixed paradigm, uncertainty in time estimation will increase as a function of the time interval being estimated (i.e. scalar theory, Gibbon, 1977), it follows that RTs will also increase with longer durations. The scenario instead changes for the variable foreperiod paradigm. Here, as time goes by during the trial and the target has not yet appeared after the shorter foreperiods, participants may infer that it will surely occur after the longest ones, provided that there are no catch trials, which explains the RT advantage for long foreperiod trials. This pattern of data can be

formally described in terms of a mechanism monitoring the hazard function, that is the conditional probability that the target will occur given that it has not yet occurred, and exploiting it to optimize preparation (Nobre et al., 2007).

Converging evidence from fMRI (Vallesi, McIntosh, Shallice, et al., 2009; A. Vallesi, McIntosh, & Stuss, 2009), transcranial magnetic stimulation (Vallesi, Shallice, & Walsh, 2007), and neuropsychological studies (Stuss et al., 2005; Trivino, Correa, Arnedo, & Lupianez, 2010; Vallesi et al., 2007) points to the involvement of prefrontal areas, in particular the right dorsolateral prefrontal cortex, in the variable foreperiod effect. However, these studies most often used few discrete foreperiod durations (e.g., two), which makes it difficult to model how behavioural and neural responses are shaped by the hazard function. Crucial in this regard is the seminal work by Janssen and Shadlen (2005). The authors trained rhesus monkeys to make eye movements to peripheral targets presented in a foreperiod task. Foreperiod durations were drawn from either a bimodal or unimodal continuous distribution. They found that monkeys' RTs and the firing rate of neurons in the lateral intraparietal area both correlated with the respective hazard functions of unimodal or bimodal duration distributions. Hence, this study provided strong evidence that temporal preparation is accomplished by combining both prior knowledge about foreperiod duration and the elapse of time (i.e., hazard function).

Janssen and Shadlen's (2005) findings have been replicated in humans in both fMRI (Bueti, Bahrami, Walsh, & Rees, 2010) and EEG studies (Herbst, Fiedler, & Obleser, 2018). Bueti and colleagues (2010) tested participants using the same task as Janssen and Shadlen and found that activity in V1 and extrastiate visual areas, together with the parietal cortex and motor regions (SMA, cerebellum), correlated with the hazard function. More recently, Herbst and colleagues (2018) showed that the EEG signal obtained from three different foreperiod distributions was modulated by the associated hazard function and that the signal tracking the hazard function was reconstructed in the supplementary motor area.

Another popular paradigm to study temporal preparation is the temporal orienting task (Kingstone, 1992). In this paradigm, which represents the temporal analogue of Posner's spatial orienting task (Posner, Snyder, & Davidson, 1980), the

warning signal acts as an explicit cue that predicts with high probability (e.g., 75%) the specific foreperiod duration (i.e., short versus long) after which the target would occur. Temporal orienting effects are typically reflected at the short time interval by faster and more accurate responses to validly-cued targets as compared to targets occurring earlier than expected. At the long time interval, temporal orienting effects are usually smaller or even absent because participants will reorient their attention to the long interval if the target has not appeared early as expected, which counteracts the negative consequences of an invalid temporal expectation (Correa, Lupianez, Madrid, & Tudela, 2006; Coull & Nobre, 1998).

Temporal orienting of attention is usually associated with greater activity in the left inferior parietal cortex (Cotti, Rohenkohl, Stokes, Nobre, & Coull, 2011; Coull & Nobre, 1998; Davranche, Nazarian, Vidal, & Coull, 2011).

Summing up foreperiod and temporal orienting studies, it is clear that there should be some functional differences in the way temporal expectations are developed in each task. Namely, temporal orienting tasks use fixed and constant cues that indicate a priori the likely moment in time at which the target might occur. Conversely, in variable foreperiod paradigms temporal expectations evolve over time and need to be updated. This key difference between temporal orienting and foreperiod paradigms led Coull and colleagues (2016) to surmise that, in Bayesian terms, the temporal predictability afforded by the two "can be considered as equivalent to prior and posterior probability, respectively". To this end, the authors ran an fMRI study in which they compared the benefits of temporal orienting (the "prior" in their reasoning) and foreperiod ("posterior") effects. Results showed that the left inferior parietal cortex was engaged by both the temporal cue (prior) and the hazard function (posterior), whereas the right inferior frontal cortex was only engaged by the hazard function. Despite interesting, however, a direct modeling of the data within a Bayesian framework was missing in that fMRI study and, to our knowledge, in all the other fMRI studies that have so far tested temporal preparation (e.g., Vallesi et al., 2009a). In the following paragraphs, we will make indeed evident that in Bayesian terms updating and posteriors are, in a strict sense, terms that cannot be related to the concepts used by Coull and colleagues (2016).

Before going into the details of how we modeled temporal updating using a Bayesian approach, we will briefly touch upon the main features of Bayesian models.

1.2 Bayesian belief updating

Imagine being in a well-lighted room. Looking at the objects in the environment, you have the naïve impression that what you perceive is an exact copy of what is around you. However, it is enough that the light grows dim to realize that you start perceiving with uncertainty. The fact that we continuously deal with uncertain information becomes much more evident if we move from vision to hearing or possibly even more to time. Consequently, a key function of our brain is to infer the possible causes of the world from uncertainty. This leads to the idea that brain processes have a probabilistic nature. In this regard, Bayesian frameworks are gaining increasing popularity to explain cognitive brain functions.

The strength of a Bayesian approach is that it provides a way to formalize inferential mechanisms necessary to process information under uncertainty. According to the Bayesian brain hypothesis, information is "represented by a conditional probability density function over the set of unknown variables– the posterior density function" (Knill & Pouget, 2004). For example, when you feel an unseen object in a bag, the brain tries to infer the causes of your sensation (i.e., which object you are touching) based on a model of the interior of the bag. This inferential process is formally expressed using Bayes' rule as:

$$P(A|B) \propto P(B|A)P(A). \tag{1.1}$$

From the formula, your beliefs about which object your are touching are expressed as a *posterior* distribution, P(A|B), that is the probability of many possible objects of being the object you feel given the available sensory information. These beliefs are derived by combining the relative *likelihood* of feeling that sensation given different possible objects, P(B|A), with our *prior* beliefs about the probability

of different objects of being in the bag, P(A). In sum, a Bayesian observer represents beliefs as probability distributions interpreting new information with respect to prior knowledge.

Experimentally, a Bayesian approach can be applied through the implementation of an ideal observer to make predictions about behavioral or neural data. A Bayesian ideal observer is a hypothetical participant who, using Bayesian inference, performs a specific task in an optimal way, consistently with the specified information and constraints (Geisler, 2011). As a result, we can look into the "mind" of our ideal participant in order to derive measures about the internal representation and processing of task information. These measures can be used, then, as a benchmark to predict behavior and brain activity of people performing the same task.

The Bayesian framework has been used to investigate several brain processes across many cognitive domains, such as visual processing, multisensory integration, sensorimotor integration, and decision-making (see Penny, 2012; Vilares & Kording, 2011; Yuille & Kersten, 2006, for reviews). One key aspect in this literature, and particularly relevant for the present dissertation, is Bayesian belief updating. Based on the episode reported at the beginning of the chapter, it is clearly important to have accurate beliefs about events in order to predict environmental contingences and behave in a more efficient way. Hence, the significance of studying the processes involved in maintaining appropriate beliefs about the environment. In the Bayesian framework, the brain iteratively derives updated beliefs (posterior) from prior beliefs given new observations (i.e., Bayes' rule). Although in stable environments belief updating is negligible (i.e., differences between prior and posterior are very small), the importance of this process becomes evident when the probabilistic structure of the events is unknown or changeable. To make an example about possible experimental situations as reported in O'Reilly and Mars (O'Reilly & Mars, 2015), participants at the beginning of an uncued Posner task (Posner et al., 1980) have no useful beliefs about the probability associated with target location. To improve their performance, participants need to update beliefs on a trial-by-trial basis to accurately predict target location, thus, enhancing information processing

and response selection. Despite belief updating is at the core of the Bayesian brain hypothesis (Knill & Pouget, 2004), researchers have started only recently to investigate the mechanisms underlying belief updating in the brain. In the remainder of the paragraph, we will briefly review the literature on fMRI and EEG studies of Bayesian belief updating.

Concerning fMRI, Bayesian belief updating has been investigated in different cognitive domains. In the spatial domain, for example, Vossel and colleagues (2015) investigated the neural mechanisms underlying Bayesian belief updating in the deployment of spatial attention. To this aim, the authors implemented a Posner cueing task in which they varied cue validity rate and applied a hierarchical Bayesian learning model (Mathys et al., 2014) to quantify trial-by-trial belief updating. The results showed the involvement of three brain regions in Bayesian belief updating, namely, right frontal eye fields (FEF), right temporo-parietal junction (TPJ) and putamen. Furthermore, they showed that effective connectivity from TPJ to other brain areas was modulated by updating.

Other fMRI studies that used a normative Bayesian learner describing belief updating showed that the anterior cingulate cortex (ACC) reflected increased uncertainty during evidence accumulation in decision making (Behrens, Woolrich, Walton, & Rushworth, 2007; Stern, Gonzalez, Welsh, & Taylor, 2010). In the field of decision making, Waskom and colleagues (2017) devised a context-dependent perceptual decision task in which participants were cued to make a decision either on the color of random dot stimuli or their motion. Frequency about the cued dimension varied during the task. Violations of expectations were associated with increased activity in bilateral inferior frontal sulcus (IFS), bilateral intraparietal sulcus (IPS), posterior cingulate cortex (PCC) and middle superior parietal lobe (mSPL).

A common aspect in all the fMRI studies presented so far is the fact that belief updating has been driven by surprising events that violated prior expectations. This is intuitive since generally speaking a surprising event leads to an update of our prior beliefs. However, not always surprise (violations of expectations) is associated with updating and actually surprise and updating represent two distinct constructs that have been often conflated. We illustrate this point through the “white noise”

paradox (Barto, Mirolli, & Baldassarre, 2013; Itti & Baldi, 2005). Imagine yourself watching a "snow" television screen. Each frame is very surprising given the high number of possible random combinations of pixel patterns, each of which is associated with a low probability of occurrence. Notwithstanding this, surprise does not lead to a relevant change in the agent's internal model since a random noise will end up being increasingly more expected.

As it will become clear later, the dissociation between surprise and updating lies at the core of our experimental designs and associated models. We are aware of only three previous studies that have differentiated surprise and updating. The first one was conducted by O'Reilly and colleagues (2013) using a spatial saccadic task. Participants were signaled whether target violating prior expectations were informative or not to predict future target locations. This gave rise to two types of trials violating participants' prior beliefs: updating trials and surprise-only trials. The authors found that ACC was involved in belief updating, while superior parietal lobule responded to the immediate consequences of violation of expectations (i.e., reprogramming actions). In the second study, Schwartenbeck and colleagues (2016) dissociated between surprise and updating to characterize the role of dopamine signaling in response to unexpected events. To the aim, they implemented a task in which participants had to infer which one of two simultaneously presented cue modalities (visual or auditory) predicted a monetary outcome. Participants were instructed that the predictive/non-predictive status of the two modalities did not change on a trial-by-trial basis but periodically. For each modality, there were one bad and one good tone/shape, which predicted, respectively, monetary loss and win with a cue validity rate of 90%. Importantly, half of the trials were useful for inference (one modality predicted a win while the other a loss), while the other half were uninformative (both modality predicted win or loss). In this way, they dissociated between surprise events (e.g., the 10% invalid trials in the uninformative trials), and updating that could occur only in informative trials. Differently from O'Reilly and colleagues (2013), no motor response was required at cue onset, and the surprise/update values of the stimuli were not explicitly signaled. The authors found that updating involved dopamine-rich midbrain regions along with bilateral

inferior frontal cortex, bilateral posterior parietal cortex, and ACC, while surprise modulated activity in pre-supplementary motor area (pre-SMA) and dorsal anterior cingulate cortex (dACC).

The last study attempting to decorrelate surprise and updating was conducted by Kobayashi and Hsu (2017). They implemented a version of the Ellsberg three-color urn problem in which participants exactly knew the total number of balls in an urn and the number of balls of one color (called “risky color”), but they did not know the number of balls of the other two colors (called “ambiguous colors”). At the end of each trial, participants received $ 10 if a resolution draw from the urn matched a winning color presented at the beginning of the trial. Before the resolution draw, participants viewed a ball drawn from the urn and, then, returned to the urn. This task allowed distinguishing not only update from surprise, but further differentiating belief update about the urn composition from value update about the chance of winning. Concerning belief update, only the draw of an ambiguous-color ball was informative because seeing a risky-ball was redundant since their number was already known. Concerning value updating, it should occur only when the specified winning color was ambiguous (“ambiguous gamble”), because in case the winning color was the risky one (“risky gamble”), participants already knew the chance of winning. Summarizing then, belief update could occur only after ambiguous-ball draws, while value update only after ambiguous gambles. Surprise was associated with every draw, since each color, including the risky color, had its level of expectancy violation. The authors found that belief updating modulated activity in bilateral middle frontal gyrus, bilateral inferior parietal sulcus (IPS) and precuneus. Value updating was associated with activity in right ventromedial prefrontal cortex, anterior and middle cingulate, and left superior temporal gyrus. Surprise was associated with activity in bilateral anterior insula.

Overall, these three studies found dissociations between belief updating and surprise highlighting the importance of tasks that allow decorrelating these types of information in order to better characterize the associated processes. However, despite some broad commonalities, there were differences in the precise localization of updating and surprise, likely due to difference in the tasks and in the required

processes. As suggested by Kobayashi and colleagues (2017), given the low number of studies attempting to dissociate between surprise and update, we need more studies in various tasks and domains to assess the existence of domain-general and domain- specific correlates of belief updating.

Concerning the EEG literature on Bayesian belief updating, the majority of the studies have focused on the P3 event-related potential (ERP) component (Kolossa et al., 2013; Kolossa, Kopp, & Fingscheidt, 2015; Kopp, 2008; Mars et al., 2008). The attention to this component was likely driven by its amplitude sensitivity to stimulus probabilities. According to the influential context-updating theory (Donchin & Coles, 1988), indeed, the P3 (a parietally-distributed positive deflection usually emerging around 250-500; for an overview, see Polich, 2003) is an index of the revision of an internal model in order to maintaining "its mapping of probabilities" (p. 367) accurate. Despite the clear similarity between former interpretations of P3 and Bayesian inference (Kopp, 2008), a Bayesian approach to the study of this component is relatively recent. Using an ideal Bayesian observer, Mars and colleagues (2008) modeled beliefs about stimulus occurrence in a choice RT task in which the relative frequency of four stimuli was manipulated between blocks. The authors found that trial-by-trial fluctuations in P3 amplitude could be explained by surprise conveyed by the stimuli. A similar result was obtained by Kolossa and colleagues (Kolossa et al., 2013) in a two-choice RT task, in which fluctuations in P3 amplitude were well explained by a Bayesian observer that updated beliefs with some memory constraints (this aspect will be discussed in details in Chapter 4) and alternation expectancies (Squires, Wickens, Squires, & Donchin, 1976).

Further studies have investigated whether different P3 subcomponents might be dissociated in terms of updating and surprise. To this aim, Kolossa and colleagues (2015) implemented a special urn-ball task that allowed manipulating probabilities at two levels. At the beginning of each trial participants were presented with a tableau containing ten urns of two different types, each of which containing ten balls of two different colors. The two probabilistic manipulations involved the proportion of urn types and the proportion of ball colors within each urn type. They referred to these

two proportions as prior probability (urn type) and likelihood (ball color). After the tableau presentation, four balls were sequentially drawn with replacement from one randomly selected urn. Afterwards, participants were required to infer which type of urn had been selected. This paradigm allowed the authors to distinguish updating of beliefs about "hidden state" (beliefs about which type of urn was being sampled from) from updating of beliefs about future observations (beliefs about which ball would have been drawn). Results showed that three subcomponents of the "late positive complex" (Sutton & Ruchkin, 1984), namely P3a, P3b and Slow Wave (SW), were differently influenced by updating and surprise. First, they confirmed previous findings about the modulation of the P3b amplitude (also referred to as P3) by surprise. Updating of beliefs about hidden states was the best predictor of the P3a amplitude, a component with a more frontocentral topography than the P3b and with earlier peak latency. Last, updating of beliefs about future observations was the best predictor of the SW activity emerging after the P3. The association between P3a amplitude and belief updating has been supported by Bennett and colleagues (2015) in a perceptual learning task. It is important to highlight here that despite all the EEG studies described so far have investigated surprise and updating, these two were not explicitly decoupled in their respective tasks.

This brief review of the literature on Bayesian belief updating demonstrates the great value of Bayesian models to gain direct insights into the cognitive processes and neural mechanisms underlying different functions. The present dissertation aims at joining and further extending this previous work by exploring a pivotal dimension of our life, that is, time. As already mentioned, research on temporal preparation has investigated how the brain tracks the temporal hazard of target onset starting from a prior foreperiod distribution. However, how prior distributions are formed and updated is still an unsettled question. In the last paragraph of the Introduction, we will describe how we applied a Bayesian approach to investigate updating and surprise associated with temporal expectations.

1.3 Project overview

The **overarching goal** of the present PhD project was to pinpoint the neural mechanisms by which beliefs about event timing are dynamically updated. Moreover, as mentioned above, although updating usually occurs in the presence of surprising events, processes involved in belief updating are probably different from those responding to mere violations of expectations (Kobayashi & Hsu, 2017; O'Reilly et al., 2013; Schwartenbeck et al., 2016). Therefore, our second and interrelated research goal was to understand whether, and to which extent, it would be possible to dissociate the neural mechanisms specifically involved in updating from those dealing with surprising events that do not require an update of internal models. To accomplish our research goals, we capitalize on both state-of-art methodologies (i.e., functional magnetic resonance imaging and electrophysiology) and Bayesian computational modeling.

The present dissertation is composed of an fMRI and two EEG studies that constitute our three main chapters.

In the fMRI study, we implemented a temporal preparation task that was devised following the spatial saccadic planning task by O'Reilly and colleagues (2013). Briefly, in their task participants had to make speeded saccades to visual colored targets that appeared at different locations on a circular perimeter. Participants could predict target locations since most of them appeared at an angle, α, drawn from a Gaussian distribution whose mean and standard deviation were kept constant during each block of trials, but abruptly changed between blocks. Blocks were not temporally separated, but participants were explicitly instructed that a change in the target color signaled the beginning of a new block. On few trials (interspersed with the other trials), the target appeared at a random location. Importantly, these "one-off" targets were always grey, signaling to the participants that the current trial was not the start of a new block (update trial), that is, no update had to be done. In sum, although update and one-off trials were both surprising, update and surprise were dissociated using the color manipulation.

In our temporal preparation task, we also differentiated between update and surprise trials by using the same color manipulation as in O'Reilly and colleagues (2013) and by varying foreperiod durations instead of target location. Hoping that the reader will forgive us for the following spoiler, the results of the fMRI study confirmed the validity of our manipulation and showed that two cognitive-control networks (Dosenbach, Fair, Cohen, Schlaggar, & Petersen, 2008) were differentially involved in updating of temporal expectations and in dealing with surprising events.

Having successfully dissociated between updating and surprise in our first fMRI study, the next step was to investigate the temporal dynamics of the two processes. Do updating and surprise act on similar or distinct processing stage(s)? What is the time course associated with the two? To address these questions, in Study 2 participants performed our adapted foreperiod task (basically the same task as in Study 1) while their electroencephalographic activity was recorded. We dissociated surprise and updating at the P3 level, but interestingly, we found modulations also at early processing stages.

While in Studies 1 and 2, the color manipulation explicitly encouraged participants to form and update temporal expectations about the foreperiod duration, in Study 3 we wanted to make a step further trying to answer the following question: how updating of temporal expectations is accomplished when Bayesian inference is implicitly rather than explicitly driven? To this aim, we took away the color manipulation from our version of the foreperiod task. In doing so, we aimed to investigate whether EEG signatures would differ or not when changes in the underlying distributions are explicitly signalled or not. Results from Study 3 showed both similarities and differences between implicitly- and explicitly-driven temporal inferences, further confirming the role of the P3 in belief updating.

Chapter 2

Temporal belief updating and surprise modulate cognitive control networks

2.1 Introduction

The ability to predict when a relevant event might occur is critical to survive in our dynamic and uncertain environment. This cognitive ability, usually referred to as temporal preparation, allows us to prepare fast and accurate responses to forthcoming stimuli by anticipating their likely timing of occurrence: from safely crossing a busy road during rush hours, to timing turn taking in a conversation, to catching something in mid-air, are all examples of how important and ubiquitous temporal preparation is in our everyday life (e.g., Correa, 2010; Coull & Nobre, 1998; Nobre et al., 2007).

Temporal preparation has been traditionally investigated in simple and choice response time (RT) tasks in which a variable time interval (i.e., foreperiod) separates warning and target stimuli (for reviews, see Niemi & Näätänen, 1981; Vallesi, 2010). Previous research has shown that temporal preparation can be modeled by the hazard function, which refers to the conditional probability that an

event will occur, given it has not yet occurred (Bueti et al., 2010; Janssen & Shadlen, 2005). This implies that an observer may build up temporal expectations according to both elapsed time and distribution of possible foreperiods. In a very influential study by Janssen and Shadlen (2005), monkeys were trained to anticipate the timing of a "go" signal preceded by a warning signal. Within a block of trials, the foreperiod was drawn from either a unimodal or a bimodal distribution. Single-cell recording showed that the firing rate of neurons in the intraparietal area reflected the hazard function associated with the valid distribution. Such a finding was then corroborated in humans in both functional magnetic resonance imaging (fMRI) and electroencephalographic (EEG) studies (for fMRI: Bueti et al., 2010; for EEG: Herbst et al., 2018; Trillenberg, Verleger, Wascher, Wauschkuhn, & Wessel, 2000). However, while these studies demonstrated the use of an internal predictive model (i.e., the underlying foreperiod distribution), the question of how such a model is learnt is still unsettled. Here, we sought to characterize the neural mechanisms that allow forming and updating temporal expectations. Specifically, we applied the Bayesian brain framework to quantitatively describe belief updating about foreperiod distributions.

According to the Bayesian brain hypothesis (Doya, Ishii, Pouget, & Rao, 2007; Friston, 2005; Kersten, Mamassian, & Yuille, 2004; Knill & Pouget, 2004), the brain weighs current evidence (*likelihood*) on the basis of expectations about the environment (*prior* beliefs) and updates such beliefs into *posterior* ones. Given an agent's beliefs, those events fulfilling our prior expectations can be predicted to optimize behavior. Conversely, those events violating our expectations are surprising, thus leading to behavioral costs and to an update of the internal predictive models. It is important to note that "a surprising observation is not necessarily associated with improving an agent's beliefs about the environment" (Schwartenbeck et al., 2016). A classic example of this concept is provided by the "white noise" paradox (Barto et al., 2013; Itti & Baldi, 2005). Imagine yourself watching a "snow" television screen. Each frame is very surprising given the high number of possible random combinations of pixel patterns, each of whom is associated with a low probability of occurrence. Notwithstanding this, surprise does

not lead to a relevant change in the agent's internal model since a random noise will end up being increasingly more expected. The white noise paradox represents an extreme case as, normally, a surprising event does cause a revision of the agent's beliefs. Two information theory measures have been used to disentangle surprise and updating. The surprise associated with a particular event is formalized in terms of Shannon's information (I_S), as follows (Itti & Baldi, 2005; Shannon, 1948):

$$I_S(o) = -\log p(o|prior), \tag{2.1}$$

that is, the negative log probability of the observation, o, given the prior expectations, prior. According to this formula, an event that is highly unlikely elicits high surprise when it occurs.

The updating of the internal model is, instead, formalized in terms of the Kullback-Leibler divergence (D_{KL}) from prior to posterior beliefs (Baldi & Itti, 2010; Itti & Baldi, 2009):

$$D_{KL}(post||prior) = \sum_{o} \log\left(\frac{o|prior}{o|post}\right) p(o|prior). \tag{2.2}$$

Although surprise and updating are likely to co-occur (i.e., they are correlated), they may reflect distinct processes that might be dissociated. To this aim, O'Reilly and colleagues (2013) developed a saccadic planning task in which target locations could be predicted by inferring the underlying spatial distribution. Crucially, participants were explicitly informed whether information from surprising trials was useful to infer future target locations (updating trials) or whether it was not (one-off trials). This manipulation allowed decomposing surprise (I_S) and updating (D_{KL}) of spatial locations at two levels of cognitive operations: 1) between-trial processes associated with the updating of an internal predictive model to

predict future events, and 2) within-trial processes involved in facing unexpected events (e.g. reprogramming a response to an unpredicted stimulus). Using this paradigm, the authors managed to observe separated brain regions for surprise and updating, namely, posterior parietal and anterior cingulate cortex, respectively.

In the present study, we combined computational modeling and fMRI to investigate the neural mechanisms associated with updating of temporal expectations and the effect of temporally unexpected, surprising, events. Despite temporal preparation is a fundamental feature of cognitive brain functions, to the best of our knowledge, the present study represents the first attempt to investigate the neural mechanisms underlying the optimization of prior temporal expectations. To achieve our goal, we adapted the spatial paradigm by O'Reilly and colleagues (2013) to a foreperiod temporal preparation task. In order to quantify surprise and updating, we used an ideal Bayesian observer that enabled to capture participants' beliefs in terms of probability distributions and to model belief updating using the Bayes' rule. Surprise and updating were employed as parametric explanatory variables of both whole-brain fMRI and functional connectivity analyses to elucidate how Bayesian inference about temporal expectations is implemented in the brain.

2.2 Methods

2.2.1 Participants

The study included an initial sample of 26 participants. Data from two participants were discarded because of excessive head movements (see details on paragraph 2.2.6). Additionally, one participant was excluded due to falling asleep (11% of no response) and another one due to low compliance with task instructions (49% of overall accuracy; the participant reported that he changed his strategy during the session, but this fact led to a lot of anticipations). Therefore, the final sample comprised 22 participants [10 females; mean age: 26 (SD=4), range: 20-34 years old]. All of them were right-handed, as assessed with the Edinburgh Handedness Inventory (Oldfield, 1971) with an average score of 89.1 (SD=11.7,

range: 60-100), reported no history of neurological or psychiatric disorders, normal color vision and normal or corrected-to-normal visual acuity (MRI-compatible glasses were used when appropriate). The procedures involved in this study were approved by the Bioethical Committee of the Azienda Ospedaliera di Padova. Participants gave their written informed consent before the experiment, in accordance with the Declaration of Helsinki, and they were reimbursed 25 euros for their time.

2.2.2 Task and procedure

The foreperiod task was implemented in MATLAB (The MathWorks, Inc., Natick, Massachusetts, United States) using the PSYCHOPHYSICS TOOLBOX 3 (Brainard & Vision, 1997; Kleiner et al., 2007; Pelli, 1997). As mentioned in the Introduction, we modeled our temporal task after the spatial one developed by O'Reilly and colleagues (2013). Each trial began with the presentation of an uninformative warning signal that consisted of a black fixation cross. The warning signal was displayed centrally against a gray background and remained on the screen for the whole duration of the foreperiod for that trial. After the foreperiod elapsed, a target appeared and participants were instructed to respond to the onset of the target as quickly as possible. The target was a colored circle with a diameter equal to the length of the cross arms, centrally presented for 1500 ms (Fig. 2.1). Participants responded by pressing a button of an MRI-compatible response box with their index finger. Half of the participants used their right hand and the other half their left hand. The inter-trial interval was a blank screen that was presented for a total duration drawn from a Poisson distribution having a lambda of 2 and shifted of 2 sec (i.e., from 2 to 12 sec).

In 80% of trials, called "normal trials", the foreperiod duration (FP) was drawn from a Gaussian distribution with a mean and standard deviation that remained fixed during a block, but that abruptly changed across blocks. Blocks were not temporally separated, such that the first trial of a new block followed directly the last trial of the previous block. In the remaining 20% trials, called "uniform trials", the foreperiod duration was drawn from a uniform distribution with

boundaries 200 and 3000 ms (Fig. 2.1B). Consequently, the generative probability density function over foreperiod duration was:

$$p(FP) = .80\ p\big(FP \big| FP \sim \mathcal{N}(\mu, \sigma)\big) + .20\ p\big(FP \big| FP \sim \mathcal{U}(200\ ms, 3000\ ms)\big) \qquad (2.3)$$

Importantly, participants were instructed to use the color of the target in order to distinguish the beginning of a new block from uniform trials. More specifically, in normal trials each target could be filled with one of four colors (vermillion, reddish purple, bluish green, and blue). A given color (i.e., blue) was kept constant over a block of trials and changed only when a new block started. Differently from normal trials, in uniform trials the target color was always white. Every time a white circle was encountered, participants were instructed to respond to it but avoid using the current temporal information (either earlier or later than the actual foreperiod distribution) to anticipate the next target occurrence. Rather, a change in color signaled that the current trial was the beginning of a new block and, thus, participants could use information from the last foreperiod to strongly update their expectation in order to predict subsequent target onsets. Summarizing, the color manipulation allowed creating essentially three types of trials: “update” trials in which the unpredicted FP led to strongly updating the internal model in order to anticipate the next target onset, “predictable” trials in which the target onset could be easily predicted using information from previous trials in the current block, and “uniform” trials that, even if they were breaking current temporal expectation, did not require any updating of foreperiod distribution.

The experiment was composed of 33 blocks with a total number of trials equal to 350 (the number of trials in a block was in the range 7-13, mean = 9.82, sd = 1.24). For each new block, the mean of the Gaussian distribution from which the normal foreperiods were extracted was at least 3 standard deviations far away from the previous block. Moreover, 16 blocks had a mean lower than the previous block, and 16 blocks a mean higher than the previous one. Overall, the 32 Gaussian distribution (not considering the first block) were derived from an orthogonal

combination of 7 means (500, 800, 1100, 1400, 1700, 2000 and 2300 ms) and 4 standard deviations (20, 40, 60 and 80 ms). In total there were 6 fMRI runs. Participants were informed that each new run started using the same foreperiod distribution as that used on the previous block (at the beginning of a new run, the number of trials before the new block was in the range 4-6, mean = 5.2, sd = 0.8).

Before the fMRI session, participants practiced the task outside the MRI bore. They performed a shorter version of the experimental task comprising four blocks. In the first two blocks, they were presented with normal foreperiods only in order to familiarize themselves with the arbitrary association between colors and foreperiod distributions. In the subsequent blocks, we introduced uniform foreperiods and carefully explained the difference between them and the normal ones.

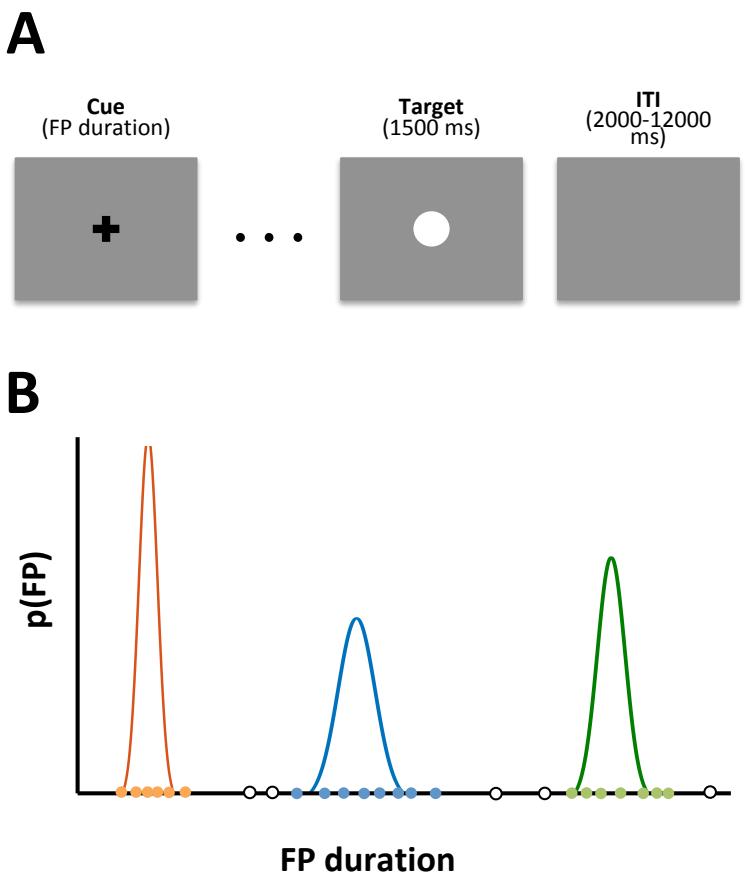

Figure 2.1 | (A) Example of a trial. On each trial participants began observing a neutral cue, which was displayed for a variable foreperiod (FP) duration. At the end of the FP a colored dot appeared and was displayed for 1500 ms. Participants were required to respond as soon as possible to the onset of the target. **(B)** Distribution of FP duration in three blocks. Predictable trials were characterized by FP durations extracted from a normal distribution (color dots indicate onset of targets after normal foreperiods) whose mean and standard deviation were kept constant within a block. Uniform trials were characterized by FP extracted from a uniform distribution (white dots).

2.2.3 Normative Bayesian learner

On a trial-by-trial basis, we modeled participants' expectations for target onset by assuming they updated their beliefs as an ideal Bayesian observer. This model (adapted from: O'Reilly et al., 2013) aims to iteratively infer the parameters μ and σ of the Gaussian distribution underlying normal foreperiods. After each new observation, it estimates the posterior probability for each possible pair of parameters μ and σ (i.e. the posterior probability over parameter space) as described below. Following the instructions given to participants, no updating occurred after uniform trials, so the posterior probability over parameter space at trial *n* was:

$$p(FP{\sim}\mathcal{N}(\mu,\sigma)|FP_{1:n}) = p(FP{\sim}\mathcal{N}(\mu,\sigma)|FP_{1:n-1}). \quad (2.4)$$

After update and predictable trials, the parameter space probability was updated using the Bayes' rule:

$$p(FP{\sim}\mathcal{N}(\mu,\sigma)|FP_{1:n}) \propto p\big(FP_n\big|FP{\sim}\mathcal{N}(\mu,\sigma)\big){\times}p(FP{\sim}\mathcal{N}(\mu,\sigma)|FP_{1:n-1},\varphi), \quad (2.5)$$

where the variable φ indicates that the type of trial determined the prior used. In predictable trials, the prior on trial *n* was the posterior obtained on trial *n-1*. However, the change of color in update trials explicitly signaled the start of a new distribution and, as a consequence, previous observations were no more useful in estimating the posterior probability. For this reason, the prior in update trials was a uniform distribution computed as:

$$p(FP{\sim}\mathcal{N}(\mu,\sigma)|FP_{1:n-1}) = {}^{1}\!/_{300\times15}\,, \quad (2.6)$$

where 300×15 indicates the size of the employed parameter space (i.e. the combination of all the means from 10 to 3000 ms and standard deviations from 10 to 150 ms in steps of 10 ms).

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

$$p(FP_{n+1}|FP_{1:n}) = p(predictable_{nb+1}) \sum_{\mu_{n+1},\sigma_{n+1}} \left(FP_{n+1}\middle|FP_{n+1}\sim\mathcal{N}(\mu_{n+1},\sigma_{n+1})\right) \times p\left(FP_{n+1}\sim\mathcal{N}(\mu_{n+1},\sigma_{n+1})\middle|FP_{1:n}\right) + p(uniform_{nb+1} \cup update_{nb+1})\, p(FP_{n+1}|\mathcal{U}(10\ ms, 3000\ ms)), \quad (2.7)$$

where $p(predictable_{nb+1})$ and $p(uniform_{nb+1} \cup update_{nb+1})$ 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). For simplicity, the combined probability to have an uniform or an update trial was set to the true proportion of those trial types at a given trial within a block, smoothed using a moving average in order to have a monotonic increase in the probability of having an update trial (e.g., the probability of having an update/uniform trial on tb+1=14 was higher than on tb+1=13 and so on). The probability of a predictable trial was equal to $1-p(uniform_{nb+1} \cup update_{nb+1})$. The output of the model is presented in Fig. 2.2A.

2.2.4 Model-based measures of updating and surprise

As mentioned in the Introduction, two measures from information theory were used to formally distinguish between the updating of temporal expectations and the surprise of observing the target after a specific foreperiod. These measures were computed with reference to our model as follows:

Updating. Following Itti and Baldi (2009), we quantified the updating of the internal predictive model as the Kullback-Leibler divergence (D_{KL}; Fig. 2.2.B) between prior and posterior on trial *n*:

$$D_{KL}(FP_n) = \sum_{FP} p(FP_n|FP_{1:n-1}) \log \frac{p(FP_n|FP_{1:n-1})}{p(FP_{n+1}|FP_{1:n})}; \quad (2.8)$$

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

$$I_S(FP_n) = -\log h(FP_n), \quad (2.9)$$

where $h(FP_n)$ is the probability of FP of the prior on trial *n* conditional on the elapsed time (Fig. 2.2.C):

$$h(FP_n) = \frac{f(FP_n)}{1 - F(FP_n)}, \quad (2.10)$$

where $f(FP_n)$ is the prior probability $p(FP_n|FP_{1:n-1})$, and $F(FP_n)$ is the cumulative probability $\int_o^{FP_n} f_{FP}(ms)\, dms$.

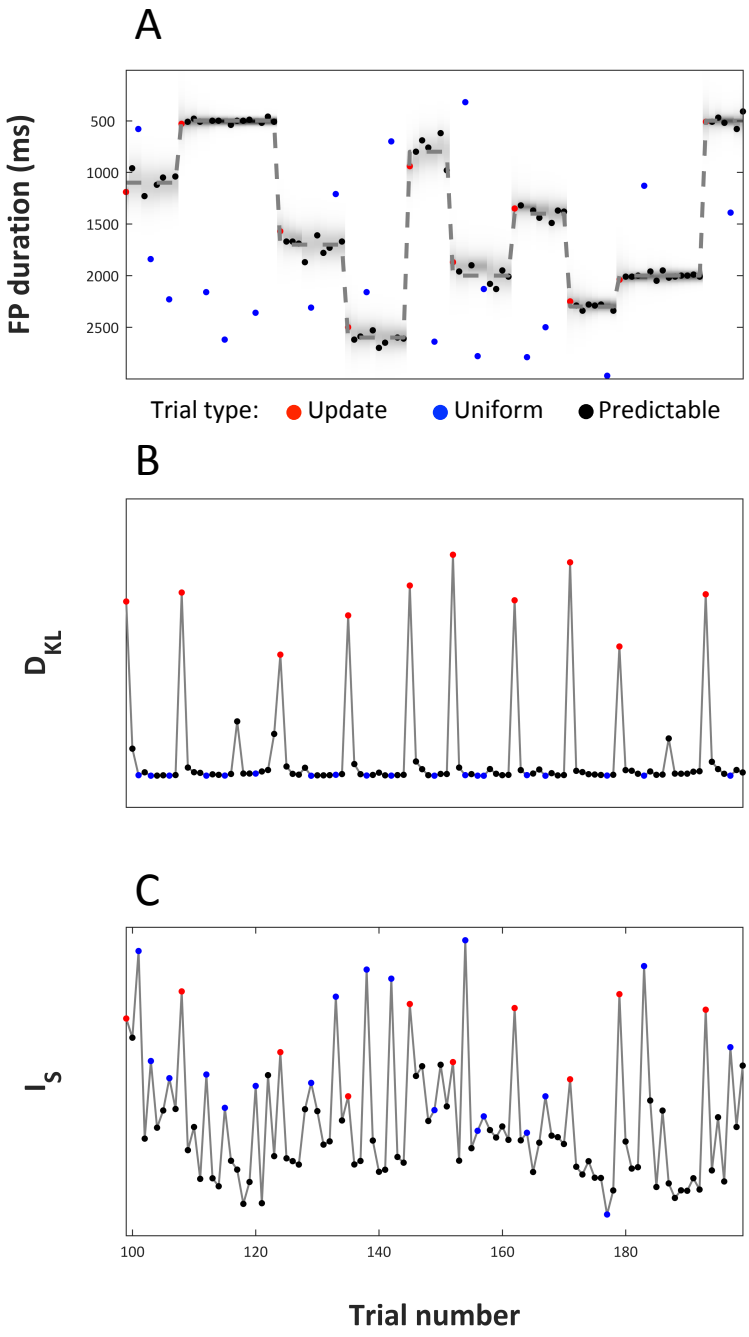

Figure 2.2 | Bayesian learner and model-based regressors. All panels show the data from 100 trials. Dot colors indicate trial types as reported in the legend. **(A)** Plot of the state of the normative Bayesian learner. On y axis is FP duration. The dashed line indicates the mean of the generative Gaussian distribution from which update and predictable foreperiods 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 regressors for updating (D_{KL}) and surprise (I_S).

2.2.5 Behavioral data analysis

Data from error trials (with anticipated or without responses) and post-error trials were excluded from analysis. Reaction times (RTs) were log-transformed to mitigate the influence of non-normal distribution and skewed data. Following the procedure proposed by Baayen and Milin (2010), log-transformed RTs were analyzed by conducting Linear Mixed Models (LMM), using the *lme4* library (Bates, Mächler, Bolker, & Walker, 2014) in R (R Core Team, 2015). In our main analysis, we investigated the behavioral correlates of surprise and updating by using I_S and D_{KL} as regressors of interest. A full LMM was specified as follows: I_S and D_{KL} (and their interaction) as well as TRIAL, that represents the rank-order of a trial, and log-RT at the preceding trial (PRECEDING RT), were entered as fixed-effects predictors. The random structure include correlated by-subject random intercepts and by-subject random slopes for TRIAL, PRECEDING RT, I_S and D_{KL}. All these continuous predictors were standardized using Z-score in order to have the same scale, which allows comparing them statistically. The variables TRIAL and PRECEDING RT were included to control for the temporal dependencies that usually occur between successive trials (Baayen & Milin, 2010). Specifically, TRIAL was included to capture possible effects of learning and fatigue, while PRECEDING RT was used to take into account the RT autocorrelation between subsequent trials. Using the function *step* from the *lmerTest* library (Kuznetsova, Brockhoff, & Christensen, 2015), a stepwise variable selection was performed starting from the full model to eliminate non-significant effects from the full LMM.

2.2.6 fMRI data analysis

Data acquisition. Structural and functional images were acquired using a 3T Ingenia Philips whole body scanner (Philips Medical Systems, Best, The Netherlands) equipped with a 32-channel head-coil, at the Neuroradiology Unit of the University Hospital of Padova, Italy. Functional data were obtained using a whole head T2-weighted echo-planar image (EPI) sequences (repetition time, TR: 2000 ms; echo time, TE: 30 ms; 39 axial slices with ascending acquisition; voxel size: 3 × 3 × 3 mm;

flip angle, FA: 76°; field of view, acquisition matrix: 84 × 84). Excluding the four dummy scans for stabilization of the T1-saturation effect, the functional acquisitions consisted of 8 minutes of resting state activity, which will not be discussed in this thesis, followed by a total of 39.4 minutes of task related activity. To correct fMRI images for spatial distortion, at the beginning of each of the six runs, two spin echo EPI images with reversed phase encoding directions were acquired. These images are geometrically matched (same field of view and voxel size) with the functional images (Glasser et al., 2013). After functional session, high resolution T1- and T2-weighted anatomical images (T1w: TR/TE: 8.10/3.72 ms; 180 sagittal slices; FA: 8°; voxel size: 1 × 1 × 1 mm; acquisition matrix: 256 × 256; T2w: TR/TE: 2500/249 ms; 180 sagittal slices; FA: 90°; voxel size: 0.97 × 0.97 × 1 mm; acquisition matrix: 256 × 256) were collected. In order to avoid head movement during scanning, small foam cushions and sponge pads were placed around the participant's head. Subjects also wore earplugs to reduce acoustic noise.

MRI preprocessing. First, spatial distortion of functional data were corrected using the susceptibility-induced off-resonance field estimated from the two oppositely phase-encoded spin echo EPI images as implemented in the FSL (FMRIB Software Library, version 5.0.7) toolbox "topup" (Andersson, Skare, & Ashburner, 2003; S. M. Smith et al., 2004). This step improves the following coregistration step between fMRI and structural image (Glasser et al., 2013). Functional data were then slice-timing corrected using the middle slice as the reference frame, rigidly realigned to the first volume and spatially smoothed using a Gaussian kernel with a full-width at half-maximum (FWHM) of 5 mm using SPM12 (Statistical Parametric Mapping software; Wellcome Department of Cognitive Neurology, London, UK; http://www.fil.ion.ucl.ac.uk/spm). Participant's head movements were quantified by means of framewise displacement (FD) index which represents the sum of the absolute values of the derivatives of the translational and rotational realignment parameters (Power, Barnes, Snyder, Schlaggar, & Petersen, 2012). Subjects with mean FD above two standard deviations from the mean of all subjects (group mean = 0.09 mm, standard deviation = 0.02 mm) were excluded. The deformation field that mapped the individual functional data to standard Montreal Neurological

Institute (MNI) template was derived combining several steps, all implemented with FSL (Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012; S. M. Smith et al., 2004). Usually a typical workflow involves the coregistration of the functional image to the T1-weighted anatomical image and the warp of the structural image to a template. Here, a T2-weighted anatomical image was used as an intermediate target since it has the same acquisition modality of fMRI data, but the same high-resolution with clear region boundary contours of T1-weighted anatomical images. First, T1-weighted anatomical image was bias-field corrected and a non-linear transformation to MNI template was estimated (T1>MNI). Both T2- and T1- weighted structural images were skull-stripped and then a 6-parameter transformation from the former to the latter was computed (T2>T1). At the end, a 12-parameter affine transformation from the first volume of the functional data to the T2-weighted skull-stripped anatomical image was estimated (fMRI>T2) and combined with the T2>T1 and T1>MNI transformations. The resulting transformation was then used to map the results obtained at individual level in the functional space to the MNI template.

Whole-brain analysis. Statistical analyses were carried out using SPM12 (Ashburner et al., 2014) to identify the volumes of interest (VOIs) for the functional connectivity (FC) analysis. For each participant, first-level analysis was performed into the subject space (i.e. not normalized) using two general linear models (GLMs). For each GLM, the task was modeled with three regressors that were the main effect of the FP, the main effect of target onset, and either D_{KL} or I_S, estimates of head movements were also included as six additional regressors of no interest. Slow signal drifts were removed using a 128 s high-pass filter. The main effect of the foreperiod was model as a boxcar starting from the cue onset and with duration equal to the FP length. The main effect of target onset was modeled as a delta function at the target onset modulated by the model-based regressor, D_{KL} or $I_{S,}$ respectively. All these regressors were convolved with the hemodynamic response function. As in the behavioral analysis, these two parametric modulators (PMs) were standardized using Z-score and orthogonalized with regard to target onset. The decision of running two GLMs instead of a single GLM including both PMs was made to keep variance that these two PMs likely share. Since this analysis was preliminary to FC, keeping this

shared variance is important to obtain a more exhaustive picture of the networks involved in processing and differentiating surprising information [1]. For each participant and each GLM, a t-contrast was computed for each PM versus zero (i.e. baseline). At the group level, individual participants' Z-statistic maps were normalized to MNI template as described in the previous section. Then, for each GLM, group-level maps were generated with random-effect models using participants' contrast maps. Group statistics were assessed for cluster-wise significance using a cluster-defining threshold of $p < .001$ and a cluster significance threshold of $p < .05$ (FWE-corrected). Furthermore, a third GLM with the hazard rate of the target onset (i.e. $h(FP_{onset})$) as PM was run in order to replicate previous findings on hazard rate (Bueti et al., 2010).

Functional connectivity analysis. Task-related functional connectivity analysis was computed using the correlational psychophysiological interaction (cPPI) toolbox (Fornito, Harrison, Zalesky, & Simons, 2012). In classical PPI analyses (Friston et al., 1997), the activity time course from a specific seed region is extracted and multiplied by a task regressor of interest to isolate task-specific modulations in the functional coupling between the seed region and other brain regions. This approach is regression-based, in the sense that for each pair of time series the seed region activity is used as a predictor of the activity in the other regions. This implies that PPI is a form of effective connectivity (Friston et al., 1997), thus, it is suitable when there are clear hypotheses about which region may modulate activity in other regions (Fornito et al., 2012). Since we had no such specific predictions, we employed the cPPI approach, which provides a measure of functional connectivity that does not require directional assumptions. Briefly, for any given pair of brain regions their time course is multiplied by the task-regressor to obtain two PPI terms. Then, the partial correlation between the two PPI terms is estimated while controlling for possible confounds (e.g., task-unrelated connectivity, noise). In a nutshell, starting from a set of regions the cPPI analysis returns a functional connectivity matrix of pair-wise covariations in task-specific modulations of neural activity. We estimated two cPPI correlation matrices, one for each of the two model-based regressors, D_{KL} and I_S. For

[1] The orthogonalization of the regressors gave similar results.

each cPPI analysis, volumes of interest (VOI) were defined by generating 6-mm-radius spheres centered on the maximum peak of each significant cluster observed in the whole-brain analysis. Then, the time course from each VOI in each subject was extracted using the *spm_regions* function in SPM. For each matrix, the correlation between pairs of VOIs was estimated after partialing out the coviariance with the remaining VOIs to ensure that each correlation is specific to each pair. P-values for the two cPPI correlation matrices were adjusted for multiple-testing by controlling the false discovery rate (FDR) at the .05 level.

2.3 Results

2.3.1 Behavioral results

Backward elimination of non-significant effects resulted in a model specified as the following *lme4*-notation formula:

$$\log(RT) \sim \text{TRIAL} + \text{PRECEDING RT} + I_S * D_{KL} + (\text{TRIAL} + \text{PRECEDING RT} + I_S \mid ID). \quad (2.11)$$

Visual inspection of the residuals showed that the model was a bit stressed. As suggested by Baayen and Milin (2010), trials with absolute standardized residuals higher than 2.5 standard deviations were considered outliers and removed (2.4% of the trials). After removing outlier trials, the model was refitted and this time it achieved reasonable closeness to normality. Table 2.1 shows the statistical results of the type III ANOVA. A significant interaction was found between I_S and D_{KL}. Fig. 2.3 shows that RTs increased with increasing surprise (I_S) and this effect was augmented with increasing D_{KL}.

Table 2.1 | Analysis of variance with type-III sums of squares.

Fixed Effects	Sum Sq	Num. *df*	Den. *df*	*F*	*p*	β
TRIAL	0.39	1	21.0	10.0	.005	0.07
PRECEDING RT	0.66	1	21.0	17.0	< .001	0.10
I_S	3.70	1	22.1	94.8	< .001	0.25
D_{KL}	0.33	1	6819.8	8.6	.003	-0.05
I_S :D_{KL}	0.99	1	6823.5	25.5	<. 001	0.09

Notes: F-statistics and associated p-values were calculated using Kenward-Roger's approximation of degrees of freedom (Halekoh & Højsgaard, 2014). Additionally, standardized regression coefficients (β) are shown.

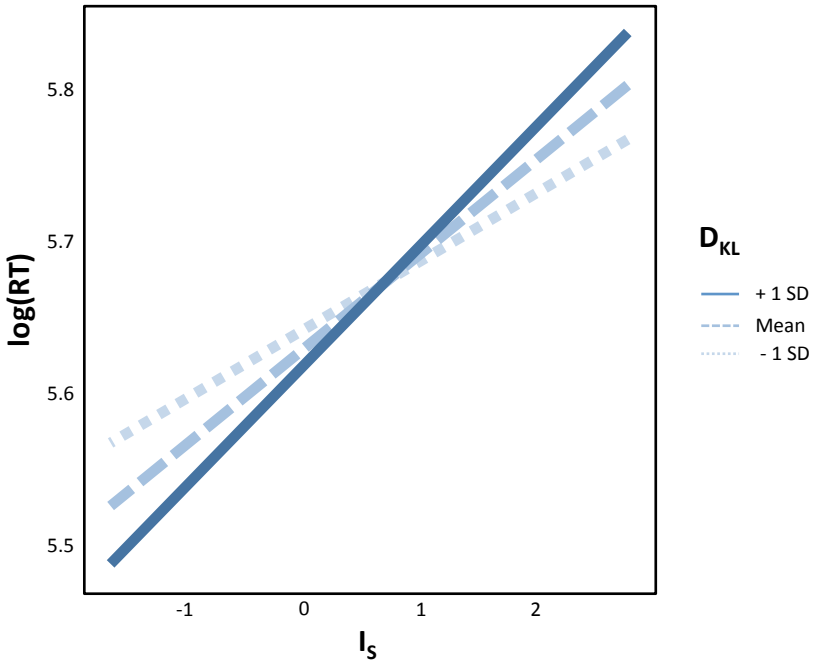

Figure 2.3. | Interaction plot for log-transformed RTs. The plot shows the effect of surprise (I_S) at three levels of updating (D_{KL}), high (+1SD), medium (Mean) and low (-1SD). The plot was done using the R package *jtools*.

2.3.2 Whole-brain fMRI results

All whole-brain fMRI results are reported in Table 2.2. Updating (D_{KL}) significantly modulated activity in a set of lateral frontal and parietal regions, as well as in medial parietal regions, along with a cluster around the left fusiform gyrus (lFFG) and the left cerebellum (Fig. 2.4A). Lateral frontal and parietal regions

included bilateral inferior frontal gyri (IFGs) and posterior parietal cortices (PPCs) around the intra-parietal sulci (IPSs). Medial parietal activations included the posterior cingulate cortex (PCC) and the precuneus.

Surprise (I_S) significantly modulated activity in the right IFG, in bilateral PPCs around the IPSs, in the dorsal anterior cingulate cortex (dACC) including the pre-supplementary motor area (pre-SMA), in bilateral anterior insula (aINS), and in the lFFG (Fig. 2.4B).

Results from cPPI analysis showed that these set of brain regions formed two task-related large-scale networks. Fig. 2.5 shows, for each network, functional connection strengths between nodes, which were significantly modulated by D_{KL} and $I_{S,}$ respectively.

Concerning the temporal hazard associated with target onset, this regressor modulated activity in bilateral lingual cortex, in the cuneus, and in bilateral superior temporal gyrus (Fig. 2.6).

Table 2.2 | Significant cluster activations in SPM analyses.

Anatomical region	MNI			Peak *Z*	Cluster level	
	x	y	z		*p*	Size
Regions modulated by D_{KL}						
L. Fusiform Gyrus	-42	-60	-12	4.52	.001	240
	-38	-56	-6	4.20		
	-36	-62	-34	4.09		
L. Posterior Parietal Cortex	-34	-64	44	4.37	< .001	801
	-48	-44	48	4.08		
	-44	-64	44	3.99		
R. Inf. Frontal Gyrus	50	20	28	4.33	< .001	315
	44	28	20	4.06		
	36	26	20	4.03		
R. Posterior Parietal Cortex	34	-62	34	4.29	< .001	371
	34	-66	46	4.14		
	26	-62	36	4.08		
L. Inf. Frontal Gyrus	-40	4	30	4.08	.016	142
Precuneus	-8	-68	46	3.93	.012	152
	4	-68	46	3.73		

	12	-68	44	3.58		
Post. Cingulate Cortex	-2	-34	26	3.74	.043	112
Regions modulated by I_S						
L. Ant. Insula	-36	22	4	5.36	< .001	515
	-38	14	6	4.59		
	-32	26	-2	4.13		
L. Posterior Parietal Cortex	-34	-60	46	5.26	< .001	932
	-40	-42	46	4.42		
	-36	-42	38	4.08		
R. Ant. Insula	34	24	2	4.90	.003	206
Dorsal Ant. Cingulate Cortex	-4	10	48	4.61	< .001	342
	-8	-4	64	3.65		
	10	16	40	3.26		
R. Posterior Parietal Cortex	34	-68	46	4.50	<.001	356
	30	-62	38	4.03		
	30	-48	44	3.75		
L. Fusiform Gyrus	-38	-60	-12	4.49	.008	170
R. Inf. Frontal Gyrus	40	26	20	4.22	.003	201
Regions modulated by h						
R. Lingual Cortex	24	-74	-4	4.65	.004	180
	16	-64	-4	3.87		
	16	-74	-12	3.50		
L. Sup. Temporal Gyrus	-50	-38	20	4.36	.001	224
	-58	-20	10	4.02		
	-58	-32	18	3.83		
R. Cuneus	14	-82	26	4.23	.003	188
	16	-90	14	3.30		
	22	-88	20	3.29		
L. Lingual Cortex	-10	-80	-2	4.00	.025	122
	-8	-72	-2	3.80		
	-14	-70	-8	3.51		
R. Sup. Temporal Gyrus	58	-12	10	3.94	.009	151
	56	4	-2	3.83		
	52	-16	4	3.58		
L. Cuneus	-16	-88	28	3.73	.032	115
	-4	-86	22	3.73		
	-10	-92	18	3.32		

Note: L. and R. indicate left and right.

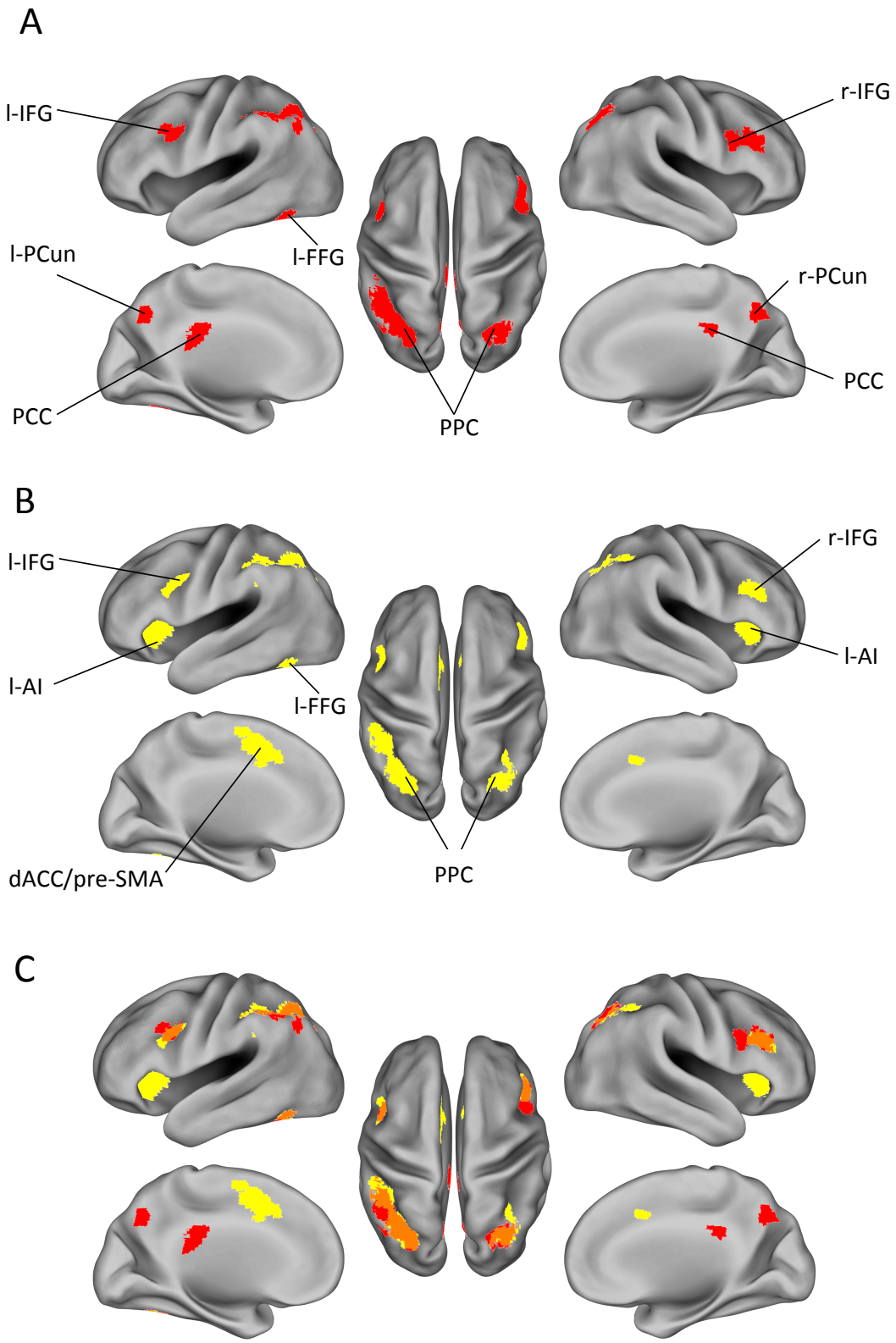

Figure 2.4 | Whole-brain analysis results. (A) Regions significantly modulated by updating (D_{KL}). **(B)** Regions significantly modulated by surprise (I_s). **(C)** Overlapping of activation between D_{KL} and I_s. Abbreviations: IFG: inferior frontal gyrus; PPC: posterior parietal cortex; Pcun: precuneus; PCC: posterior cingulate cortex; FFG: fusiform gyrus. AI: anterior insula; dACC: dorsal anterior cingulate cortex; pre-SMA: pre-supplementary motor area.

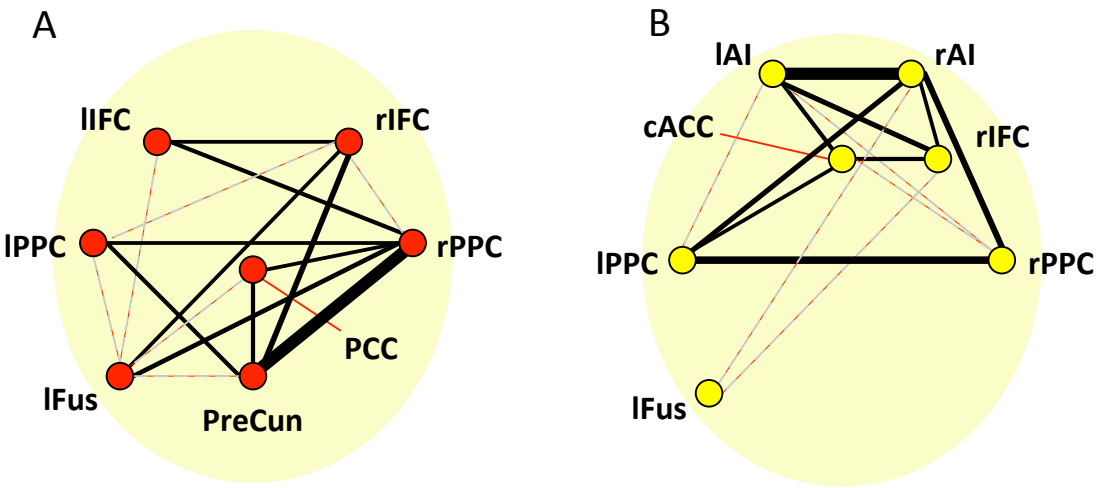

Figure 2.5 | cPPI analysis results. Lines represent significant correlations (FDR =.05) between nodes modulated by **(A)** updating (D_{KL}) and **(B)** surprise (IS). Dashed lines represent significant correlations with p > .01. Line width is proportional to the strength of the correlation.

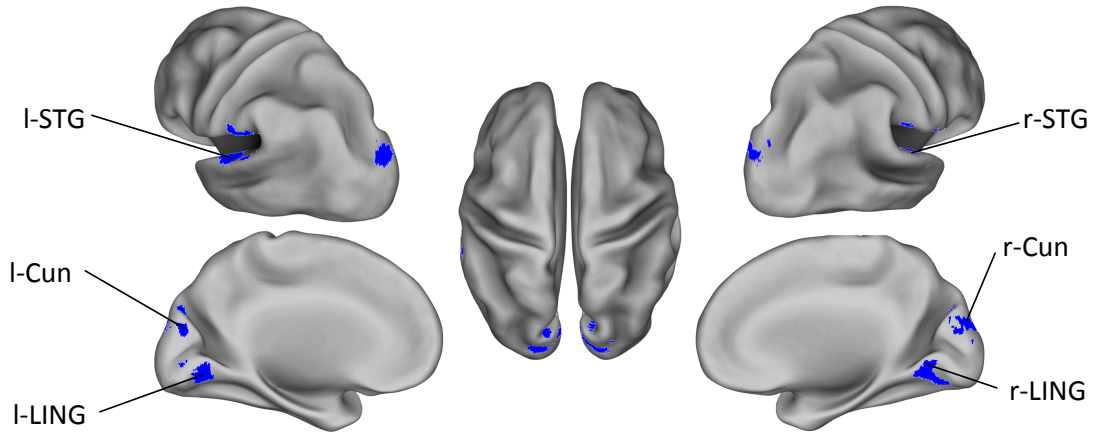

Figure 2.6 | Whole-brain analysis for the temporal hazard (h). Abbreviations: STG: superior temporal gyrus; Cun: cuneus; LING: lingual gyrus.

2.4 Discussion

The present study sought to identify the neural correlates of between-trials updating of temporal expectations and within-trials surprise about variable foreperiod durations. While previous research focused on how the brain tracks temporal expectations during the foreperiod (i.e., hazard function), the mechanisms by which the brain forms and updates such expectations have not been properly

defined. The combination of Bayesian computational modeling of temporal beliefs and single-trial analyses of both fMRI activity and functional connectivity showed the engagement of two sets of brain regions, which differentially encoded updating and surprise. Remarkably, such regions belong to two functional networks that prior work has found to play a key role in cognitive control: the fronto-parietal (FPN) and cingulo-opercular (CON) networks (Dosenbach et al., 2008).

Updating, measured as the Kullback-Leibler divergence (D_{KL}) between prior and posterior beliefs, was mainly associated with regions that are part of the FPN such as the posterior parietal cortex (PPC), the precuneus, the posterior cingulate cortex (PCC) and the inferior frontal gyri (IFG). Surprise, defined as the Shannon's information (I_S) on the hazard rate of target onset, was encoded in regions of the CON such as the dorsal anterior cingulate cortex (dACC), the presupplementary motor area (pre-SMA) and the anterior insulae (aINS), and in some regions of the FPN, namely, the IPL and the right lPFC.

According to recent models of cognitive control (Cocchi, Zalesky, Fornito, & Mattingley, 2013; Crittenden, Mitchell, & Duncan, 2016; Dosenbach et al., 2008), the FPN is involved in the trial-by-trial adjustment of task-relevant information to implement top-down control. An example is offered by the task-switching paradigm in which instantiation/maintenance of the proper task-set is required on each switch/repeat trial (Waskom, Kumaran, Gordon, Rissman, & Wagner, 2014). This property of the FPN is in line with the idea of a Bayesian brain that updates its predictive models after each new observation. Accordingly, Waskom and colleagues (2017) showed that the FPN gradually responded to violation of predictions about forthcoming task-sets in a task-switching-like paradigm. Our study corroborates their findings as we also found the involvement of the FPN when predictions were violated. However, as outlined in the Introduction, events violating our expectations may lead to both surprise and updating. It follows that the FPN activation found by Waskom and colleagues (2017) could reflect either one or both of these cognitive processes, as their paradigm did not differentiate between them. In this regard, our findings critically extended those previous results by disentangling surprise and updating in the FPN. Updating was associated with all the areas belonging to the

FPN. The right PPC and IFG, which are congruent with the areas included in the *ventral* network proposed by Corbetta and Shulman (2002), additionally responded to surprise. The ventral system is thought to deal with salient and unexpected events (Corbetta & Shulman, 2002). It, thus, makes sense to speculate that in our paradigm such areas responded to two types of salient information, one needed for responding quickly to current relevant goals (i.e., within-trial response reprogramming triggered by surprise) and the other one for reconfiguring predictive models. While these lateral fronto-parietal areas were activated for both surprise and updating, other medial parietal areas (Precuneus and PCC) were specifically modulated by updating. This finding is in accord with previous studies showing the involvement of such areas in task-set reconfiguration (Chiu & Yantis, 2009), integration of information between brain systems (Fornito et al., 2012; Leech, Braga, & Sharp, 2012), environmental reward outcomes (Hayden, Nair, McCoy, & Platt, 2008), and encoding of statistical properties of changing environments (Pearson, Heilbronner, Barack, Hayden, & Platt, 2011).

Unlike the FPN, areas belonging to the CON were exclusively associated with surprise. According to the "dual-network" model (Dosenbach et al., 2008), whereas the FPN would implement control on a trial-by-trial basis, the CON would be in charge of maintaining relevant task-goals across trials. Other models, however, have proposed that the AI along with the ACC form a "salience network", which is involved in the transient identification of relevant stimuli in order to guide behavior (Seeley et al., 2007). More in details, Menon and Uddin (2010) proposed that the AI, which receives multimodal sensory inputs, detects salient stimuli and sends transient control signals to the ACC and associated pre-SMA, which in turn send motor outputs to respond to such salient stimuli. This model is more in line with our surprise-related findings than the sustained role of the CON proposed in the dual-network model. Indeed, following Menon and Uddin (2010), we can speculate that in our task too the salience-network detected low-probable targets in order to enhance a non-well temporally prepared response.

Of note, updating and surprise both elicited activity in a cluster of regions around the left fusiform gyrus (lFFG) and the cerebellum. The contribution of the

lFFG could be explained by its putative role in color perception (Zeki et al., 1991). In our task, indeed, cue color was the only information provided to participants in order to discriminate between update and surprise trials. The involvement of the cerebellum could be instead understood within the dual-network framework as it has been suggested that the FPN and CON interact with each other via the cerebellum (Dosenbach et al., 2008). Unfortunately, the field of view used in our fMRI protocol was not optimal for a complete acquisition of the cerebellum and nearby regions and the related results should be interpreted with caution.

As mentioned in the Introduction, only few studies tried to dissociate updating and surprise. O'Reilly and colleagues (2013) separately manipulated surprise and updating in a saccadic planning spatial task. Unlike our study, they observed significant activation of the ACC/pre-SMA for updating and enhanced activity in the PPC for surprise. The discrepancy between their and our results might be due to different task demands. Concerning the correlation between surprise and PPC, the peak of activity (MNI coordinates: [-18 -60 58]) fell in a region called IPS3 (Mars et al., 2011), a homolog of the monkey's saccadic planning area LIP. The authors suggested that surprise-related activity in the IPS3 was elicited by the need of a saccadic reprogramming towards unexpected locations. Thus, the absence in our study of a significant activation in this region might be probably explained by the lack of saccadic motor planning demands. Different task demands can also account for the divergent pattern of ACC/pre-SMA activations reported in our paradigm. For example, it has been shown that SMA and adjacent regions are key areas in tracking the hazard function of predictable temporal events (Herbst et al., 2018). Since in our task surprise only, but not updating, was computed on the hazard rate of target onset, it is plausible that variations in these regions were better captured by I_S than D_{KL}.

In contrast to O'Reilly and colleagues (2013), our findings on surprise replicated those by Schwartenbeck and colleagues (2016). These authors also attempted to dissociate surprise and updating with a focus on the role of dopamine in belief updating. ROI analyses of midbrain activity showed these dopamine-rich regions to be modulated by updating. Conversely, in line with our results, whole-

brain analyses showed surprise to be encoded in medial frontal regions including the dACC and pre-SMA. A last study, in which updating and surprise were dissociated, was conducted by Kobayashi and Hsu (2017). Using and adapted Ellsberg three-color urn task (see General Introduction) the authors found that surprise modulated activity in bilateral insula while updating of belief about the urn content was associated with activity in bilateral middle frontal girus and IPS, and in precuneus. Our findings are consistent with their results on both updating and surprise.

Concerning the areas responding to the temporal probability of target onset, we found that regions located in the auditory and visual cortices were sensitive to the hazard function. These findings corroborate those by Bueti and colleagues (2010), which showed the involvement of sensory visual areas in tracking or at least shadowing elapsed time. Interestingly, it has also been shown that auditory temporal expectations modulated activity in both auditory and visual areas (Bueti & Macaluso, 2010), a result that points to the existence of crossmodal associations in temporal preparation. Here, we further lend support to this idea by showing the involvement of auditory regions in the deployment of visual temporal expectations.

In sum, our fMRI data showed that updating of internal models and surprise about the timing of relevant events do rely on the work of two influential cognitive control networks. In this regard, the value of our study can be appreciated along two directions. On a general level, it sheds new light into the understanding of the differential role of the FPN and CON in higher-order cognition. More specifically, to our knowledge, this is the first study that unveils the neural mechanisms underlying the formation and adjustment of temporal predictive models. Indeed, we showed that our brain encodes temporal probabilities in an optimal Bayesian fashion and that regions involved in updating can be at least in part differentiated from regions dealing with surprising events.

Chapter 3

Electrophysiological correlates of temporal belief updating and surprise

3.1 Introduction

In recent years, Bayesian inference is gaining increasing popularity in cognitive neuroscience (Doya et al., 2007; Friston, 2012; Kersten et al., 2004; Knill & Pouget, 2004). Recent theories propose that, like other adaptive systems, the brain tries to infer the causal structure of the environment in a Bayesian optimal fashion. This means that we derive beliefs according to Bayes' rule in order to predict environmental contingencies. The Bayesian brain hypothesis has been applied to many cognitive domains including perception, language, motor preparation and decision making (Chater & Manning, 2006; Wolpert, 2007; Yuille & Kersten, 2006). As shown in Chapter 2, we successfully used this approach in an fMRI study to model temporal expectations during a foreperiod task. We identified two networks partially distinctly associated with two levels of processes elicited by surprising events: 1) the fronto-parietal network was mainly involved in updating of temporal expectations; 2) the cingulo-opercular network was related to processes dealing with unexpected events. These findings, thus, showed that our foreperiod task was well suited to

isolate neural correlates of processes related to these two levels of cognitive operations. However, the poor temporal resolution of fMRI does not tell us much about the temporal dynamics underlying Bayesian inferences of temporal expectations. The recording of electroencephalographic (EEG) activity may help characterize in a more direct manner the nature of the cognitive processes associated with surprise and updating. To this end, in the present study we combined our previously used foreperiod task with the excellent temporal resolution of EEG.

Recent EEG research suggests a role of Bayesian inference in modulating electrophysiological activity associated with perception and learning (for an overview, see: Kopp et al., 2016a). One of the most studied event-related potentials (ERPs) in this regard is the P3 (more specifically, various components belonging to the P3 family). Traditionally, it has been consistently shown that "surprising events elicit a large P300 component" (Donchin, 1981, p. 498), which has led to the hypothesis that the P3 amplitude is inversely related to the observer's subjective probability of the event (Donchin & Coles, 1988). Although an explicit link between the P3 amplitude and the Bayesian brain hypothesis has been proposed relatively recently (Friston, 2005; Kopp, 2008), since its discovery terms related to information processing and inference such as expectancy, uncertainty, subjective probability, surprise, or updating have been often employed to functionally characterize this ERP component (Donchin, 1979, 1981; Squires et al., 1976; Sutton, 1979). For example, according to the influential "context-updating" hypothesis (Donchin & Coles, 1988) "the P300 is elicited by processes associated with the maintenance of our model of the context of the environment" (p. 370). This theoretical account for the P3 is highly compatible with Bayesian inference (Kopp, 2008): the prior could be seen, indeed, as a conceptual cognate of the internal context model and updating as the process that guarantees the optimal maintenance of this model.

Corroborating the similarity between former interpretations of the P3 and the Bayesian brain hypothesis, recent studies have shown that trial-by-trial fluctuations in the P3 amplitude could be explained by means of an ideal Bayesian observer (e.g., Mars et al., 2008; see also Chapter 2). Some studies have consistently

reported that some P3 sub-components can be differentiated in terms of distinct measures of surprise: Shannon's information (Is) and Kullback-Leibler divergence (DKL). In particular, a parietally-distributed P3b has been associated with Is (Kolossa et al., 2013; Mars et al., 2008), while an earlier and more fronto-central P3a with DKL (Kolossa et al., 2015). However, the precise nature of the processes captured by these two measures of surprise is still elusive. Concerning DKL, even if it plausibly represents a formal measure of processes underlying Bayesian updating, it is unclear whether its neurophysiological correlates (e.g., P3a) reflect the actual updating of internal models or they just index more general attentional processes also playing a role in updating. As regards Is, although its neurophysiological correlates (e.g., P3b) have been related to uncertainty or surprise, previous studies have not specified the nature of the underlying processes.

The present study aimed at investigating the neurophysiological signatures of processes formalized by Is and DKL by making a distinction between two levels of cognitive operations. As detailed above and also in Chapter 2, the first level is given by between-trial processes involved in the updating of predictive models. The second level is related to within-trial processes coping with surprising events. The implementation of a temporal preparation task is very useful to differentiate these two levels of analysis since they are affected by two different types of probabilistic information conveyed by the target. On the one hand, DKL uses the probability of target onset with respect to the prior (likelihood). On the other hand, Is is associated with the probability of target onset given that it has not yet occurred within the trial (i.e., hazard rate). Accordingly, our manipulation allowed making sure that our DKL is reflecting updating, while Is is genuinely associated with processes involved in surprise (e.g., reprogramming a response) but, critically, not in updating.

In order to depict the temporal dynamics of the processes involved in updating and surprise, here we tried to move a step forward with respect to previous ERP studies on this topic. These studies investigated the effects of continuous variables such as Is and DKL on EEG measures extracted from a-priori time windows (i.e., mean amplitude and/or peak) and electrodes of interest. The

main disadvantage of this approach is that it cannot provide a temporally detailed waveform (N. J. Smith & Kutas, 2015). Moreover, it is plausible that processes evoked by updating and surprise occur in close temporal proximity and, accordingly, that the resulting EEG signal may reflect some overlap between the two that could be confounded. Finally, other modulations above and beyond the P3 could be associated with updating and surprise but they have been probably overlooked in previous studies. To face all these limitations, in the present study we adopted a regression-based mass univariate approach that considered the entire spatio-temporal EEG domain (Ehinger & Dimigen, 2018). In doing so, we aimed at providing a more temporally defined and comprehensive picture about the ERPs reflecting updating and surprise than what already done in previous EEG studies. Importantly, to the best of our knowledge, our study represents the first attempt to delineate the neural correlates of Bayesian updating of predictive models about the timing of forthcoming events.

3.2 Methods

3.2.1 Participants

The study included an initial sample of 26 participants. One participant was excluded due to low compliance with task instructions (22% of responses were anticipations) and replaced with an additional participant. Therefore, the final sample still comprised 26 participants [10 males; mean age: 23.4 (SD = 3), range: 19-33 years old]. All of them were right-handed [EHI average score: 81.8 (SD=16)] except one who showed weak right-handedness (EHI = 20), they reported no history of neurological or psychiatric disorders, normal color vision and normal or corrected-to-normal visual acuity. The procedures involved in this study were approved by the Bioethical Committee of the Azienda Ospedaliera di Padova. Participants gave their written informed consent before the experiment, in accordance with the Declaration of Helsinki, and they were reimbursed 20 euros for their time.

3.2.2 Task and procedure

The foreperiod task was the same as the one employed in Study 1 (Chapter 2). However, since in an EEG paradigm we do not need ITI as long as in fMRI paradigms, in this task we could double the number of trials. Overall, participants performed 72 blocks (plus an initial block excluded from the analyses) whose Gaussian distributions were derived from an orthogonal combination of 9 means (500, 700, 900, 1100, 1300, 1500, 1700, 1900, and 2100 ms) and 8 standard deviations (25, 50, 75, 100, 125, 150, 175 and 200 ms). The block sequence is presented in Table 3.1). The increase in the number of trials allowed us to have a more balanced paradigm in terms of distances between subsequent means. Furthermore, we could increase variability in the precision of the Gaussian distribution also by increasing the number and the duration of possible standard deviations. Foreperiod durations in uniform trials were drawn from a uniform distribution with boundaries 250 and 2500 ms.

Differently from the task in the previous study the ITI duration was randomly jittered between 1.25 and 1.75 s and the target was displayed for 1 second. Each block consisted of a number of trials between 8 and 13 (mean: 10.5). The blocks were grouped in four runs with self-paced breaks in between. Before the EEG montage, participants practiced the task. The practice session and the instructions were identical to those provided in Study 1.

Table 3.1 | Block-list.

Block	μ	σ	Block	μ	σ	Block	μ	σ
0	500	175						
1	1900	125	25	1300	50	49	1500	100
2	1500	175	26	1100	50	50	1100	200
3	500	100	27	1700	50	51	2100	125
4	1100	25	28	1300	25	52	1300	200
5	900	75	29	1100	175	53	500	200
6	1300	100	30	2100	175	54	2100	100
7	700	150	31	700	75	55	1500	125
8	1700	75	32	1100	100	56	1100	75
9	1300	150	33	500	175	57	1700	100
10	1900	25	34	700	25	58	900	25
11	1700	25	35	900	200	59	700	125
12	2100	150	36	2100	75	60	1500	200
13	1300	75	37	1500	150	61	500	125
14	1500	50	38	900	175	62	1300	175
15	1700	150	39	2100	50	63	1900	175
16	700	50	40	1700	125	64	700	200
17	1100	150	41	900	50	65	2100	25
18	1900	50	42	1100	125	66	1900	200
19	2100	200	43	1900	75	67	500	50
20	500	75	44	1700	175	68	900	150
21	900	100	45	500	25	69	1900	100
22	1300	125	46	700	175	70	1500	75
23	700	100	47	1900	150	71	1700	200
24	1500	25	48	900	125	72	500	150

Notes: For each block, mean (μ), and standard deviation (σ) of the generative FP distribution are indicated and expressed in ms. Block 0 was not included in the analyses.

3.2.3 EEG data acquisition

The EEG was recorded using BrainAmp amplifiers (Brain Products, Munich, Germany) from 64 Ag/AgCl electrodes that were mounted on an elastic cap (EASYCAP GmbH, Germany) according to the international 10-20-system. Electrooculographic activity was recorded through an electrode placed under the right eye and was also monitored through the scalp electrodes placed in the proximity of both eyes. Before recording, impedance for all electrodes was checked

and adjusted until it was lower than 10 $k\Omega$ before testing. All electrodes were referenced to FCz during the recording, and an electrode positioned at AFz served as the ground. EEG activity was digitized at a sampling rate of 500 Hz.

3.2.4 Normative Bayesian learner and regressors

The model was the same as the one described in Study 1 (see 2.2.3). The only difference concerned the parametric space. Due to the new standard deviations, the size of the parameter space was 300×25. As in Study 1, surprise and updating were quantified using I_S and D_{KL}, which were derived as in 2.2.4.

3.2.5 EEG data analysis

EEG pre-processing. Offline EEG processing was performed using custom MATLAB scripts using functions from the EEGLAB environment (version 13.6.5b; Delorme & Makeig, 2004).

As a pre-preprocessing step for subsequent ICA, a band-pass filter using a one-pass non-causal zero-phase Kaiser windowed sinc FIR filter [cut-off frequencies = 2 and 40 Hz, transition bandwidth = 4 and 20 Hz for the high- and low-pass filters, respectively, passband ripple = .001] was applied to continuous EEG data. The high cut-off for the high-pass filter was applied to remove low-frequency drifts in order to improve the results of ICA (Winkler, Debener, Muller, & Tangermann, 2015). The *clean_rawdata* function was used to remove noisy channels (channel criterion = .8) and short-time bursts (burst criterion = 20 SD) from the data. A maximum of 3 channels per subject (mean = 0.6, sd = 1) were removed. Then, the FastICA algorithm (Hyvarinen & Oja, 2000) was employed to obtain ICA weight matrices and sphereing matrices.

Whilst the use of such an extreme high-pass filter cut-off stabilizes ICA solution, it is known that high-pass filtering may attenuate ERP effects and introduce distortions. For this reason, the ICA solution calculated on 2-Hz high-pass filtered data was then applied on continuous EEG data band-pass filtered using a one-pass non-causal zero-phase Kaiser windowed sinc FIR filter with cut-off frequencies of 0.1

(transition bandwidth = 0.2) and 40 Hz (transition bandwidth = 20) for the high- and low-pass filter, respectively. Indeed, the 0.1-Hz cut-off frequency seems a good trade-off between waveform distortions and statistical power (Tanner, Morgan-Short, & Luck, 2015). Before applying the ICA solution, noisy channels identified in 2-Hz high-pass filtered data were removed. Subsequently, the EEGLAB extension SASICA (Chaumon, Bishop, & Busch, 2015) was used to guide the identification and exclusion of artifact independent components (e.g., blinks, eye movements, muscle activity, misconnected channels). Finally, removed channels were interpolated using spherical splines (Perrin, Pernier, Bertrand, & Echallier, 1989) and continuous EEG data were re-referenced to the average of all EEG electrodes.

Inferential statistics. First-level (subject-specific) analysis was performed using the *Unfold* toolbox (Ehinger & Dimigen, 2018) in MATLAB. This toolbox allows performing regression-based EEG analysis by integrating a mass-univariate approach with linear deconvolution. Deconvolution tries to disentangle overlapping electrophysiological responses from subsequent events (e.g., from stimulus onset and from button press). This aspect is very helpful in our paradigm to analyze the neural response associated to target onset. Indeed, targets were preceded by the foreperiod interval in which several processes likely occurred. Since targets appeared at different stages of these foreperiod processes depending on their status (i.e., predictable vs. surprising), the resolution of electrophysiological correlates of such processes differently affected target-related ERPs. Deconvolution helped us to face this problem by isolating specific ERPs associated with target onset. For a methodological description of the "Unfold" analysis steps the reader is referred to the original paper (Ehinger & Dimigen, 2018). For the analysis we specified three events: cue onset, target onset and button press. Target onsets were modeled according to the following Wilkinson-notation (Wilkinson & Rogers, 1973) formula: $y = 1 + D_{KL} + I_S$, where *1* represents the intercept term, and I_S and D_{KL} are our model-derived parametric regressors for updating and surprise, respectively. Since we were not interested in modeling cue onset and button press, these events were modeled using only an intercept term ($y = 1$). The design matrix was time expanded from -1 s to 1 s around each event using a set of spline functions (1 s is the

target stimulus duration and, therefore, the maximum time allowed to respond). Before fitting the model, artifact intervals were identified using a peak-to-peak voltage threshold of 75 μV and removed from the design matrix (i.e., set to 0).

Second-level analysis was performed using the *ept-TFCE* toolbox (Mensen & Khatami, 2013) 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 one-sample *t*-test (number of permutations = 200000, alpha-level = .001).

3.3 Results

3.3.1 Behavioral results

Response time analysis is described at 2.2.5. Backward elimination of non-significant effects resulted in a model specified as the following Wilkinson-notation formula:

$$\log(RT) \sim \text{TRIAL} + \text{PRECEDING RT} + I_S * D_{KL} + (\text{TRIAL} + \text{PRECEDING RT} + I_S \mid \text{ID}). \quad (3.1)$$

Visual inspection of the residuals showed that the model was a bit stressed. As suggested by Baayen and Milin (2010), trials with absolute standardized residuals higher than 2.5 standard deviations were considered outliers and removed (1.5% of the trials). After removing outlier trials, the model was refitted and this time it achieved reasonable closeness to normality. Table 3.2 shows the statistical results of the type III ANOVA. A significant interaction was found between I_S and D_{KL}. Fig. 3.1 shows that RTs increased with increasing surprise (I_S) and this effect was augmented with increasing D_{KL}.

Table 3.2 | Analysis of variance with type-III sums of squares.

Fixed Effects	Sum Sq	Num. *df*	Den. *df*	*F*	*p*	β
TRIAL	0.15	1	24.9	4.10	.054	-0.04
PRECEDING RT	1.94	1	24.2	52.32	< .001	0.14
I_S	9.14	1	26.4	246.50	< .001	0.26
D_{KL}	2.89	1	18367.8	77.91	< .001	-0.10
I_S :D_{KL}	4.58	1	18368.2	123.64	<. 001	0.12

Notes: F-statistics and associated p-values were calculated using Kenward-Roger's approximation of degrees of freedom (Halekoh & Højsgaard, 2014). Additionally, standardized regression coefficients (β) are shown.

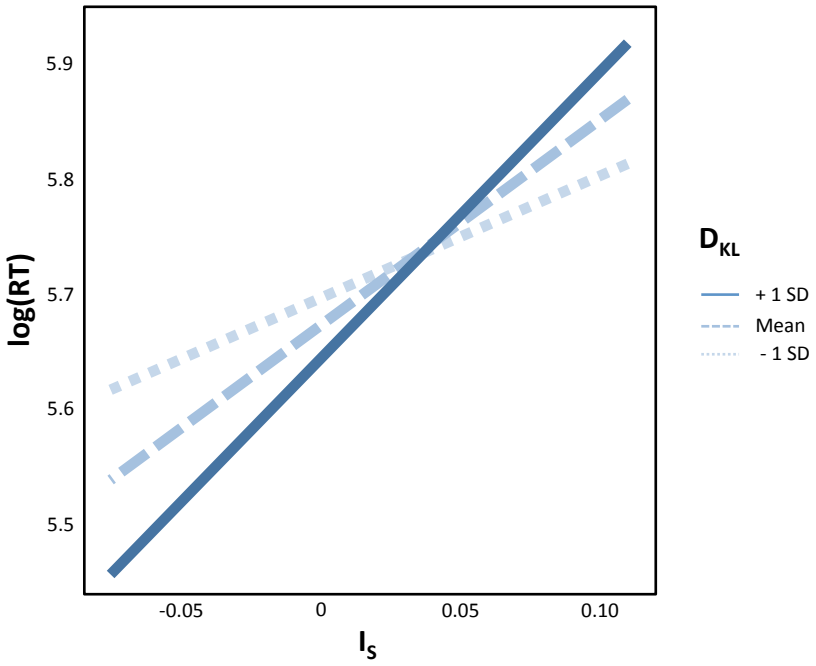

Figure 3.1 | Interaction plot for log-transformed RTs. The plot shows the effect of surprise (I_S) at three levels of of updating (D_{KL}), high (+1SD), medium (Mean) and low (-1SD). T

3.3.2 Electrophysiological results

The results of the TFC analyses on surprise and updating showed that both processes were associated with several electrophysiological modulations displaying distinct spatio-temporal characteristics.

Concerning surprise (Is), as portrayed in Fig 3.2 (warm color) the first significant effect was an early positive deflection emerging in the 70-130 ms time window, which was distributed over posterior scalp electrodes (P4, P6, PO3, POz, PO4, PO8, O1, Oz, O2). Correspondingly, a negative modulation (Fig. 3.2, cold color) was observed over frontal electrodes (Fp1, Fpz, Fp2, AF7, AF3, AF8, AF4, F7, F5, F6). Fig. 3.4A shows the corresponding topographic maps and the effects averaged across participants for these results. The second modulation, represented in Fig. 3.2 (warm color) was a larger and wider-spread positivity that developed over centro-parietal electrodes (Cp1, Cp3, Cpz, Cp2, Cp4, P1, P3, Pz, P2, P4, PO3, POz, PO4) and lasted for a longer time window ranging from 360 ms to 680 ms. As depicted in the topographic map (Fig. 3.4B), this positivity was surrounded by a negativity over lateral frontal electrodes (Fp1, Fpz, Fp2, AF7, AF3,AF8, AF4, F7, F5, F8, F6, FT9,FT7, FT10, FT8, FC5, FC6, T7, T8, TP9, TP10).

The results of the TFCE analysis on Updating (D_{KL}) are portrayed in Fig. 3.3: It is clear that, in contrast to surprise, updating was associated with a more complex electrophysiological pattern. Specifically, there was a first double-peak waveform that was distributed over fronto-central scalp regions (FP1, Fpz, Fp2, AF7, AF3, AF4, AF8, F7, F5, F3, F1, Fz, F6, F4, F2, FC5, FC3, FC1, FCz, FC4, FC2, C3, C1, Cz, C4, C2) lasting from 50 ms to 290 ms (Fig. 3.5A). Between the first and second peak, there was a positive modulation distributed over parietal electrodes (CP1, P1, P3, CPz, Pz, CP2, CP4, P2, P4, P6) from 100 ms to 200 ms (Fig. 3.5C). Such modulations were followed by a later (from 320 ms to 480 ms) positivity (Fig. 3.5B), which was significantly present over occipital electrodes (TP9, TP10, P7, P6, P8, PO3, PO7, PO4, PO8, O1, Oz, O2). The last modulation (Fig. 3.5C) was a positivity that developed at 450 ms and lasted until 880 ms over centro-posterior electrodes (CP1, CP3, CP5 CPz, CP2, CP4, CP6, P1, P3, P5, Pz, P2, P4, P6, PO3, POz, PO4, PO8).

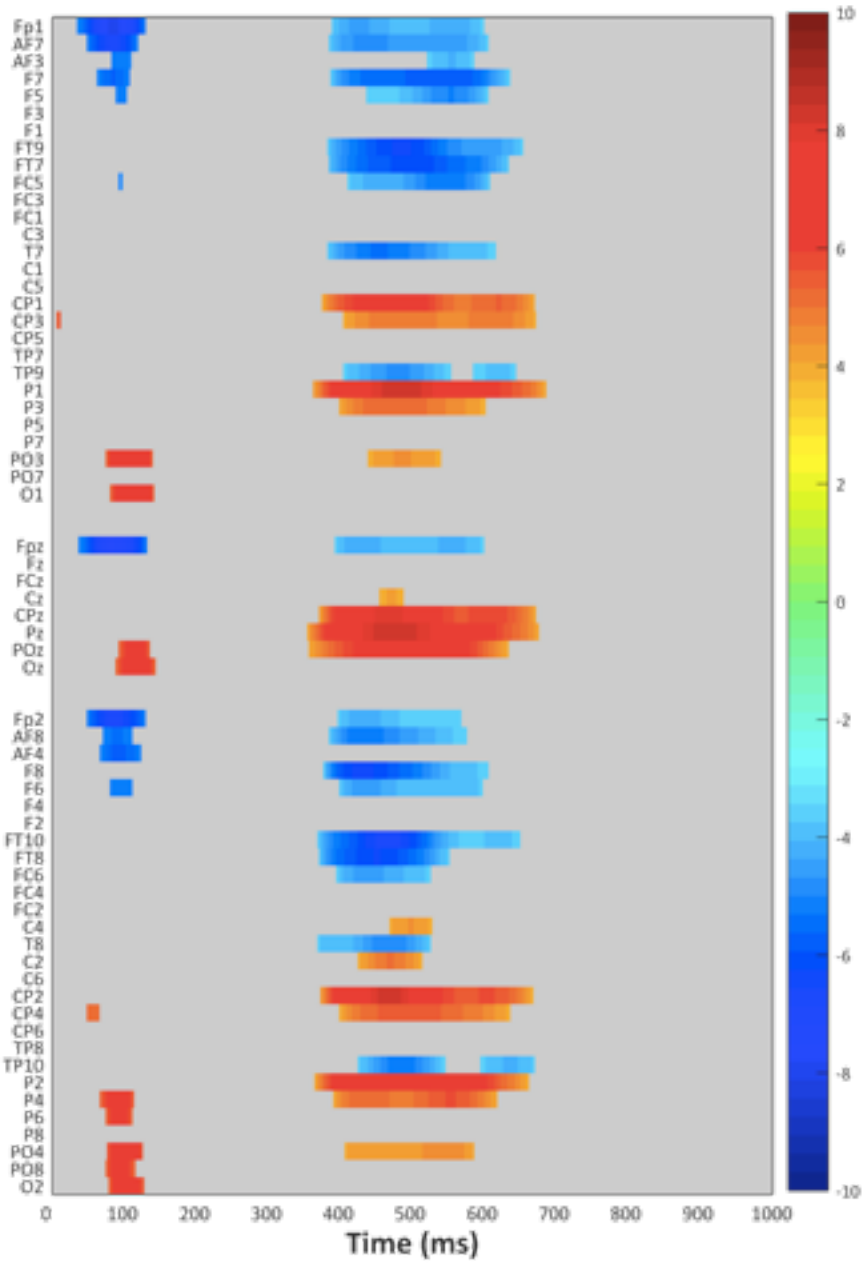

Figure 3.2 | Raster diagram showing significant effect elicited by surprise (I_S) according to TFCE analysis. Rectangle in warm and cold colors indicates electrodes/time points significantly modulated by I_S positively and negatively, respectively. The colorbar on the right indicates *t* values. Gray rectangles indicate electrodes/time points that were not significantly modulated. Note that electrodes are organized along the y-axis somewhat topographically (Groppe, Urbach, & Kutas, 2011). Electrodes on the left side of the scalp are grouped on the top part of the diagram, midline electrodes are shown in the middle, and right electrodes are grouped at the bottom. Within those three groupings, the y-axis top-to-bottom corresponds to scalp anterior-posterior.

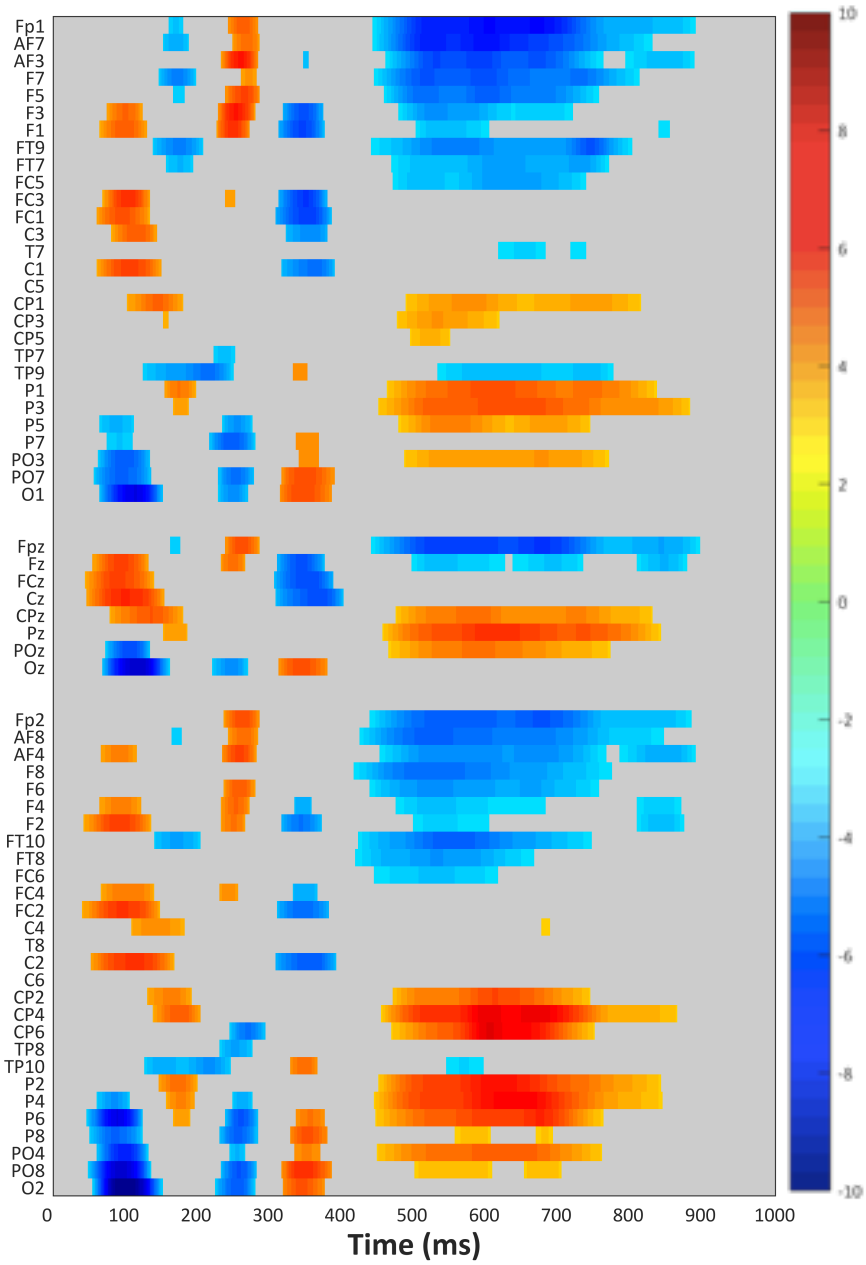

Figure 3.3 | Raster diagram showing significant effects elicited by updating (D_{KL}) according to TFCE analysis.

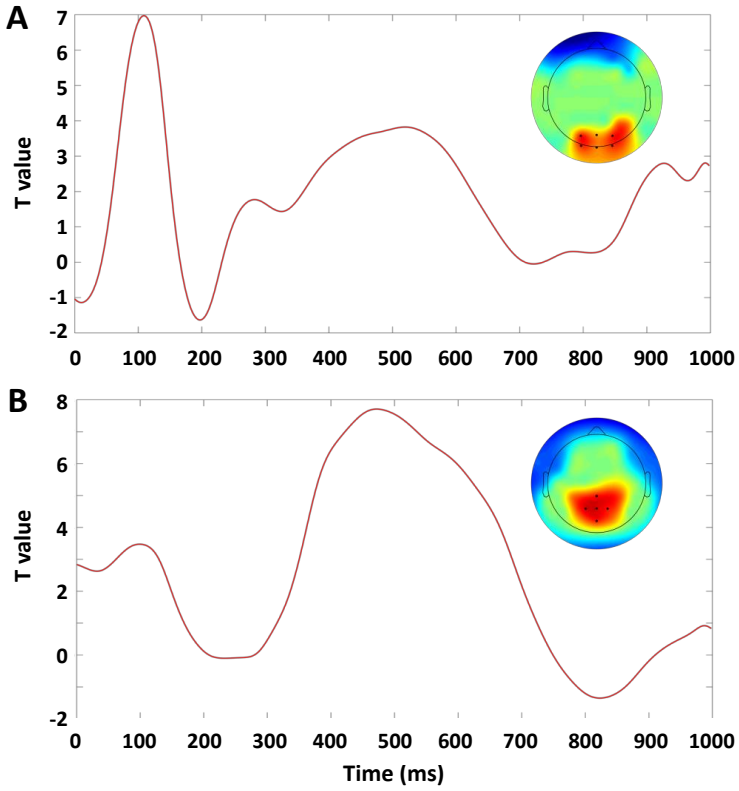

Figure 3.4 | Electrophysiological results: surprise (I_S). (A) The trace plot depicts the average *t* value pooled over the electrodes PO3, POz, PO4, O1, OZ, O3. These electrodes are indicated as black circles in the topographical map on the right. The topographical map shows the *t* values averaged in the time window ranging from 80 ms to 120 ms. The color scale is the same as that of the raster diagram. **(B)** The trace plot depicts the average *t* value pooled over the electrodes Cz, CP1, CPz, CP2, Pz. These electrodes are indicated as black circles in the topographical map on the right. The topographical map shows the *t* values averaged in the time window ranging from 450 ms to 550 ms.

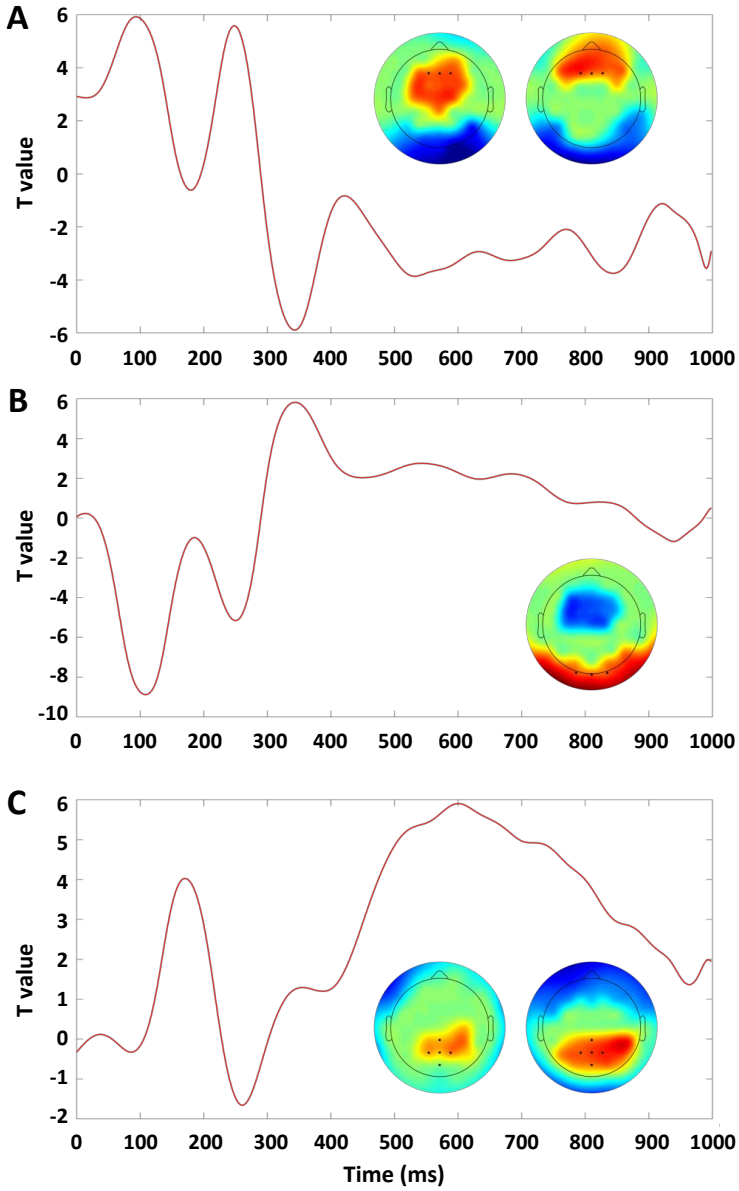

Figure 3.5 | Electrophysiological results: updating (D_{KL}). (A) The trace plot depicts the average *t* value pooled over the electrodes F1, Fz, F2. These electrodes are indicated as black circles in the topographical maps on the right. The topographical map on the left shows the *t* values averaged in the time window ranging from 80 ms to 120 ms. The topographical map on the right shows the *t* values averaged in the time window ranging from 240 ms to 2600 ms. (**B)** The trace plot depicts the average *t* value pooled over the electrodes O1, Oz, O2. The topographical map shows the *t* values averaged in the time window ranging from 340 ms to 360 ms. (**C)** The trace plot depicts the average *t* value pooled over the electrodes Cz, CP1, CPz, CP2, Pz. The topographical map shows the *t* values averaged in the time window ranging from 550 ms to 650 ms. The topographical map on the left shows the *t* values averaged in the time window ranging from 160 ms to 200 ms. The topographical map on the left shows the *t* values averaged in the time window ranging from 550 ms to 650 ms.

3.4 Discussion

In the present study, we investigated the electrophysiological correlates of updating and surprise in a modified foreperiod temporal preparation task that allowed separating updating of temporal expectations from processes dealing with surprising events. Corroborating previous results, we replicated the modulation of late components (as indexed by a P3-like potential) by both surprise and updating. Moreover, our channel- and time-uninformed mass univariate approach revealed that probabilistic and inferential processes also acted on earlier processing stages. In brief, surprise was associated with only two significant modulations, an earlier phasic modulation and a later and longer one. Updating, instead, triggered a more complex electrophysiological pattern composed of a first series of early and fast deflections followed by a later and more sustained component. Below, we elaborate on each one of these results.

As concerns early processing stages, surprise elicited a positive component peaking at around 100 ms, whose occipital scalp distribution resembled a P1 waveform. The common finding in studies of temporal attention, in which a symbolic cue predicts the likely timing of target onset, is that early components are not modulated by valid temporal expectations, unless a demanding discrimination task is used (see Correa, for a discussion on this issue). At first glance, our finding of a larger P1 for surprising events could thus seem at odds with such previous studies. However, it has been shown that P1 also responds to color processing (Forder, He, & Franklin, 2017). Taking into account that higher surprising events (i.e., both update and surprise trials) were always associated with a color change in the target, it might be possible then that the surprise-related P1 modulation found here was just reflecting a low-level perceptual encoding rather than an attentional one. This explanation is indirectly supported by our previous fMRI study (see Chapter 2), in which we found activation of the fusiform gyrus for the surprise modulator.

In sharp contrast with surprise, updating acted on several early processing stages. As it is evident from Fig. 3.3, there were a first double-peak waveform distributed over fronto-central sites and a parietal distributed P2-like deflection followed by a later P2-like posterior positivity, which was more pronounced over

occipital electrodes. Overall, this pattern of electrophysiological modulations strengthens the idea that updating involved more high-level perceptual and memory processes than surprise. Indeed, similar frontal ERPs as the ones reported here have been observed in perceptual decision-making studies (e.g., Go-NoGo), in which they have been considered as markers of prefrontal activity reflecting top-down perceptual processing, such as perceptual awareness (Di Russo et al., 2017). Concerning the parietal positive modulation, Kononowicz and van Rijn (2014), have shown that the amplitude of a similar P2 component was proportional to the distance between standard and comparison intervals in a temporal discrimination task. These findings may suggest that our parietal P2-like component was an index of some comparison mechanisms involved in updating. Concerning the posterior modulation, previous cuing and priming studies have shown that the P2 was enhanced for invalid compared to valid targets, while it was attenuated for target repetitions, respectively (Freunberger, Klimesch, Doppelmayr, & Holler, 2007). These findings, thus, suggest that the P2 may index the recruitment of top-down mechanisms underlying the revision of participants' expectancies. In any case, it is important to note here that it is quite difficult to ascribe a univocal functional meaning to each one of the early modulations elicited by updating since, to our knowledge, there are no previous similar studies of updating of temporal expectations against which to compare our results. Further analyses such as principal component analysis might help in reducing the dimensionality of these findings and their interpretation. What it is critical to outline, however, is that surprise and updating acted differentially on the first stages of target processing, which confirms that our modulators managed to capture two distinct classes of cognitive operations already early along the processing stream.

Our results on the late electrophysiological responses extend previous findings on the relation between P3 and Bayesian inference in the temporal domain and also help further characterize the cognitive processes underlying the P3. Concerning surprise, we found a significant positive modulation that, according to its timing and scalp distribution, can be described as a P3b component. The strong modulation of the P3b by surprise replicates previous studies (Kolossa et al., 2013;

Kolossa et al., 2015; Mars et al., 2008). In this regard, it should be acknowledged that prior work has not provided an exhaustive definition of the cognitive processes captured by surprise. We believe that more information on this issue can be drawn from our results, as detailed in what follows. First of all, the nature of our task, which used a color manipulation to differentiate between surprise and updating, allowed minimizing the possibility that the surprise-related P3b found here reflected processes involved in the revision of an internal model. Such a conclusion is bolstered by the fact that we computed surprise on the target hazard rate rather than on its prior probability (i.e., the expected foreperiod duration at the beginning of the trial). As explained in the Introduction and in Chapter 2, the information derived by the hazard function is indeed very useful for the current trial (i.e., for processes subserving immediate behavioral response) but less for updating of expectations about future foreperiod durations in the forthcoming trials. A second point is that previous studies typically used paradigms with two or more target types associated with distinct motor outputs (i.e., n-choice task). This inevitably creates a confound since the P3b evoked by surprise in this kind of designs may either reflect the surprise triggered by the sensory stimulus or the surprise related to the selection of the corresponding (less expected) motor response (Barceló, Perianez, & Nyhus, 2008; O'Connell, Dockree, & Kelly, 2012). Given that our task required only one motor response, we can then conclude that violations of sensory expectations per se (in our case temporal expectations) are sufficient to elicit the P3b. Nevertheless, since interactions between temporal preparation and task-demands have been shown to be captured by P3-like components (cf. Barceló & Cooper, 2018), it would be interesting for future research to compare our P3 modulation with that elicited in a two-forced choice paradigm.

Similarly to surprise, updating also elicited a positive potential distributed over parietal sites with a later and more sustained activity compared to the surprise-P3b described above (from 450 ms to 880 ms vs. from 360 ms to 680 ms, respectively). As for the results concerning surprise, we suggest that such a positivity may resemble a P3b-like component. Our finding of a significant modulation of the P3b by updating is in disagreement with previous studies reporting that updating is

usually associated with an earlier and more anterior P3a (Bennett et al., 2015; Kolossa et al., 2015). However, there is some recent evidence that may help explain the discrepancy between our results and those previous findings. In particular, Kolossa and colleagues (2015) used a modified version of the urn-ball paradigm and decomposed updating into two distinct processes labeled “Bayesian surprise” and “postdictive surprise”. Bayesian surprise represents the updating of beliefs about hidden states given current observation. Postdictive surprise represents the change in beliefs about future observations given current observation. To better understand the difference between these two measures, we make a parallel with our task. Hidden states in Kolossa and colleagues’ (2015) study could be associated with the Gaussian distribution from which normal FP can be drawn in our task (i.e. the parameter space, see 2.2.4). Thus, Bayesian surprise represents changes in the probability over the parameter space before and after each FP. This measure was not considered in our study. By contrast, postdictive surprise would correspond to the D_{KL} that we used here to represent updating. Kolossa and colleagues (2015) found that the P3a was mainly modulated by Bayesian surprise, whereas the postdictive surprise better explained a positive slow wave (SW) emerging after the P3b. In addition to these two measures, they also observed that surprise (I_S) was related to the P3b. Considering the different EEG analytic approach used here, it is reasonable to speculate that our surprise- and updating-related P3-like components may correspond to Kolossa and colleagues (2015) P3b and SW potentials, respectively. Coupled with the above-mentioned studies (Bennett et al., 2015; Kolossa et al., 2013; Kolossa et al., 2015; Mars et al., 2008), then, our findings reinforce the idea that the P3 family is a valuable index to get access to the Bayesian brain across very different paradigms and cognitive domains. In this regard, it would be interesting in future studies to compare our updating- and surprise-related P3-like components with those obtained in tasks in which uncertainty relates to “what” instead of to “when”.

In sum, in the present EEG study we isolated the electrophysiological responses specifically associated with surprise and updating during a temporal preparation task. In order to achieve this, we relied exactly on the same

manipulation employed in our previous fMRI study. However, one might argue that it is difficult to generalize our results to common real life situations in which, normally, changes in the environment are not explicitly signaled as done here. Despite this point does not undermine the validity of our paradigm, it raises an important question that needs to be answered: How updating of temporal expectations is accomplished when Bayesian inference is implicitly rather than explicitly driven? This question represents the starting point for our second EEG study, which will be described in the next chapter.

Chapter 4

Beyond explicit inference: EEG correlates of implicit updating of temporal expectations

4.1 Introduction

In the previous fMRI and EEG studies of the present thesis, we investigated the neural mechanisms by which temporal expectations are updated and separated them from those responsible for an immediate reaction to a surprising event. To disentangle updating and surprise, we used a color manipulation by which changes in the current generative foreperiod distribution were explicitly signaled. However, belief updating is not only driven by explicit cues but also it can be implicitly accomplished. In this regard, it has been shown that explicit and implicit inferential processes are differently encoded by the brain (e.g., Hayden, Smith, & Platt, 2010; Pearson et al., 2011; Yu & Dayan, 2005). Given that so far we have explored the mechanisms involved in the temporal updating induced by explicit changes in the

environment, we could not generalize our results to situations in which such changes implicitly occur.

Thus, the main goals of the present study were to investigate the electrophysiological correlates underlying "implicit" Bayesian updating of temporal expectations and to directly compare them with the explicit one. To these aims, we employed the same task as that used in the previous EEG study (Chapter 3), but without the color manipulation. Here, participants were simply instructed to respond as fast as possible to the onset of the target but they had no information about the probabilistic structure of the task. This modification in the paradigm required the implementation of a new *ideal Bayesian observer*, which should take into account not only differences in the paradigm and the participants' knowledge about the task (i.e., different task instructions), but also the long-term trialwise accumulation of evidence and memory constraints. Namely, in the previous paradigm, after the change in color, participants could discard the information carried out by the previous trial since this was no longer useful to infer the new, current foreperiod distribution. Conversely, the absence of the color in the present study led to the fact that participants were exposed to a long series of foreperiods that was experienced as a unique sequence throughout the task. This aspect was implemented in the model, as elaborated in the Method section.

As in our previous studies, we also aimed at differentiating the processes involved in updating from those dealing with surprising events. Even if here we discarded the color manipulation, we reasoned that the temporal nature of our task should still allow us to distinguish between the two. Specifically, as mentioned in the previous chapters, the two measures used to quantitatively describe updating (D_{KL}) and surprise (I_S) relied on two types of temporal information carried by the target. Briefly, surprise was computed on the hazard rate related to target onset that, as already explained, considered the passage of time during the trial. The information conveyed by the passage of time was not instead used for the calculation of trial-to-trial updating of temporal expectations. The computational distinction between the modulators for updating and surprise also holds for the present temporal

preparation task, which made us confident to observe again such a difference at the electrophysiological level.

The previous EEG study showed that surprise and updating acted on both early and late processing stages. The modulation of later components was in line with the literature (Kolossa et al., 2013; Kolossa et al., 2015; Mars et al., 2008) confirming that the P3 is sensitive to inferential processes (Kopp, 2008). In contrast, we could not draw firm conclusions on the specific role of earlier components. This is due to the fact that, since processing of the cue color necessarily affected early components, it is not clear whether such potentials truly encoded probabilistic information or they were just involved in the cue elaboration. Testing this issue is crucial to pinpoint the difference between explicit and implicit inferential processes and to better interpret the results of our previous study.

To sum up, here we tried to gain further insight into the inferential mechanisms underlying temporal expectations when they are implicitly driven and to make a comparison between explicit and implicit inferential processes.

4.2 Methods

4.2.1 Participants

The study included a sample of 28 participants [17 females; mean age: 24.5 (SD = 4), range: 19-33 years old]. All of them were right-handed [EHI average score: 82.7 (SD=19), range: 40-100], they reported no history of neurological or psychiatric disorders, normal color vision and normal or corrected-to-normal visual acuity. The procedures involved in this study were approved by the Bioethical Committee of the Azienda Ospedaliera di Padova. Participants gave their written informed consent before the experiment, in accordance with the Declaration of Helsinki, and they were reimbursed 20 euros for their time.

4.2.2 Task and procedure

The foreperiod task was the same as the one employed in the previous EEG study (Chapter 3), except for the target color that was always black. Participants were not informed about the temporal structure or the probabilistic nature of task. They were only instructed to respond as fast as possible to the onset of the target and to avoid responding before the target had appeared.

4.2.3 Normative Bayesian learner and regressors

For the present study, we modified the normative Bayesian learner employed in the former two studies (Section 2.2.3; see: O'Reilly et al., 2013) in order to reflect the features of the present paradigm. Before describing the new model, we will present the differences in task structure between the present and the previous two studies that were considered for the implementation of the model.

As stated by O'Reilly and colleagues (2013), the model was optimal, meaning that it provided the best estimate of the generative foreperiod distribution given the data sequence (the same experienced by the participant) and that it took into account participants' knowledge about the structure of the task based on the given instructions. Since participants were informed that white targets appeared after foreperiods of random duration, the model did not update after uniform foreperiods (see eq. 2.4). Moreover, since participants were instructed that the change of target color signaled the beginning of a new distribution, observation from previous distributions were blanked. Consequently, after update trials, the model estimated the posterior starting from a uniform distribution. All these assumptions are no more valid in the present paradigm since the type of trial (i.e., update, predictable and uniform) was not explicitly signaled by the color cue (all the targets were identical). Therefore, the new model estimated posterior distribution in the same way after each trial.

The absence of a cued transition between blocks implied that at a given trial *n* all the observations from 1 to *n* were available for estimating the posterior. This raised the question of the extent to which previous observations are used to

estimate. For example, it could be possible that *all* past events concur in the estimation, with an equal weight for distant and recent observations. However, several studies have shown that the brain integrates past experience accordingly to a temporal gradient over past observation (see: Harrison, Bestmann, Rosa, Penny, & Green, 2011). For instance, it has been shown that in the oddball paradigm, the recent trial history influences the P3b amplitude to a greater extent (Squires et al., 1976). Recently, Kolossa and colleagues implemented a model of trial-by-trial P3 fluctuations in a simple two-choice response time task (Kolossa et al., 2013), which takes into account both short-term and long-term memory decay processes. Since this model provided a superior account of P3b amplitude with respect to other previous P3b models, we supplied our model with those two short-term and long-term memory forgetting factors.

Let us call

$$P_{FP}(n) = p(FP \sim \mathcal{N}(\mu, \sigma) | FP_{1:n}), \tag{4.1}$$

the estimated foreperiod posterior probability over the parameter space $\mathcal{N}(\mu, \sigma)$ (see Section 2.2.3). on trial *n* (the parameter space had a size of 300×200, that is, the combination of all the means from 10 ms to 3000 ms and standard deviations from 10 ms to 2000 ms in steps of 10 ms). Based on Kolossa and colleagues (2013), in the current study the posterior was computed as follows:

$$P_{FP}(n) = \alpha_L \cdot P_{L,FP}(n) + \alpha_S \cdot P_{S,FP}(n), \tag{4.2}$$

that is the weighted sum of $P_{L,FP}$ and $P_{S,FP}$, which represent the posteriors estimated accordingly to long-term and short-term decaying memories about prior observations, respectively. The weighting parameters α_L and α_S represent the relative contributions of $P_{L,FP}$ and $P_{S,FP}$ in forming the posterior, and they hold $\alpha_S = 1 - \alpha_L$ and $0 \leq \alpha_S \leq 1$. The two posteriors $P_{L,FP}$ and $P_{S,FP}$ can be expressed as:

$$P_{i,FP}(n) = Pr_{i,FP}(n-1)^{\gamma_i} L_{i,FP}(n), i \in [L, S], \quad (4.3)$$

where $L_{i,FP}(n)$ is the likelihood $p\left(FP_n \middle| FP \sim \mathcal{N}(\mu, \sigma^2)\right)$ at trial *n*, and $Pr_{i,FP}(n-1)^{\gamma_i}$ is the power prior (Ibrahim, Chen, Gwon, & Chen, 2015) $p(FP \sim \mathcal{N}(\mu, \sigma^2) | FP_{1:n-1})$ raised to the power γ_i, which represents the exponential forgetting factor of the long-term (γ_L) or short-term (γ_S) decaying memory. From Kolossa and colleagues (2013), the short-term and long-term decay factors were respectively defined as

$$\gamma_S = e^{-\frac{1}{\beta_S}} \quad \text{e} \quad \gamma_L = e^{-\frac{1}{\beta_{L,n}}}. \quad (4.4)$$

Both forgetting factors are exponential, but only the long-term memory depends on the trial number as follows:

$$\beta_{L,n} = e^{-\left(\frac{1}{\tau_1} \cdot n + \frac{1}{\tau_2}\right)}. \quad (4.5)$$

This aspect determines that forgetting becomes much sharper when the number of trials increases (for a detailed description of the two exponential forgetting factors, we refer the reader to: Kolossa et al., 2013).

Posteriors were, then, translated from the parameter space over time, as follows:

$$P_{i,FP}(n) = \sum_{\mu_n, \sigma_n} \left(FP_n \middle| FP_n \sim \mathcal{N}(\mu_n, \sigma_n^2)\right), \quad (4.6)$$

Model-based measures of updating (D_{KL}) and surprise (I_S) were calculated as in section 2.2.4.

The values for the model parameters, αL, τ1, τ2, and βS, were identified by finding the parameter combination underlying the Bayesian observer who better explained RTs. Since the calculation of the model evidence for all possible combinations of parameters with a reasonable resolution is computationally too expensive, we adopted an iterative selection procedure as illustrated in Table 4.1. At each iteration, the ideal Bayesian observer from each parameter combination provided the two D_{KL} and I_S regressors that were used to estimate the following model (lme function in MATLAB):

$$\log(RT) \sim I_S * D_{KL} + (1 \mid ID). \quad (4.7)$$

The model with the combination of parameters (Table 4.1) that better explained RTs in terms of AIC was selected for the EEG analysis. Behavioral results for the best-fitting model are presented in Table 4.2.

As a final note, despite the temporal structure of the task was the same of the paradigm in the previous EEG study, the estimated foreperiod distributions in the current task were characterized by higher uncertainty (Fig. 4.1). This aspect led to a more gradual updating process (D_{KL}; see Fig. 4.1).

Table 4.1 | Ranges of free model parameters for each iteration and optimized parameters from the RTs best-fitting model.

	α_L	τ_1	τ_2	β_S
Iteration 1	0.1-0.9	10-100	0.1-1	1-10
Iteration 2	0.4-0.6	88.75-111.25	0.21-0.44	0.01-2.12
Iteration 3	0.47-0.52	102.81-108-44	0.35-0.41	0.27-0.80
Optimized parameters	**0.51**	**107.03**	**0.40**	**0.53**

Notes. At iteration 1 we found the best combination from a subset of parameters, which included 5 values per parameter linearly space in the indicated range. At subsequent iterations, for each parameter a new range of 5 values was centered on the best value found in the previous iteration and had a length equal to the distance between subsequent values in the previous iteration.

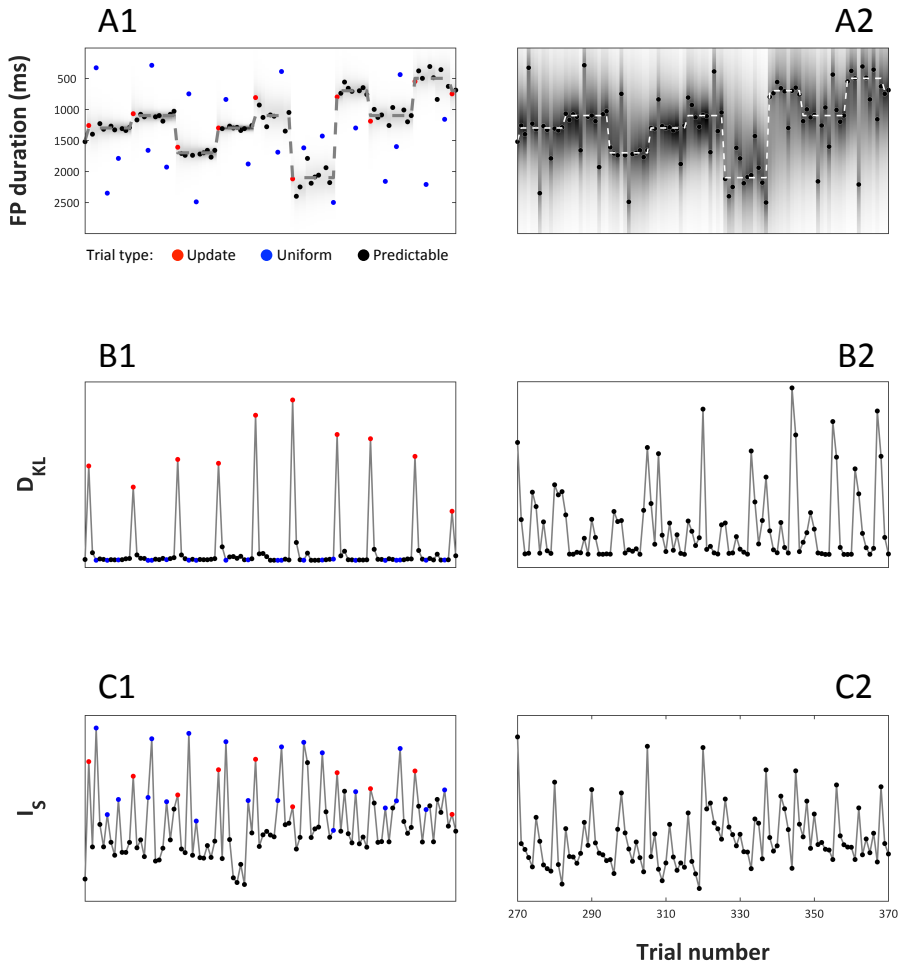

Figure 4.1 | Model and regressors of Study 2 (left column) and Study 3 (right column). All panels show the data from 100 trials. For Study 2, dot colors indicate trial types as in the legend. **(A1-2)** Plot of the state of the normative Bayesian learner. On the y axis is FP duration. The dashed line indicates the mean of the generative Gaussian distribution from which update and predictable trials 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). Note that the estimation in Study 3 was more uncertain (higher variance) than in Study 2. **(B1-2, C1-2)** Model-based regressors for updating (D_{KL}) and surprise (I_S).

Table 4.2 | Linear mixed-effects model fit by Maximum Likelihood for the final model

Fixed effects	**Estimate**	**SE**	**Den. *df***	***t***	***p-value***
Intercept	5.721	0.021	20621	268.03	<.001
IS	0.071	0.002	20621	31.73	<.001
DKL	-0.004	0.002	20621	-2.12	.034
IS*DKL	0.009	0.002	20621	5.83	<.001

4.2.4 EEG data analysis

EEG data acquisition and pre-processing were performed as described in the previous chapter (3.2.3).

Inferential statistics. First-level analysis was performed using the *Unfold* toolbox (Ehinger & Dimigen, 2018) following the same model specification of the previous EEG study (see Section 3.2.3). Second-level analysis was performed using the *ept-TFCE* toolbox (Mensen & Khatami, 2013) in MATLAB. Estimated D_{KL} and I_S parameters in the data space channels × epoch time points (0 - 1000 ms) were tested using a threshold-free cluster enhancement (TFCE) one-sample *t*-test (number of permutations = 200000, alpha-level = .001).

Between-study comparison. We further tested differences between Study 2 and Study 3 for updating (D_{KL}) and surprise(I_S) using a TFCE two-sample independent *t*-test (number of permutations = 200000, alpha-level = .001). We also tested for spatio-temporal regions of significant between-study equivalence for updating (D_{KL}) and surprise (I_S) using equivalence testing (Rogers, Howard, & Vessey, 1993; Schuirmann, 1987). In particular, we performed the so-called two one-sided (t-)tests for equivalence (TOST, see D. Lakens, 2017; Daniël Lakens, Scheel, & Isager, 2018; Montefinese, Ambrosini, & Roivainen, 2018). Although it is never statistically possible to conclusively show the absence of any effect, this approach allows to reject the presence of meaningful effects by testing whether the observed effect size for a non-significant test is close enough to zero (or, in other words, too small to be of practical importance or meaningful). In the current study, we set equivalence bounds at effects we had 80% power to reject, given our sample size and an alpha level of 0.05 (D. Lakens, 2017), that is, corresponding to an effect size *d* of .78.

4.3 Results

4.3.1 Electrophysiological results

The results of the TFCE analyses on surprise and updating in the current study are shown in Fig. 4.2 and 4.3, respectively.

Concerning updating (D_{KL}), as portrayed in Fig. 4.2 (warm color), there were two significant effects (Fig. 4.4A). The first significant modulation was a positive deflection emerging in the 130-320 ms time window, which was distributed over parietal electrodes (P1, P3, PO3, Cz, CPz, Pz, POz, CP2, P2). This modulation was followed by a second larger positive deflection emerging in the 480-900 ms time window again over parietal electrodes (CP1, P1, P3, CPz, Pz, POz, CP2).

Concerning surprise (I_S, Fig. 4.3), the first significant effect (Fig. 4.4B) was a positive modulation in the 260-490 ms time window over centro-frontal electrodes (F3, F1, FC5, FC3, FC1, C1, C5, CP1, CP3, P1, P3, Fz, FCz, Cz, CPz, F4, F2, FC4, FC2, C4, C2, CP2), which was surrounded by a negativity mainly located at lateral posterior electrodes (TP7, TP9, P7, PO7, O1, Oz, T8, TP8, TP10, P6, P8, PO8, O2). A second significant effect was a slow positive deflection (Fig. 4.4C) observed over frontal sites (Fp1, AF7, AF3, F7, F5, Fpz, Fp2, AF8, AF4, F8, F6, FT10, FT8, FC6), which started from 510 ms. Correspondingly, a negative modulation was observed over parietal channels (CP3, P1, P3, P5, PO3, CPz, Pz, POz, CP2, P2).

4.3.2 EEG results, between-study comparison

Concerning updating, Fig. 4.5 shows the results of the TFCE (left panel) and TOST (right panel) analyses for spatio-temporal regions of significant between-study differences and equivalence, respectively. The first significant difference (Fig. 4.7A-B) interested a first double-peak waveform that was enhanced in Study 2 (cold color). The difference was distributed over fronto-central electrodes (Fp1, AF7, AF3, F7, F5, F3, F1, FT9, FT7, FC5, FC3, FC1, C3, T7, C1, Fpz, Fz, F2, FC4, FC2) and lasted from 50 to 300 ms. The second difference regarded a positive deflection at parietal electrodes in the 200-320 time. As shown in Fig. 4.7C, this positive component was more

pronounced and longer lasting in Study 3 than in Study 2. Fig. 4.7C also shows in the 340-720 ms time interval, the significant equivalence of the parietal positive deflections (P3b-like) we found in both studies.

The results of the TFCE and TOST analyses on surprise (I_S) are portrayed in Fig. 4.6. A first difference emerged at posterior electrodes (Fig. 4.8A) in the 70-120 ms time window. This difference interested a positive deflection that was present only in Study 2. A second difference emerged at frontocentral electrodes (FC3, C1, C3,C5, CP2) around 300-450 ms. As shown in Fig. 4.8B, the positive deflection over fronto-central electrodes described above was present only in Study 3. A last difference emerged around 400-800 ms, over parieto-occipital channels (P1, P3, PO3, O1, CPz, Pz, POz, Oz, P2, P4, PO4, PO8). As shown in Fig. 4.8C, the difference interested a positive deflection that was higher and larger in Study 2.

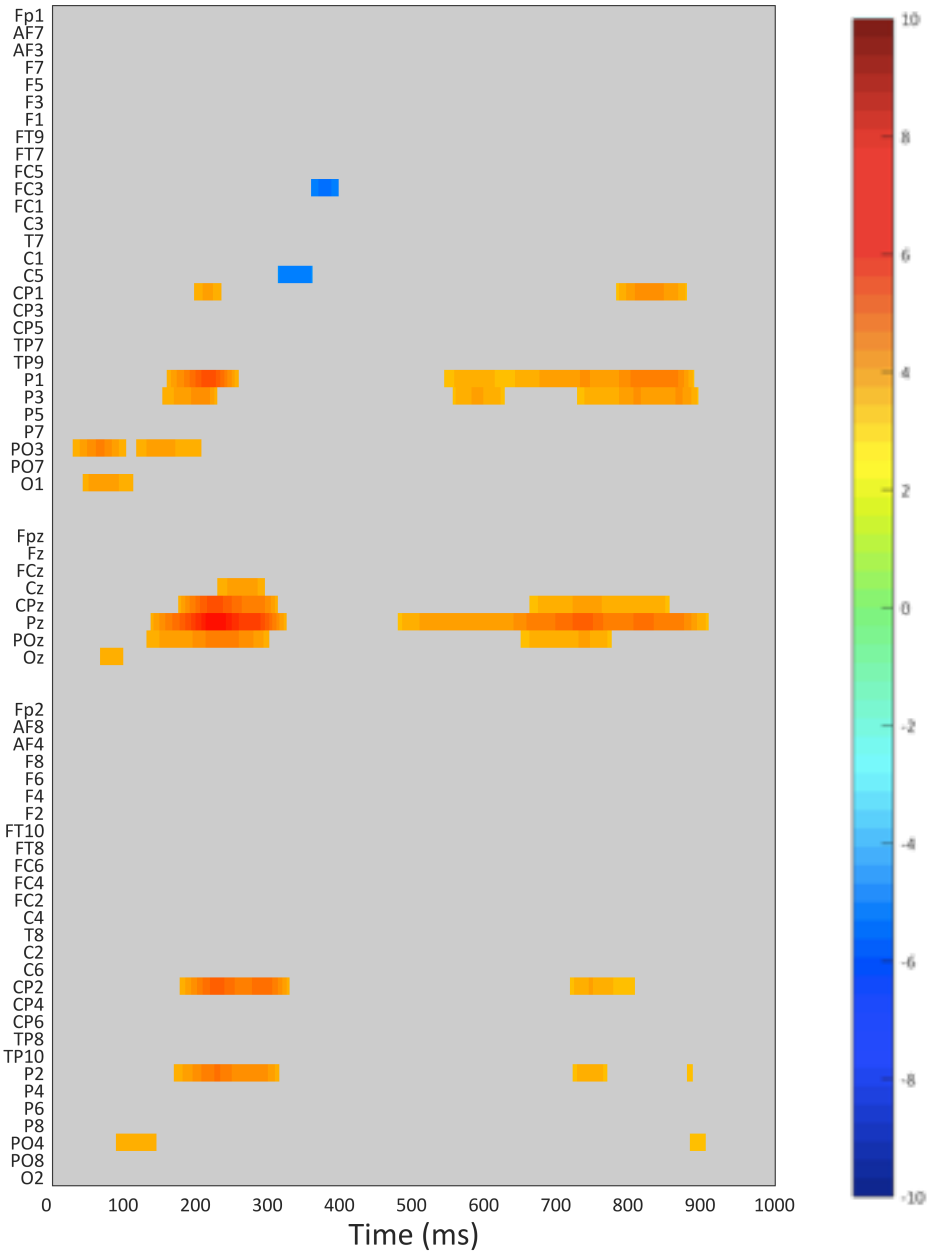

Figure 4.2 | Raster diagram showing significant effects elicited by updating (D_{KL}). For other conventions see Fig. 3.2

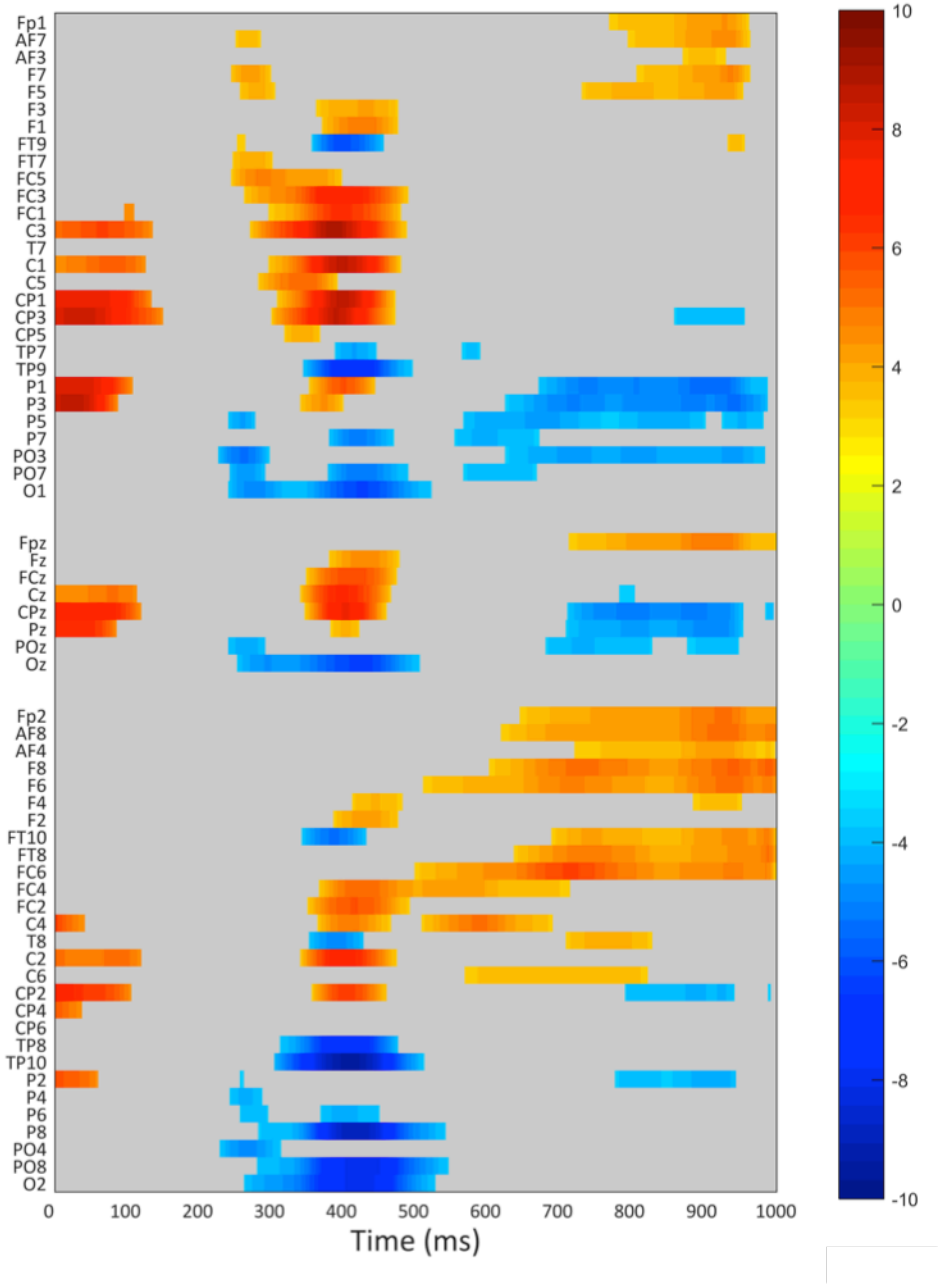

Figure 4.3 | Raster diagram showing significant effects elicited by surprise (I_S).

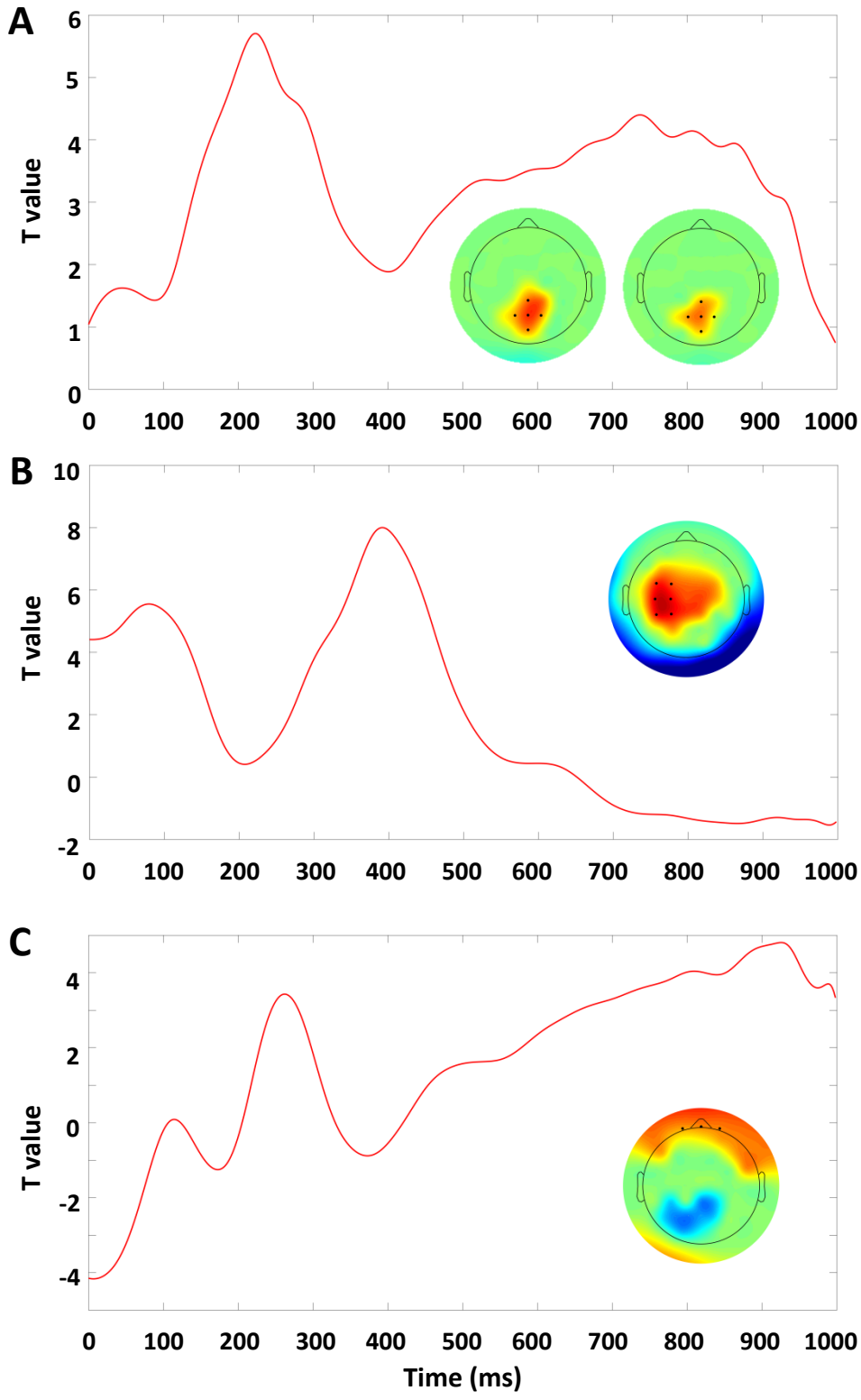

Figure 4.4 | Electrophysiological results. (A) The trace plot depicts the average *t* value for updating (D_{KL}) pooled over electrodes Cz, Cp1, Cpz, Cp2, Pz. The topographical maps show the *t* values averaged in the time windows 180-280 and 690-790 ms. **(B)** The trace plot depicts the average *t* value for surprise (I_S) pooled over electrodes FC3 FC1 C3 C1 CP3, CP1. The topographical map shows the *t* values averaged in the time window 350-450 ms. **(C)** The trace plot depicts the average *t* value for surprise (I_S) pooled over electrodes Fp1, Fpz, Fp2. The topographical map shows the *t* values averaged in the time window 850-950 ms. For other conventions, see Fig. 3.2.

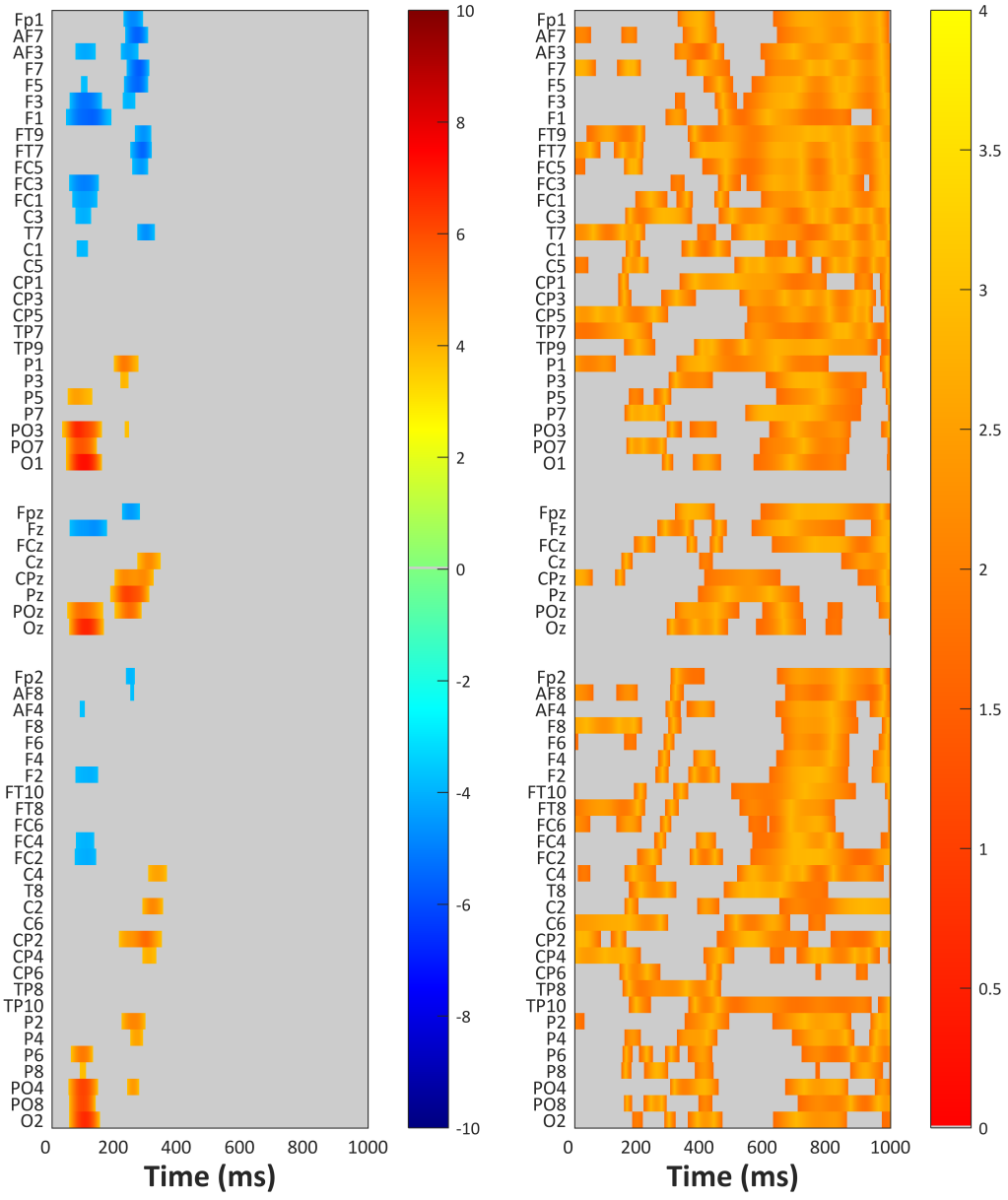

Figure 4.5 | Between-study comparison results: Updating (D_{KL}). Results from the TFCE two-sample independent *t*-test are shown in the **left panel**. Warm color indicates electrodes/time points in which the effect was significant more positive in Study 3 than in Study 2. Cold color indicates electrodes/time points in which the effect was significant more positive in Study 2 than in Study 3. Results from the TOST equivalence test are presented in the right panel. Warm color indicates electrodes/time points in which is possible to reject the presence of meaningful differences.

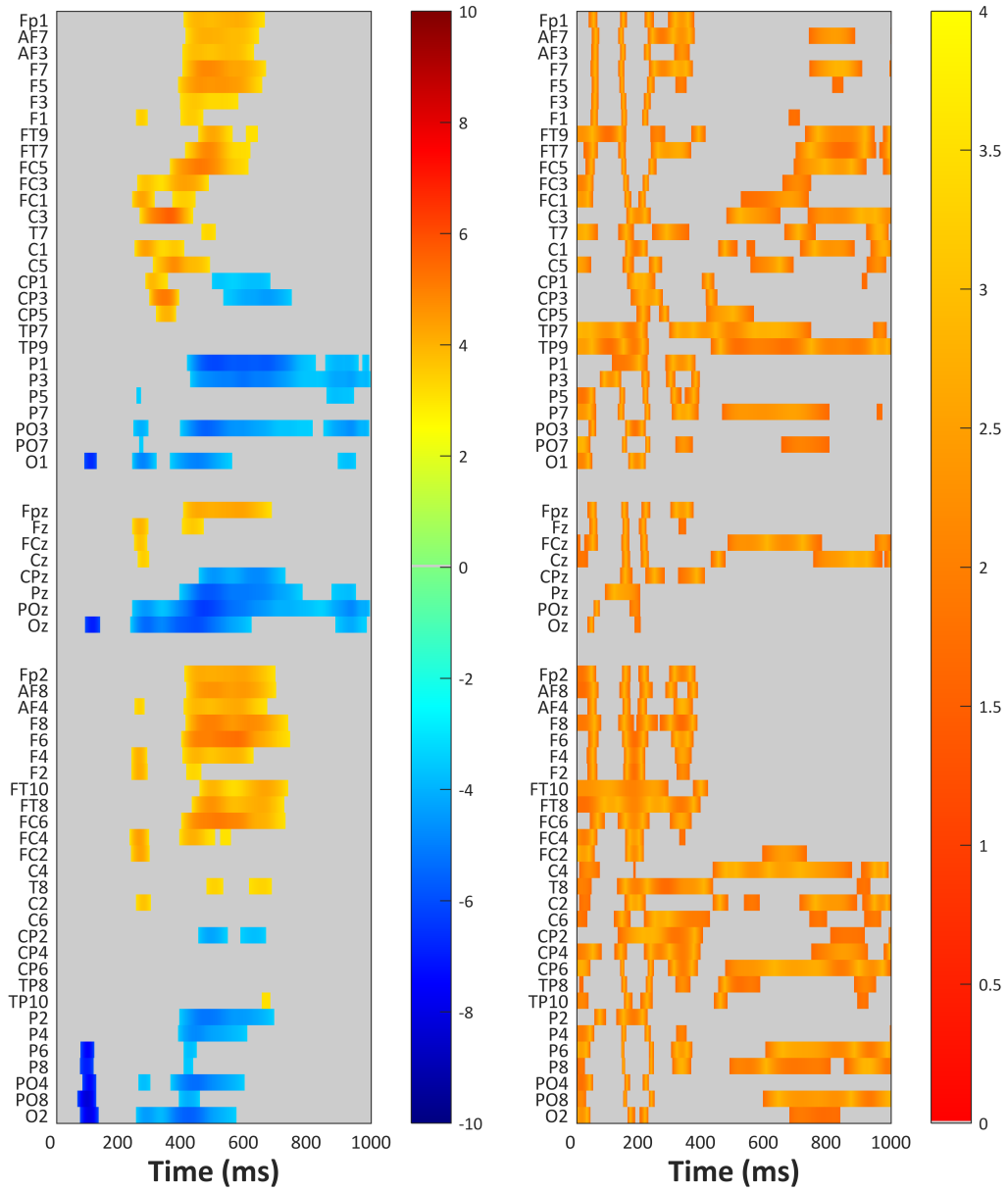

Figure 4.6 | Between-study comparison results: Surprise (I_S). Results from the TFCE (left panel) and TOST analyses.

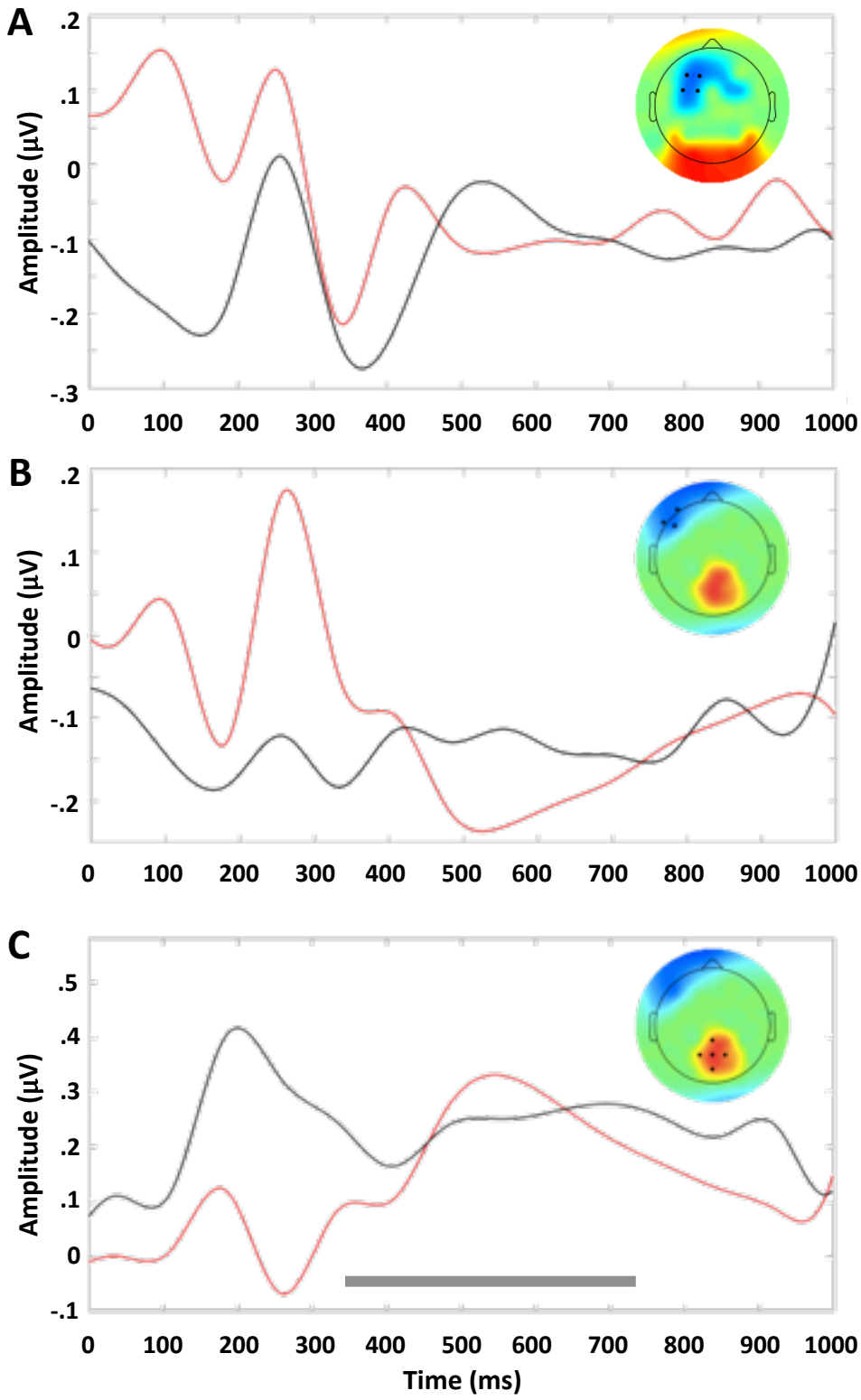

Figure 4.7 | Between-study electrophysiological differences: Updating (D_{KL}). The trace plots depict the mean effects (µV) of D_{KL} in Study 2 (red line) and in Study 3 (black line)**.** The topographical map shows the *t* values (mean effect differences between Study 3 – Study 2). **(A)** The trace plots depicts the mean effects pooled over electrodes F3, F1, FC1, FC3. The topographical map shows the *t* values averaged in the time window 80-120 ms. **(B)** The trace plots depict the mean effects pooled over electrodes AF7, F7,F5. The topographical map shows the *t* values averaged in the time window 240-280 ms. **(C)** The trace plots depicts the mean effects pooled over electrodes Cz, CP1, CPz, CP2, Pz. The topographical map shows the *t* values averaged between 240-280 ms. The gray bar indicates the equivalence interval for the P3b-like components.

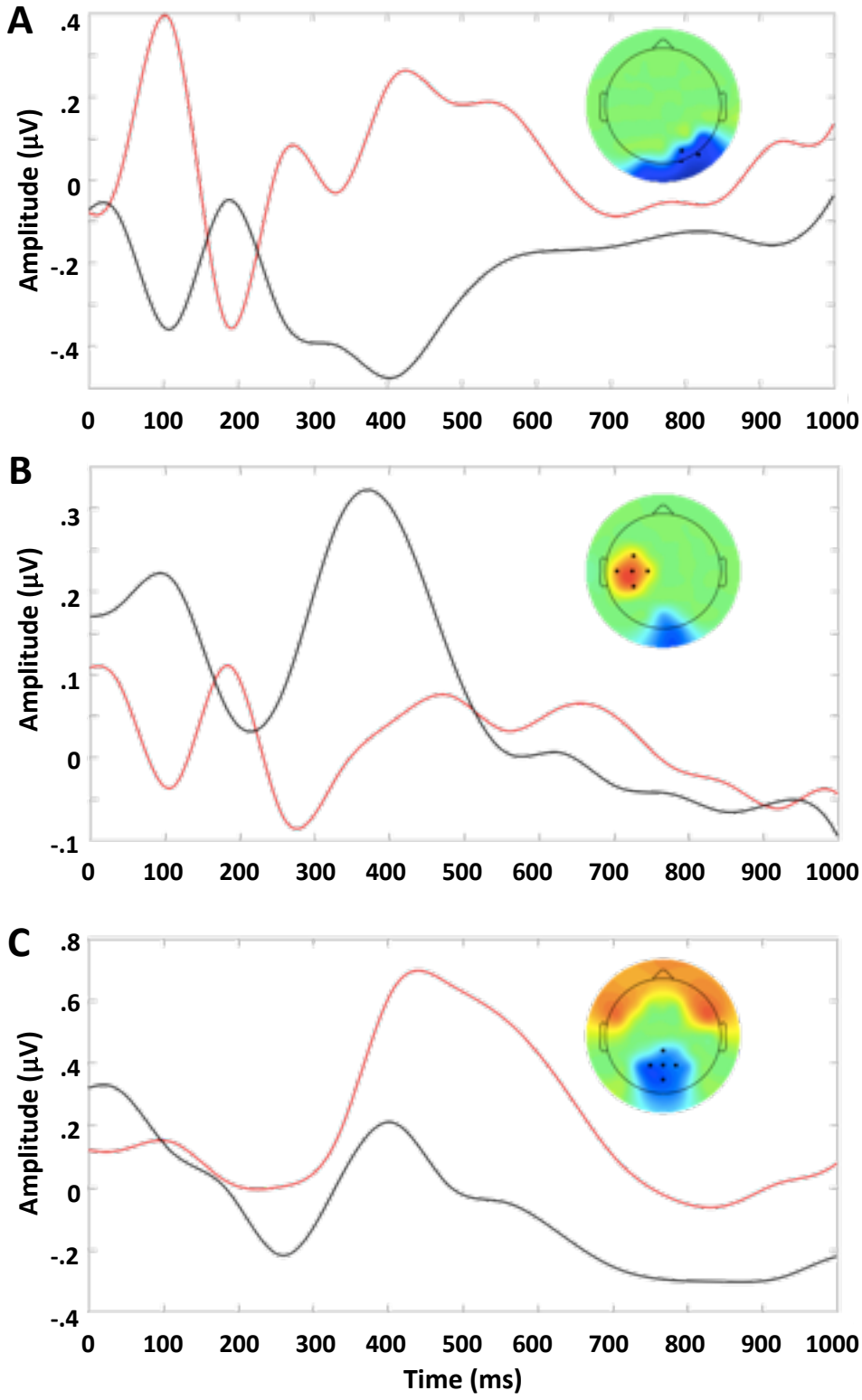

Figure 4.8 | Between-study electrophysiological differences: Surprise (I_S). For conventions see Fig. 4.7. **(A)** The trace plots depict the mean effects pooled over electrodes PO4, PO8, O2. The topographical map shows the *t* values averaged in the time window 80-120 ms. **(B)** The trace plots depict the mean effects pooled over electrodes FC3, C5, C3, CP3. The topographical map shows the *t* values averaged in the time window 340-380 ms. **(C)** The trace plots depict the mean effects pooled over electrodes Cz, CP1, CPz, CP2, Pz. The topographical map shows the *t* values averaged in the time window 450-550 ms.

4.4 Discussion

The main aim of the present EEG study was to investigate how surprise and updating are encoded when changes in the temporal probabilities of the events are implicitly inferred. As shown in Fig. 4.1B, the absence of the cue color manipulation used in our previous studies made the updating process more gradual over time by eliminating the abrupt shift in the estimated distribution. This aspect was critical to test implicit inferential processes and to draw a direct comparison with the explicit ones addressed in our previous study.

To summarize the main findings, surprise elicited a first fronto-central positivity and a later slower positivity at frontal sites. Updating was instead associated with a P2-like component followed by a P3-like modulation, both distributed over parietal sites. Overall, these results confirmed that, even without the explicit color change, it was still possible to distinguish surprise and updating at the electrophysiological level.

Beginning with our surprise-related findings, the analyses showed a significant positive modulation peaking at about 400 ms over fronto-central electrodes. According to its timing and topographical distribution (see Fig. 4.4B), such a component could be seen as belonging to the P3a waveform (Polich, 2003). Traditionally, the anterior P3a has been associated with rare and novel attention-capturing distractors across different tasks (e.g., oddball paradigm; Squires, Squires, & Hillyard, 1975). However, it has been shown that the P3a may be also modulated by rare target stimuli in the oddball paradigm (Spencer, Dien, & Donchin, 2001). All in all, these previous findings provide support for our P3a modulation by low-probable (surprising) target onset. It is worth mentioning that a surprise-related P3a was not found in our previous EEG study where, in contrast, surprising events elicited a more posterior and later positivity (labeled as P3b-like component). One difference between the two studies relies on the different amount of uncertainty experienced during the task, with more uncertain temporal expectations in the present study compared to the previous one. From a computational point of view, the increase in uncertainty is represented by the larger standard deviation in the prior (see Figure 4.1A). It follows, then, that here the general task context was

overall more demanding than the other EEG study. These key differences in the task structure are in favor of the claim that the anterior P3a and posterior P3b should not be seen as two distinct components but, rather, as the reflection of the activity of a common "multiple-demand" network (Duncan, 2013), which is differently distributed on the basis of the specific task demands (Barceló & Cooper, 2018). Of course, our assumption would benefit from further studies manipulating the degree of uncertainty in a parametric fashion. In this respect, some studies of belief updating emphasized a distinction between unexpected and expected uncertainty (Yu & Dayan, 2005). Expected uncertainty arises from a known unreliability of a stable environment. For example, in Study 2 expected uncertainty is due to the probabilistic nature of the causes of FP duration (i.e. the generative probability density function given by the weighted sum of a Gaussian distribution and a uniform distribution as expressed in Eq. 2.3). Unexpected uncertainty relates to strong violations of an internal model, which usually arises from unsignaled changes in the causes of the context that invalidates predictions based on previous observations and signals for a revision of the internal predictive model. In our studies, strong violations were caused by both update and uniform trials, but they likely led to different types of uncertainty depending on the presence/absence of the color manipulation. In Study 2, the changes in the underlying FP distribution were explicitly signaled, thus, it is unlikely they induced uncertainty beyond the one caused by the stochasticity inherent in the context (expected uncertainty). Also violations in uniform trials, since signaled, were "expected", thus, contributing to the unreliability of the context. In contrast, the absence of the color manipulation in Study 3 might have led to unexpected uncertainty after update and uniform trials. This suggests that the two studies might differ in the degree of unexpected uncertainty. However, the high context volatility in our tasks, due to the frequent changes in the underlying FP distributions, might have induced participants to also internally represent the rate of context change (Behrens et al., 2007). This implies that in Study 3 uncertainty due to changes in the FP distribution might have been also expected and that the differences between the two studies might regard differences in the level of expected uncertainty.

The second significant modulation elicited by surprise was a late and slow positive potential emerging around 650 ms over frontal electrodes. It is quite difficult to attribute a specific functional meaning to such a component since, to our knowledge, it has been usually reported in fields far away from our topic (e.g., late positive potential in emotion and affective processing, Brown, van Steenbergen, Band, de Rover, & Nieuwenhuis, 2012). For example, a larger post-N400 frontal positivity (pN400FP) has been observed in response to unfulfilled expected sentence continuations (DeLong, Urbach, Groppe, & Kutas, 2011). Accordingly, it has been proposed that the pN400FP could index the violation of word predictions in a sentence (Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007). Drawing a parallelism with our task, it is tempting to extrapolate some similarities between these language studies and the present one in the sense that in our task surprising events violated an expected (temporal rather than verbal-semantic) prediction as well. At any rate, we are aware that our interpretation is highly speculative and hence it should be taken with great caution also in light of the fact that a slow frontal positivity was not found in our previous study.

As a final remark on surprise, here we did not observe any significant modulation of the occipital P1 as reported previously. The significant differences in this component between the two studies further suggest that the P1 in Study 2 was likely related to the cue color manipulation and not to the process under study (see Discussion of Chapter 3).

Concerning updating, the first significant modulation was a P2-like component arising around 250 ms over parietal sites. As mentioned in the previous study (see Discussion of Chapter 3), the involvement of this potential for updating of temporal expectations fits well with previous findings showing that the P2 amplitude reflects the distance between standard and comparison intervals in temporal discrimination tasks and is not affected by the hazard rate of elapsed time (Kononowicz & van Rijn, 2014). It is noteworthy that in the current paradigm the updating-related P2 effect was significantly higher and longer lasting than the previously found P2. We suggest that the P2 indexes the computation of the posterior. In the previous study, after update trials, the posterior temporal

expectation was computed starting from a uniform distribution since participants were explicitly instructed about the transition from one block to the other one (O'Reilly et al, 2013). This means that they could disregard the temporal expectations built up on the previous distribution to derive the posterior. Conversely, in the current study the posterior was continuously computed starting from the prior. This more complex computation of the posterior might have been captured by larger P2 amplitude.

As in our previous study, we again found a significant P3b modulation associated with updating. For most of the involved time points and electrodes, there was no substantial difference in the P3b between the two studies, which suggests that the same updating processes occurred for both explicit and implicit temporal inferences. This finding strongly supports the view that the P3b represents a key electrophysiological correlate of updating of temporal expectations in general.

Another feature of the present study results is that we did not find significant modulations by updating of the early potentials, namely, the frontal double-peak component and the occipital P2, observed in the former study. Only the frontal double-peak component was significantly different between the two studies. Taking into account that such components have been formerly related to top-down perceptual stimulus evaluation (Berchicci, Spinelli, & Di Russo, 2016; Di Russo et al., 2017), it is possible that they were involved in the processing of the explicit cue.

In our studies we focused on processes associated with target processing by analyzing deconvolved stimulus-locked ERPs. However, by using recent trial-by-trial decomposition techniques (i.e., residue iteration decomposition, RIDE; Ouyang, Sommer, & Zhou, 2015), recent studies (Brydges & Barcelo, 2018; Verleger, Grauhan, & Smigasiewicz, 2016) have shown distinct target P3-like potentials that are time-locked not only to the stimulus, but also to the response or to neither of the two (i.e., capturing trial-by-trial latency variability in neural activity). Therefore, it could be interesting to adopt this approach to disentangle updating- and surprise-related target P3-like components in the form of stimulus-locked, response-locked or non-phase locked EEG response.

Before concluding, it should be acknowledged that the model that better explained participants' RTs included both short-term and long-term memory decay as compared to a model that did not take into account any recency effect. This finding lends support to Kolossa and colleagues (2013), who showed that the P3 amplitude is indeed better accounted for by a model that considers both short-term and long-term memory decay.

To sum up, in the present study we isolated the electrophysiological correlates of surprise and updating by highlighting both commonalities and differences in implicit and explicit temporal inferential processes.

Chapter 5

General discussion

The present project tackled an aspect of temporal preparation that has been often raised in the literature but never directly investigated so far, that is, the formation and revision of prior temporal expectations. Specifically, our aim was mainly twofold. The general aim was to identify the neural and electrophysiological correlates of both explicit and implicit belief updating about the time of occurrence of an event. Considering that belief updating takes places prominently after events violating our prior expectations, a second related aim was to disentangle processes involved in updating from those merely responding to surprising events. To the best of our knowledge, there are no previous studies that have investigated these issues in the field of temporal preparation.

To address our research questions, we developed a foreperiod temporal preparation task in which the generative foreperiod distribution was non-stationary throughout the task. This implies that participants had to constantly update beliefs in order to speed up response to target onset. To investigate explicit processes, in the first two studies we manipulated the color of the target in order to decompose updating and surprise. By contrast, to explore implicit processes, in our last study we got rid of the color manipulation. Despite the absence of the color, it was still possible to differentiate updating and surprise because of the temporal nature of

our task. Indeed, we measured updating and surprise based on two different probabilistic processes. More in detail, to characterize updating of prior temporal expectations, we developed an ideal observer that quantitatively described trial-by-trial Bayesian belief updating. At each trial, updating was quantified as the Kullbach-Liebler divergence (D_{KL}) between prior and posterior. Surprise was measured as the Shannon's information (I_S), which represents the violation of participants' expectations according to the hazard function (i.e., the probability that an event will occur given that it has not yet occurred). Hence, while updating (D_{KL}) was associated with Bayesian inference, surprise (I_S) was related to hazard inference.

In what follows we shall provide a general overview of the main findings obtained in the present project together with a consideration of both values and challenges that, we hope, will stimulate more exciting work in this field of study.

The fMRI findings showed both common and differential involvement of two cognitive control networks for updating and surprise: the fronto-parietal network (FPN; Dosenbach et al., 2008) and the cingulo-opercular network (CON; Dosenbach et al., 2008; Menon, 2015). Concerning updating, it modulated activity and functional connectivity in regions belonging to the FPN, namely bilateral lateral frontal cortex, bilateral posterior parietal cortex, posterior cingulate cortex, and precuneus. The reliability of such results is supported by the high concordance with previous fMRI studies that have investigated belief updating (e.g., Gläscher, Daw, Dayan, & O'Doherty, 2010; Kobayashi & Hsu, 2017; Schwartenbeck et al., 2016; Waskom et al., 2017). As an example, Waskom and colleagues (2017) reported that inferior frontal sulcus, intra-parietal sulcus (IPS), precunesus and posterior cingulate cortex (PCC) responded to prediction error in a context-dependent perceptual decision making task. Although prediction error is a measure more similar to I_S than D_{KL}, our results make it unlikely that their findings reflected surprise alone. Further support for our findings comes from a similar involvement of the FPN in belief updating in those tasks that have decomposed surprise and updating (Kobayashi & Hsu, 2017; Schwartenbeck et al., 2016). However, there is also evidence that argues against our results on belief updating (O'Reilly et al., 2013; Vossel et al., 2015). In particular, O'Reilly and colleagues (2013) found that updating was mainly located to

ACC (an area also reported in Schwartenbeck et al., 2016). The discrepancy between O'Reilly and colleagues and our study could seem quite counterintuitive considering that our paradigm was modeled after their task. Yet, we believe that such a difference is particularly telling in the light of the fact that they used a saccadic planning task requiring visuo-motor learning and saccadic preparation. This may imply that belief updating in space and time could rely on different brain areas, lending support to the hypothesis that predictions for *where* and *when* are functionally different (Coull & Nobre, 1998). Future studies that manipulate updating in the two dimensions during the same task are necessary to provide direct evidence on this question.

Our fMRI results on surprise showed the involvement of regions belonging to the CON, including bilateral insula, dorsal ACC and pre-SMA. Such areas are coherent with two out of the three studies that separated updating and surprise (Kobayashi & Hsu, 2017; Schwartenbeck et al., 2016). It is worth noting that these two studies were also in line with our results on updating. Our localization of surprise in the CON is consistent with previous studies reporting that the CON responds to salient stimuli (Menon & Uddin, 2010).

In order to integrate our fMRI findings with previous literature on belief updating and violation of expectation, a tantalizing speculation would be that updating and surprise might be at the core of the dissociation between FPN and CON. If this were true, our hypothesis could provide a valuable key to further understand the functional role that these two networks play in cognitive control.

Turning back to previous fMRI studies tracking temporal hazard, our study lends support to the involvement of sensory areas, which can be modulated by attention deployed in anticipation of a forthcoming stimulus (Bueti et al., 2010; Bueti & Macaluso, 2010; Vallesi, McIntosh, Shallice, et al., 2009). Regarding the pivotal role that has been attributed so far to the right prefrontal cortex in monitoring temporal contingencies (Coull et al., 2016; Vallesi, McIntosh, Shallice, et al., 2009), our findings showed that such a region responded to both violation of expectations and updating of prior beliefs. This suggests that its relation to the hazard function might deal more with the detection and resolution of expectancy

violation than with the simple tracking of the passage of time. However, it should be noted that since our analyses were target-locked, we cannot draw any conclusion on the (monitoring) role of the right prefrontal cortex during the course of the foreperiod. The same reasoning applies to other studies that have investigated temporal preparation by looking at the BOLD activity time-locked to the cue stimulus (Cotti et al., 2011; Coull & Nobre, 1998; Davranche et al., 2011). In this regard, our study critically adds to this previous literature not only by showing how prior temporal expectations are updated but also by providing a finer-grained analysis of the processes that take place at target onset.

Once identified the neural correlates of surprise and updating, which supports the success of our color manipulation in separating the two, the next step was to investigate the temporal dynamics associated with surprise and updating. In doing so, we were inspired by the literature on the P3 and the Bayesian brain hypothesis (Bennett et al., 2015; Kolossa et al., 2015; Kopp, 2008; Kopp et al., 2016b; Mars et al., 2008; Seer, Lange, Boos, Dengler, & Kopp, 2016). Importantly, however, we also went beyond the P3 by exploiting recent advances in computer power and statistics to analyze the EEG data. Specifically, we used a method that combined a mass-univariate approach with deconvolution (Ehinger & Dimigen, 2018). This approach gave the following advantages over a more traditional EEG approach: first, it allowed us to isolate specific ERPs associated with target onset and to characterize the specific contribution of updating and surprise; moreover, it allowed fully exploiting the spatial and temporal resolution of the EEG that enabled us to explore non-expected effects and, at the same time, to provide a more defined and comprehensive picture of a priori expected components (i.e., P3).

In the first EEG study, we found that surprise and updating differed in that while the former only elicited two modulations (i.e., posterior P1-like and parietal P3b-like waveforms), the latter was associated with a more complex pattern (i.e., frontal double-peak, parietal P2-like and occipital P2-like earlier waveforms, and later P3b-like component). What is particularly interesting here is that we corroborated previous literature on the P3 family (Kolossa et al., 2013; Kolossa et al., 2015; Mars et al., 2008) by showing that specific P3-like components were

selectively associated with surprise and updating. Importantly, we extended these previous studies by isolating the specific contribution of each process.

Adding to the fMRI findings, the EEG data give us some hints of the time course and the scalp topography of the P3-like waveforms modulated by both surprise and updating. These two components had similar scalp distribution, but the P3-like component elicited by surprise occurred earlier and lasted shorter than the P3 elicited by updating. We might surmise that the similar topography was due to similar P3 cortical generators. In our fMRI study we found, indeed, some overlapping regions that were modulated by both updating and surprise, and which were mainly located in the PPC. This speculation finds plausible support from several studies that, by using fMRI-constrained EEG source analysis (Bledowski et al., 2004; Crottaz-Herbette & Menon, 2006; Horovitz, Skudlarski, & Gore, 2002), have shown that the PPC is a generator of the parietal P3. Concerning the difference in the time course, this might also find anatomical support in our fMRI study that identified the insula as a critical region involved in surprise. Indeed, Menon and Uddin (2010) proposed a model of attentional control in which the AI sends rapid signals through von Economo neurons (neurons with large axons which facilitate rapid broadcasting of the signal; Nimchinsky et al., 1999) to areas generating the scalp-recorded P3, including the PPC (see also Sridharan, Levitin, & Menon, 2008). In our fMRI study, we also found strong connectivity between right AI and PPC, which could suggest that the early onset of the surprise-related P3b was due to the rapid trasmission of AI signals to the PPC. These speculations would largely benefit from source reconstruction of the EEG data, albeit we are aware that a direct comparison between fMRI and source analysis should be considered with caution. In any case, source reconstruction of the EEG data could help clarify the functional meaning of our ERPs and represents a future research step that slipped off the present thesis due to time constraints.

To sum up the EEG results of Study 2, a conclusion is that surprise and updating can be also separated in terms of electrophysiological signatures. At this point, we asked whether our results truly reflected Bayesian inference or were reflecting a sort of all-or-none shift in beliefs (Nassar, Wilson, Heasly, & Gold, 2010)

caused by our explicit manipulation. Answering this question was critical to substantiate our conclusions. For this aim, we designed a more implicit task that kept the same general structure as the explicit one but that removed any explicit information about its probabilistic nature. In any case, as stated above, we were confident about the fact that we were still differentiating between updating and surprise, since the respective measures, D_{KL} and I_S, were calculated on probabilistic information of two different inferential processes (i.e., Bayesian and hazard). Confirming our hypotheses, we found that surprise and updating could be again differentiated. In this study, surprise elicited a positive modulation that according to its timing and topography can be conceived as a P3a-like component. Although this result was consistent with previous findings showing that the P3a is elicited by rare surprising stimuli (see Polich, 2003), it differed from the results from study 2. A possible explanation for this difference could be that in the last study the environment was likely experienced as more uncertain and volatile, thus making the task more demanding. According to recent views that see the P3-family as an index of the activity from the multiple demand network (a super network including both FPN and CON; Crittenden et al., 2016), which is more frontally distributed with increased task demands (Barceló, Periáñez, & Nyhus, 2008). Hence, we could speculate that the P3s in the two studies might reflect similar brain activity differently modulated in response to uncertainty. The comparison of the source reconstructed EEG data from the two studies could help to gain more insight on this issue.

As regards updating, we corroborated our previous findings by showing that a later P3b was modulated by implicitly driven updating of temporal expectations. Moreover, we provided substantial evidence for the absence of differences in P3b modulations by implicitly and explicitly driven inference. These results extend the literature on the P3 by showing that this component was truly reflecting Bayesian belief updating.

Moreover, in both studies we found that the P3b was preceded by a parietal P2-like component. However, in the last study this modulation was significantly greater with respect to the P2 found in study 2. Since this component has been

previously implicated in comparison processes involved in temporal discrimination (Kononowicz & van Rijn, 2014), here we suggest that it might reflect an index of the posterior computation. This explanation stems from the fact that in the last study posterior estimation was computationally more demanding. Indeed, in study 2, posterior computation was easier since it could be optimally derived from a uniform prior distribution. In study 3, posterior estimation required a more complex integration of prior and likelihood. It follows that the enhanced P2 modulation in study 3 might reflect this increased computational complexity.

In sum, in study 3 we were able to differentiate update and surprise even in absence of a strong explicit manipulation that allowed having surprise-only events. Moreover, by integrating findings from the two EEG studies, we discovered two EEG indexes of the inferential process underlying updating of prior temporal expectations, which respond to both explicit and implicit contextual changes.

At this point it should be noted that the use of an ideal Bayesian observer represents "just a description of optimal behavior: it does not prescribe how Bayes optimal perception, sensorimotor integration or decision-making under uncertainty emerges" (Friston, 2012). To attempt to understand *how* the brain updates temporal expectations, we need more sophisticated models that take into account constraints deriving from the anatomy and the physiology of the brain, such as its functional hierarchical architecture or neuronal dynamics. In this respect, our project should be considered as a pioneering attempt that should stimulate the implementation of further models to understand how the brain refines temporal expectations.

Related to this point, another remaining question that has to be addressed in the future is the role of neuromodulatory systems in updating of temporal expectations. In this regard, a recent study provided evidence for a role of dopamine (DA) in regulating the precision with which humans track the temporal hazard (Tomassini, Ruge, Galea, Penny, & Bestmann, 2016). At the same time, other studies have shown the involvement of catecholamine systems, including DA and norepinephrine (NE), in belief updating (Jepma et al., 2018; Jepma et al., 2016). In the light of these findings, our paradigm might help characterize the role of DA in "Bayesian" and hazard inference. More generally, our paradigm might make a

significant contribution in understanding the specific influence that the different neuromodulatory systems have on belief updating. For example, in our EEG studies, we found that P3-like modulations were sensitive indexes of Bayesian inference in different situations (e.g. in presence of evident or non-evident changes in the environment, or in higher or lower context uncertainty) and allow distinguishing between updating and surprise. In light of the association between P3 and catecholamine (i.e., DA and NE) systems (Nieuwenhuis, Aston-Jones, & Cohen, 2005), our task represents a promising tool to better characterize the involvement of these neuromodulatory systems in updating. Moreover, the simplicity of our paradigm makes it an excellent tool to explore these and other related research questions on (temporal) belief updating in patients and in animals.

To conclude, the present dissertation provides the first characterization of the cognitive brain processes involved in temporal belief updating. Considering that temporal expectations are a fundamental feature of cognitive brain functions, we hope that our work will stimulate future work to provide a more exhaustive answer to an overarching question that still puzzles us: how does the brain exploit temporal expectations to optimize behavior?

Acknowledgements

I would like to express my sincere gratitude to my supervisor Prof. Antonino Vallesi for the support to my PhD studies, for his patience, motivation, understanding and encouragement over the past three years.

My profound gratitude goes to Dr. Ettore Ambrosini, the best mentor I have ever had. Thank you for guiding me on the right path and for your friendship. I will always be thankful to you.

I'm immensely grateful to Dr. Mariagrazia Capizzi. Thank you for your special friendship and for accompanying me throughout this PhD project. I couldn't have done it without your help. Thank you again.

I need to thank all my Lex-Mea mates. I cannot begin to express my gratitude and appreciation for their friendship.

My sincere thanks also go to Prof. Kopp. I am indebted to him for sharing his expertise, and for his sincere and valuable guidance and encouragement.

I acknowledge the support of the European Research Council, which has largely funded the studies in the present thesis and my PhD scholarship through the Starting Grant LEX-MEA n° 313692 (FP7/2007-2013) to Prof. Antonino Vallesi. I also acknowledge the support of the International Travel Grant awarded by the Boehringer Ingelheim Fonds that funded my research period in the Lab of Prof. Kopp.

Mostly importantly, none of this could have happened without my family and without Laila. I love you.

References

Andersson, J. L., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage, 20*(2), 870-888.

Ashburner, J., Barnes, G., Chen, C., Daunizeau, J., Flandin, G., Friston, K., . . . Moran, R. (2014). SPM12 manual. *Wellcome Trust Centre for Neuroimaging, London, UK*.

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3*(2), 12-28.

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

Barceló, F., Perianez, J. A., & Nyhus, E. (2008). An information theoretical approach to task-switching: evidence from cognitive brain potentials in humans. *Front Hum Neurosci, 1*, 13.

Barceló, F., & Cooper, P. S. (2018). An information theory account of late frontoparietal ERP positivities in cognitive control. *Psychophysiology, 55*(3).

Barto, A., Mirolli, M., & Baldassarre, G. (2013). Novelty or surprise? *Front Psychol, 4*, 907.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Bausenhart, K. M., Rolke, B., & Ulrich, R. (2008). Temporal preparation improves temporal resolution: evidence from constant foreperiods. *Percept Psychophys, 70*(8), 1504-1514.

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nat Neurosci, 10*(9), 1214-1221.

Bennett, D., Murawski, C., & Bode, S. (2015). Single-Trial Event-Related Potential Correlates of Belief Updating. *eNeuro, 2*(5), ENEURO.0076-15.2015.

Berchicci, M., Spinelli, D., & Di Russo, F. (2016). New insights into old waves. Matching stimulus- and response-locked ERPs on the same time-window. *Biol Psychol, 117*, 202-215.

Bledowski, C., Prvulovic, D., Hoechstetter, K., Scherg, M., Wibral, M., Goebel, R., & Linden, D. E. (2004). Localizing P300 generators in visual target and distractor processing: a combined event-related potential and functional magnetic resonance imaging study. *J Neurosci, 24*(42), 9353-9360.

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial vision, 10*, 433-436.

Brown, S., van Steenbergen, H., Band, G. P. H., de Rover, M., & Nieuwenhuis, S. (2012). Functional significance of the emotion-related late positive potential. *Front Hum Neurosci, 6*.

Brydges, C. R., & Barceló, F. (2018). Functional Dissociation of Latency-Variable, Stimulus- and Response-Locked Target P3 Sub-components in Task-Switching. Front Hum Neurosci, 12, 60.

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

Bueti, D., & Macaluso, E. (2010). Auditory temporal expectations modulate activity in visual cortex. *Neuroimage, 51*(3), 1168-1183.

Chater, N., & Manning, C. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences, 10*(7), 335-344.

Chaumon, M., Bishop, D. V., & Busch, N. A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *J Neurosci Methods, 250*, 47-63.

Chiu, Y. C., & Yantis, S. (2009). A domain-independent source of cognitive control for task sets: shifting spatial attention and switching categorization rules. *J Neurosci, 29*(12), 3930-3938.

Cocchi, L., Zalesky, A., Fornito, A., & Mattingley, J. B. (2013). Dynamic cooperation and competition between brain systems during cognitive control. *Trends in cognitive sciences, 17*(10), 493-501.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience, 3*(3), 201.

Correa, A. (2010). Enhancing behavioural performance by visual temporal orienting. *Attention and time*, 357-370.

Correa, A., Lupianez, J., Madrid, E., & Tudela, P. (2006). Temporal attention enhances early visual processing: a review and new evidence from event-related potentials. *Brain Res, 1076*(1), 116-128.

Cotti, J., Rohenkohl, G., Stokes, M., Nobre, A. C., & Coull, J. T. (2011). Functionally dissociating temporal and motor components of response preparation in left intraparietal sulcus. NeuroImage, 54(2), 1221-30.

Coull, J. T. (2009). Neural substrates of mounting temporal expectation. *PLoS biology, 7*(8), e1000166.

Coull, J. T., Cotti, J., & Vidal, F. (2016). Differential roles for parietal and frontal cortices in fixed versus evolving temporal expectations: Dissociating prior from posterior temporal probabilities with fMRI. *Neuroimage, 141*, 40-51.

Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *J Neurosci, 18*(18), 7426-7435.

Crittenden, B. M., Mitchell, D. J., & Duncan, J. (2016). Task encoding across the multiple demand cortex is consistent with a frontoparietal and cingulo-opercular dual networks distinction. *Journal of Neuroscience, 36*(23), 6147-6155.

Crottaz-Herbette, S., & Menon, V. (2006). Where and when the anterior cingulate cortex modulates attentional response: combined fMRI and ERP evidence. *J Cogn Neurosci, 18*(5), 766-780.

Davranche, K., Nazarian, B., Vidal, F., & Coull, J. T. (2011). Orienting attention in time activates left intraparietal sulcus for both perceptual and motor task goals. *Journal of cognitive neuroscience, 23*(11), 3318-3330.

DeLong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology, 48*(9), 1203-1207.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods, 134*(1), 9-21.

Di Russo, F., Berchicci, M., Bozzacchi, C., Perri, R., Pitzalis, S., & Spinelli, D. (2017). Beyond the "Bereitschaftspotential": action preparation behind cognitive functions. *Neuroscience & Biobehavioral Reviews, 78*, 57-81.

Donchin, E. (1979). Event-related brain potentials: A tool in the study of human information processing. In *Evoked brain potentials and behavior* (pp. 13-88): Springer.

Donchin, E. (1981). Surprise!… surprise? *Psychophysiology, 18*(5), 493-513.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences, 11*(3), 357-374.

Dosenbach, N. U., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends Cogn Sci, 12*(3), 99-105.

Doya, K., Ishii, S., Pouget, A., & Rao, R. P. (2007). *Bayesian brain: Probabilistic approaches to neural coding*: MIT press.

Duncan, J. (2013). The structure of cognition: attentional episodes in mind and brain. *Neuron, 80*(1), 35-50.

Ehinger, B. V., & Dimigen, O. (2018). Unfold: An integrated toolbox for overlap correction, non-linear modeling, and regression-based EEG analysis. *bioRxiv 360156*.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Res, 1146*, 75-84.

Forder, L., He, X., & Franklin, A. (2017). Colour categories are reflected in sensory stages of colour perception when stimulus issues are resolved. PloS One, 12(5), e0178097.

Fornito, A., Harrison, B. J., Zalesky, A., & Simons, J. S. (2012). Competitive and cooperative dynamics of large-scale brain functional networks supporting recollection. *Proceedings of the National Academy of Sciences, 109*(31), 12788-12793.

Freunberger, R., Klimesch, W., Doppelmayr, M., & Holler, Y. (2007). Visual P2 component is related to theta phase-locking. *Neurosci Lett, 426*(3), 181-186.

Friston, K. (2005). A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci, 360*(1456), 815-836.

Friston, K. (2012). The history of the future of the Bayesian brain. In *Neuroimage* (Vol. 62-248, pp. 1230-1233).

Friston, K., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and Modulatory Interactions in Neuroimaging. *NeuroImage, 3*(6), 218-229.

Geisler, W. S. (2011). Contributions of Ideal Observer Theory to Vision Research. *Vision Res, 51*(7), 771-781.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological review, 84*(3), 279-325.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., . . . Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage, 80*, 105-124.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus Rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron, 66*(4), 585-595.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology, 48*(12), 1711-1725.

Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models–The R Package pbkrtest. *Journal of Statistical Software, 59*(9), 1-32.

Harrison, L. M., Bestmann, S., Rosa, M. J., Penny, W., & Green, G. G. R. (2011). Time Scales of Representation in the Human Brain: Weighing Past Information to Predict Future Events. *Front Hum Neurosci, 5*.

Hayden, B. Y., Nair, A. C., McCoy, A. N., & Platt, M. L. (2008). Posterior cingulate cortex mediates outcome-contingent allocation of behavior. *Neuron, 60*(1), 19-25.

Hayden, B. Y., Smith, D. V., & Platt, M. L. (2010). Cognitive control signals in posterior cingulate cortex. *Front Hum Neurosci, 4*, 223.

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.

Horovitz, S. G., Skudlarski, P., & Gore, J. C. (2002). Correlations and dissociations between BOLD signal and P300 amplitude in an auditory oddball task: a parametric approach to combining fMRI and ERP. *Magn Reson Imaging, 20*(4), 319-325.

Hyvarinen, A., & Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Netw, 13*(4-5), 411-430.

Ibrahim, J. G., Chen, M. H., Gwon, Y., & Chen, F. (2015). The Power Prior: Theory and Applications. *Stat Med, 34*(28), 3724-3749.

Itti, L., & Baldi, P. (2005). *A principled approach to detecting surprising events in video.* Paper presented at the Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on.

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision research, 49*(10), 1295-1306.

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

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). FSL. *Neuroimage, 62*(2), 782-790.

Jepma, M., Brown, S., Murphy, P. R., Koelewijn, S. C., de Vries, B., van den Maagdenberg, A. M., & Nieuwenhuis, S. (2018). Noradrenergic and Cholinergic Modulation of Belief Updating. *J Cogn Neurosci*, 1-18.

Jepma, M., Murphy, P. R., Nassar, M. R., Rangel-Gomez, M., Meeter, M., & Nieuwenhuis, S. (2016). Catecholaminergic Regulation of Learning Rate in a Dynamic Environment. *PLoS Comput Biol, 12*(10), e1005171.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annu Rev Psychol, 55*, 271-304.

Kingstone, A. (1992). Combining expectancies. *The Quarterly Journal of Experimental Psychology Section A, 44*(1), 69-104.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception, 36*(14), 1-16.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci, 27*(12), 712-719.

Kobayashi, K., & Hsu, M. (2017). Neural Mechanisms of Updating under Reducible and Irreducible Uncertainty. *J Neurosci, 37*(29), 6972-6982.

Kolossa, A., Fingscheidt, T., Wessel, K., & Kopp, B. (2013). A model-based approach to trial-by-trial p300 amplitude fluctuations. *Front Hum Neurosci, 6*, 359.

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

Kononowicz, T. W., & van Rijn, H. (2014). Decoupling interval timing and climbing neural activity: a dissociation between CNV and N1P2 amplitudes. *J Neurosci, 34*(8), 2931-2939.

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

Kopp, B., Seer, C., Lange, F., Kluytmans, A., Kolossa, A., Fingscheidt, T., & Hoijtink, H. (2016). P300 amplitude variations, prior probabilities, and likelihoods: A Bayesian ERP study. *Cognitive, Affective, & Behavioral Neuroscience, 16*(5), 911-928.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). lmerTest: tests in linear mixed effects models. R package version 2.0-20. *Vienna: R Foundation for Statistical Computing*.

Lakens, D. (2017). Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses. *Soc Psychol Personal Sci, 8*(4), 355-362.

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, 1(2), 259–269.

Leech, R., Braga, R., & Sharp, D. J. (2012). Echoes of the brain within the posterior cingulate cortex. *J Neurosci, 32*(1), 215-222. doi:10.1523/jneurosci.3689-11.2012

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. *J Neurosci, 28*(47), 12539-12545. doi:10.1523/jneurosci.2925-08.2008

Mars, R. B., Jbabdi, S., Sallet, J., O'Reilly, J. X., Croxson, P. L., Olivier, E., . . . Rushworth, M. F. (2011). Diffusion-weighted imaging tractography-based parcellation of the human parietal cortex and comparison with human and macaque resting-state functional connectivity. *J Neurosci, 31*(11), 4087-4100.

Mathys, C. D., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., & Stephan, K. E. (2014). Uncertainty in perception and the Hierarchical Gaussian Filter. *Front Hum Neurosci, 8*.

Mattes, S., & Ulrich, R. (1997). Response force is sensitive to the temporal uncertainty of response stimuli. *Percept Psychophys, 59*(7), 1089-1097.

Menon, V. (2015). Salience Network. In A. W. Toga (Ed.), *Brain Mapping* (pp. 597-611). Waltham: Academic Press.

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Struct Funct, 214*(5-6), 655-667.

Mensen, A., & Khatami, R. (2013). Advanced EEG analysis using threshold-free cluster-enhancement and non-parametric statistics. *Neuroimage, 67*, 111-118.

Montefinese, M., Ambrosini, E., & Roivainen, E. (2018). No grammatical gender effect on affective ratings: evidence from Italian and German languages. *Cogn Emot*, 1-7. doi:10.1080/02699931.2018.1483322

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *J Neurosci, 30*(37), 12366-12378.

Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological bulletin, 89*(1), 133.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychol Bull, 131*(4), 510-532.

Nimchinsky, E. A., Gilissen, E., Allman, J. M., Perl, D. P., Erwin, J. M., & Hof, P. R. (1999). A neuronal morphologic type unique to humans and great apes. *Proc Natl Acad Sci U S A, 96*(9), 5268-5273.

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

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nat Neurosci, 15*(12), 1729-1735.

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. *Proc Natl Acad Sci U S A, 110*(38), E3660-3669.

O'Reilly, J. X., & Mars, R. B. (2015). Bayesian models in cognitive neuroscience: A tutorial. In *An Introduction to Model-Based Cognitive Neuroscience* (pp. 179-197): Springer.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113.

Ouyang, G., Sommer, W., & Zhou, C. (2015). A toolbox for residue iteration decomposition (RIDE)--A method for the decomposition, reconstruction, and single trial analysis of event related potentials. J Neurosci Methods, 250, 7-21.

Pearson, J. M., Heilbronner, S. R., Barack, D. L., Hayden, B. Y., & Platt, M. L. (2011). Posterior cingulate cortex: adapting behavior to a changing world. *Trends Cogn Sci, 15*(4), 143-151.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision, 10*(4), 437-442.

Penny, W. (2012). Bayesian models of brain and behaviour. *ISRN Biomathematics, 2012*.

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalogr Clin Neurophysiol, 72*(2), 184-187.

Polich, J. (2003). Theoretical Overview of P3a and P3b. In J. Polich (Ed.), *Detection of Change: Event-Related Potential and fMRI Findings* (pp. 83-98). Boston, MA: Springer US.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *J Exp Psychol, 109*(2), 160-174.

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage, 59*(3), 2142-2154.

R Core Team. (2015). R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2015.

Rogers, J. L., Howard, K. I., & Vessey, J. T. (1993). Using significance tests to evaluate equivalence between two experimental groups. *Psychol Bull, 113*(3), 553-565.

Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J Pharmacokinet Biopharm, 15*(6), 657-680.

Schwartenbeck, P., FitzGerald, T. H. B., & Dolan, R. (2016). Neural signals encoding shifts in beliefs. *Neuroimage, 125*, 578-586.

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., . . . Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *J Neurosci, 27*(9), 2349-2356.

Seer, C., Lange, F., Boos, M., Dengler, R., & Kopp, B. (2016). Prior probabilities modulate cortical surprise responses: a study of event-related potentials. *Brain and cognition, 106*, 78-89.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J., 27*, 623-656.

Smith, N. J., & Kutas, M. (2015). Regression-based estimation of ERP waveforms: I. The rERP framework. *Psychophysiology, 52*(2), 157-168.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., . . . Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage, 23 Suppl 1*, S208-219.

Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology, 38*(2), 343-358.

Squires, K. C., Wickens, C., Squires, N. K., & Donchin, E. (1976). The effect of stimulus sequence on the waveform of the cortical event-related potential. *Science, 193*(4258), 1142-1146.

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalogr Clin Neurophysiol, 38*(4), 387-401.

Sridharan, D., Levitin, D. J., & Menon, V. (2008). A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proc Natl Acad Sci U S A, 105*(34), 12569-12574.

Stern, E. R., Gonzalez, R., Welsh, R. C., & Taylor, S. F. (2010). Updating beliefs for a decision: Neural correlates of uncertainty and underconfidence. *J Neurosci, 30*(23), 8032-8041.

Stuss, D. T., Alexander, M. P., Shallice, T., Picton, T. W., Binns, M. A., Macdonald, R., . . . Katz, D. I. (2005). Multiple frontal systems controlling response speed. *Neuropsychologia, 43*(3), 396-417.

Sutton, S. (1979). P300--thirteen years later. In *Evoked brain potentials and behavior* (pp. 107-126): Springer.

Sutton, S., & Ruchkin, D. S. (1984). The late positive complex. Advances and new problems. *Ann N Y Acad Sci, 425*, 1-23.

Tanner, D., Morgan-Short, K., & Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology, 52*(8), 997-1009.

Tomassini, A., Ruge, D., Galea, J. M., Penny, W., & Bestmann, S. (2016). The Role of Dopamine in Temporal Uncertainty. *J Cogn Neurosci, 28*(1), 96-110.

Trillenberg, P., Verleger, R., Wascher, E., Wauschkuhn, B., & Wessel, K. (2000). CNV and temporal uncertainty with 'ageing' and 'non-ageing' S1-S2 intervals. *Clin Neurophysiol, 111*(7), 1216-1226.

Trivino, M., Correa, A., Arnedo, M., & Lupianez, J. (2010). Temporal orienting deficit after prefrontal damage. *Brain, 133*(Pt 4), 1173-1185.

Vallesi, A. (2010). Neuro-anatomical substrates of foreperiod effects. *Attention and time*, 303-316.

Vallesi, A., McIntosh, A. R., Shallice, T., & Stuss, D. T. (2009). When time shapes behavior: fMRI evidence of brain correlates of temporal monitoring. *J Cogn Neurosci, 21*(6), 1116-1126.

Vallesi, A., McIntosh, A. R., & Stuss, D. T. (2009). Temporal preparation in aging: a functional MRI study. *Neuropsychologia, 47*(13), 2876-2881.

Vallesi, A., Mussoni, A., Mondani, M., Budai, R., Skrap, M., & Shallice, T. (2007). The neural basis of temporal preparation: insights from brain tumor patients. *Neuropsychologia, 45*(12), 2755-2763.

Vallesi, A., Shallice, T., & Walsh, V. (2007). Role of the prefrontal cortex in the foreperiod effect: TMS evidence for dual mechanisms in temporal preparation. *Cereb Cortex, 17*(2), 466-474.

Verleger, R., Grauhan, N., & Smigasiewicz, K. (2016). Is P3 a strategic or a tactical component? Relationships of P3 sub-components to response times in oddball tasks with go, no-go and choice responses. Neuroimage, 143, 223-234.

Vilares, I., & Kording, K. (2011). Bayesian models: the structure of the world, uncertainty, behavior, and the brain. *Ann N Y Acad Sci, 1224*(1), 22-39.

Vossel, S., Mathys, C., Stephan, K. E., & Friston, K. J. (2015). Cortical Coupling Reflects Bayesian Belief Updating in the Deployment of Spatial Attention. *J Neurosci, 35*(33), 11532-11542.

Waskom, M. L., Kumaran, D., Gordon, A. M., Rissman, J., & Wagner, A. D. (2014). Frontoparietal representations of task context support the flexible control of goal-directed cognition. *Journal of Neuroscience, 34*(32), 10743-10755.

Waskom, M. L., Frank, M. C., & Wagner, A. D. (2017). Adaptive engagement of cognitive control in context-dependent decision making. Cerebral Cortex, 27(2), 1270-1284.

Wilkinson, G., & Rogers, C. (1973). Symbolic description of factorial models for analysis of variance. *Applied Statistics*, 392-399.

Winkler, I., Debener, S., Muller, K. R., & Tangermann, M. (2015). On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. *Conf Proc IEEE Eng Med Biol Soc, 2015*, 4101-4105.

Wolpert, D. M. (2007). Probabilistic models in human sensorimotor control. *Hum Mov Sci, 26*(4), 511-524.

Woodrow, H. (1914). The measurement of attention. *The Psychological Monographs, 17*(5), i.

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron, 46*(4), 681-692.

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends Cogn Sci, 10*(7), 301-308.

Zeki, S., Watson, J. D., Lueck, C. J., Friston, K. J., Kennard, C., & Frackowiak, R. S. (1991). A direct demonstration of functional specialization in human visual cortex. *J Neurosci, 11*(3), 641-649.