Informational segmentation in event prediction: Temporal dynamics and predictive efficiency

Inaugural-Dissertation

zur Erlangung des Doktorgrades

im Fachbereich Psychologie und Sportwissenschaft

der Westfälischen Wilhelms-Universität in Münster

Vorgelegt von

Daniel Kluger

aus Oberhausen

- 2019 -

Dekan:	Prof. Dr. Guido Hertel
Erste Gutachterin:	Prof. Dr. Ricarda Schubotz
Zweite Gutachterin:	Prof. Dr. Pienie Zwitserlood
Tag der mündlichen Prüfung:	.....
Tag der Promotion:	.....

Table of contents

List of figures

List of tables

List of abbreviations

fMRI	functional magnetic resonance imaging	PCG	post-central gyrus
EEG	electroencephalography	STG	superior temporal gyrus
BOLD	blood oxygen level dependent	ANG	angular gyrus
ACC	anterior cingulate cortex	TPJ	temporoparietal junction
IFG	inferior frontal gyrus	PHC	parahippocampal cortex
ANOVA	analysis of variance	PFC	prefrontal cortex
FDR	false discovery rate	MTG	middle temporal gyrus
EPI	echo-planar imaging	RSA	representational similarity analysis
TR	repetition time	PE	prediction error
TE	echo time	CP	checkpoint
MNI	Montreal neurological institute	STD	standard trial
FWHM	full-width half-maximum	ERP	event-related potential
GLM	general linear model	TMS	transcranial magnetic stimulation
ECM	Eigenvector centrality mapping	EOG	electrooculogram
ROI	region of interest	AAHC	atomize and agglomerate hierarchical clustering
PR	percentage of recognition	TM	template map

PCC	posterior cingulate cortext	AC	adjustment cue
SFS	superior frontal sulcus	SNR	signal-to-noise ratio
BA	Brodman area	LDA	linear discriminant analysis
TFCE	threshold-free cluster enhancement	DSM	dissimilarity matrix

General acknowledgements

First and foremost I want to thank my supervisor Prof. Ricarda Schubotz for giving me the opportunity to work and pursue my PhD under her tutelage. I deeply appreciate the means, the time, and the stimulating ideas she has invested in the work we have done together. I am thankful to have had the freedom of working independently when I wanted to and her guidance to rely on when I needed it.

Current and former advisors, colleagues, and collaborators have largely contributed to the projects this thesis is built around. I would like to sincerely thank Prof. Pienie Zwitserlood for her continued support since I joined this university. Not only has she kindly agreed to co-supervise this thesis, but her insightful advice has always been much appreciated. Postdocs Axel Kohler and Moritz Wurm have taught me with extraordinary commitment and their expertise was a great pleasure to work with. I want to thank Christiane Ahlheim for her primary support and for leading the way for many others in the group, sporting an almost unreasonable level of enthusiasm for science that was highly motivational. As always, I am grateful that Monika Mertens has kept data acquisition smoothly on track throughout the years.

A lot of friends deserve an even bigger lot of thanks for always being there and doing the small things that matter: MAR and CMW for helpful comments on this thesis; DZ and LQ for self-help over dinner; NES and ICT for self-help in the absence of dinner; MLW, PS, NPM, and TPS for all the phone calls; MK and LW for never going away; JK for putting up with living with me through the hardest part; NB and AKK for memorable diversions; and MMG for pudding and apple pie.

I would like to thank my family for all their love, patience, and encouragement. My parents for teaching me to be curious and my sister for setting an example on how to handle the tough stuff. Last but maybe most importantly, I owe everything to Prof. Ulrich Mußhoff without whom I would have ended up in an entirely different place.

What a time it has been.

Daniel S. Kluger
February 2019

Abstract

Predictive brain processing is computationally efficient in that downstream predictions from internal models are used to compress incoming sensory information solely to what is unexpected. In turn, these *prediction errors* instigate rapid perceptual inference as well as slower, more global learning mechanisms through model updating on different time scales. This dissertation presents three consecutive studies aimed to extend the understanding of predictive efficiency: In study 1, we aimed to qualitatively dissociate non-reward error signals in fMRI. Furthermore, we examined strategic adaptation to contextual uncertainty. *Checkpoints* emerged as model-compliant events informing model evaluation under global uncertainty.

Study 2 employed EEG to build functional profiles of checkpoints and prediction errors, revealing P3b as a joint index of informational gain. In contrast, N400 was found to indicate the mismatch signal for prediction errors (vs checkpoints).

Again using EEG, study 3 assessed *adjustment cues* as a second class of model-compliant events which – in contrast to checkpoints – provided on-line information about local changes in probability. By means of multivariate pattern classification and representational similarity analysis, we were able to show how the global model – despite clear behavioural and functional indices of adjustment cue exploitation – takes precedence over local contingency changes. Implications on the role of prediction errors, checkpoints, and adjustment cues as well as their interplay across time scales are discussed. In conclusion, predictive processing exploits mismatch signals to adapt and – when called for – model-compliant information to evaluate internal models across time scales for the sake of computational efficiency.

1. General introduction

1.1 Relevance and outline of this dissertation

In recent years, the concept of a *predictive brain* has been sustainably revived in cognitive neuroscience (Clark, 2013). Very generally speaking, this notion postulates that it is one of the central faculties of the brain to continuously - and actively - estimate both its own state and that of the world surrounding it (Friston, 2005; Rao & Ballard, 1999). These predictive estimations arise from internal, situational models of the world shaped by past experience and are constantly checked against incoming sensory information (Huang & Rao, 2011; Lee & Mumford, 2003). Any mismatch between expected and obtained information - termed a *prediction error* - is immediately fed forward to instigate model updating (Bar, 2009). As a result, predictive processing is broadly conceptualised as a bidirectional interplay of top-down predictive codes and bottom-up error propagation (Mumford, 1992), which will be elaborated on in the next sections.

Naturally, such an account of brain function fundamentally challenges the classical view of a passive brain waiting for and then reacting to sensory input. Its unequivocal benefits, however, are evident: By restricting bottom-up signal flow solely to unpredicted (i.e., surprising) portions of information, predictive processing is computationally and energetically efficient (Friston, 2009). Numerous studies from various scientific domains have demonstrated the merit of anticipatory processing in facilitating perception (Bar et al., 2006; Todorovic et al., 2011), cognitive control (Alexander & Brown, 2011), language (Kutas & Hillyard, 1980), and motor performance (Bestmann et al., 2008). Moreover, model adaptation as an immediate consequence of false predictions has been widely suggested as a parsimonious framework for various learning mechanisms (see *1.3 Model adaptation and evaluation*).

With an increasing amount of attention gained in the neuroscientific community, critical advances have been made in apprehending predictive processing as a general principle of brain function. While these previous accomplishments will be the focus of the following sections, the extensive explanatory potential of predictive coding has raised a number of topics that have yet to be adequately addressed. In this dissertation, I will thus first provide a more detailed summary of the implications and implementations of predictive processing as well as a selective review of previous work. Introductory chapter 1 then concludes with a summary of overarching aims of this dissertation and specific research questions motivating respective experiments. What follows are three studies conducted over the course of my PhD (chapters 2 - 4). In chapter 5, central findings will be brought together in a general summary before larger-scale implications of selected results are discussed in more detail. After a critical reflection on the present body of work, I will provide a comprehensive outlook on future directions and conclude this dissertation.

1.2 The predictive brain

The idea of prediction as a core mechanism of brain function goes back to Von Helmholtz' work in the 19th century and is rooted in theoretical considerations of efference copies in oculomotor control (Von Helmholtz, 1860). Its recent reformulations, most prominently the more formalised *predictive coding* account (Friston, 2005; Mumford, 1992), are in stark contrast with traditional views of perception: The classical perspective on perception assumed primacy of bottom-up feature detection (Marr, 1982). Sensory areas such as the visual cortex were conceptualised as a hierarchy of neural feature detectors whose responses were driven by bottom-up stimulus properties (Egner et al., 2010). However, among the critical bottlenecks of this stimulus-driven view of brain function was the implication that such passive brains should cease activation in the absence of task-specific simulation

- an assumption decidedly refuted by a growing body of work on *resting state activity* and the *default mode* (for review, see Raichle & Snyder, 2007). Rather, reverting to Helmholtz' proposal, theoretical models concerned with ongoing and endogenous brain activity have successively evolved into the concept of predictive processing prevalent in modern day neuroscience (see Clark, 2013; Howhy, 2012). Based on the idea that a pro-active brain constantly tries to predict streams of sensory information, this downward flow of predictions now carries most of the computational burden (Clark, 2015). This allows short-term processing to focus exclusively on the portion of the input signal that cannot be 'explained away' by top-down predictions. As a result, the driving sensory signal effectively just provides corrective feedback prompted by these residual prediction errors (Friston, 2005; Hohwy, 2013). In other words, the classically assumed forward flow of sensory information is replaced by the forward flow of prediction error (Feldman & Friston, 2010).

The underlying principle of compressing information solely to informative bits is widely applied as a data compression strategy in auditory (MP3) and visual signal processing (JPEG). Translated into the terminology of predictive coding, the informative bits of information equal the unpredicted - or surprising - portion of sensory signals that cannot be explained by model-based assumptions (Friston, 2011). The concept of *surprise* (in a formalised rather than an intuitive sense) thus plays a central role in all accounts of predictive processing, the implications and evidence of which will be outlined in the next section (*1.3 Model adaptation and evaluation*).

Importantly, the computational efficiency of cancelling out redundant information is substantiated further by implementing a hierarchical structure. The interplay of downstream prediction and upstream prediction errors is suggested to operate at every level of the cortical hierarchy (Hohwy et al., 2008): Using high-level knowledge, top-down probabilistic generative models encode the system's 'best guess' of neural population activity within lower levels. This bidirectional cascade makes use of layer-specific *representation neurons* (carrying model assumptions) and *error neurons* (propagating

the mismatch signal to the next highest hierarchy level) for highly efficient, parallel computation (Summerfield & Egner, 2009). This way, non-informative signal is disregarded already at low levels of the processing hierarchy (i.e., primary sensory cortices). By means of such a bidirectional hierarchical structure, predictive processing can elegantly account for adjustments of the system's model across time scales, which warrants a closer look at how these adaptations are implemented.

1.3 Model adaptation and evaluation

Thinking out the mechanisms described so far, the overarching goal of any predictive brain is to acquire situational models that are as fitting as possible for as many contexts as possible: The more incoming sensory signal is accounted for by model predictions, the smaller the residual prediction error and, consequently, the computational load of processing. For updating and optimising predictions, making use of a hierarchical structure induces so-called *empirical priors*, meaning that the best model available at a given level is used to infer priors at the level below (Neal & Hinton, 1998). These priors are simply constraints reflecting current context and are therefore continuously tuned by the sensory input itself. Model adaptation thus critically depends on the occurrence of prediction errors: Following a mismatch at any hierarchy level, error-indicating activity is propagated upstream to rapidly adjust probabilistic representations at the next highest level. This way, adjustments can quickly be made to react on-line, while the error signal is simultaneously used to instigate long-term, structural changes to the model. In terms of efficient processing, the system uses such perceptual learning mechanisms to prepare the model for future encounters of the same or a similar context (Clark, 2013).

At this point, it is worth noting that model adaptation is not dependent on the domain of the error signal: As predictive processing applies the same computational strategies (i.e., 'explaining away' sensory input for prediction error minimisation) to perception, cognition, and action, learning can

take place in any of these domains following the same principles (see above). The concept of *active inference* (Friston, 2009; Friston et al., 2010) extends to actions, meaning that error signals are used to elicit targeted movements that incrementally change sensory as well as proprioceptive input (Friston, 2003). This framing of motor control may need getting used to, but the underlying logic is very similar to the predictive account of perception outlined above: Imagine yourself sitting at the kitchen table as you intend to grasp a glass of water in front of you. Active inference suggests that the desired state (i.e., your hand grasping the glass) is 'simulated' and sent downstream where the substantial mismatch with the perceived state (i.e., your hand lying on the kitchen table) is registered. Consequently, this error signal initiates movement of the sensors (i.e., the hand) to gradually minimise the discrepancy between predicted (i.e., desired) state and current sensory information.

Having established the importance of prediction errors for model updating, it is instructive to consider the implication that - at least in everyday life - no two prediction errors are exactly alike: They may differ with regard to quantity (i.e., expecting a $1 reward and getting $2 vs $200) as well as their quality (i.e., expecting a $1 reward and getting $2 vs losing $2). Quantitative differences between prediction errors have been the focus of many studies with the concept of surprise (sometimes termed *surprisal*; Tribus, 1961) introduced to reflect the improbability of a certain event (Jones, 1979). While a more formal explanation is provided in the methods sections of studies 1 and 2, the basic reasoning is this: The less probable the occurrence of an event, the more informative it is. In other words, the high-amplitude prediction error induced by highly improbable (i.e., surprising) events leads to more intense model adaptation (Egner et al., 2010; Strange et al., 2005).

As reflected by the example above, qualitatively different prediction errors have predominantly been assessed in reward-related studies. For example, *positive* and *negative* prediction errors (i.e., getting more or less reward than expected, respectively; Schultz, 1998) have long been established in reward paradigms. However, aside from 'just being right', much of our everyday experience is not

particularly rewarded. Consider throwing a ball to a friend and missing: For consecutive attempts, not only the amplitude (i.e., quantity - one or two feet), but also the direction (i.e., quality - left or right) of the error signal entails corrective adjustments. Predictive processing - if it is indeed the fundamental mechanism of brain function (see Friston, 2010) - should thus very equally differentiate and classify these kinds of prediction errors to apply apt model adjustments for future occasions. In neuroscientific research, however, investigations of such qualitatively different non-reward prediction errors have to this date been limited to a handful of studies in the cognitive (Den Ouden et al., 2010; Schiffer et al., 2012) and the perceptual domain (O'Reilly et al., 2013). Therefore, it was one of the focal starting points of the work presented in this dissertation to assess distinctive characteristics of non-reward prediction errors and their neural correlates in fMRI (for details, see *1.5 Aims of this dissertation*).

To recap, prediction errors drive learning through progressive model adaptation. Error signals can be distinguished quantitatively - e.g., by the amount of surprise they carry - or qualitatively. Finally, distinct qualities of prediction errors, besides having distinct immediate consequences for behaviour, bring about differential model adaptations to improve subsequent predictions.

Critically, some recent studies have stressed a role of probabilistic events that carry behaviourally relevant information but do not index a canonical mismatch. Using digit sequences, Kühn and Schubotz (2012) demonstrated distinct frontal activations for expected, model-compliant events at sequential positions where rare prediction errors had previously been observed. In a similar vein, Trempler and colleagues (2017) found partly joint activation patterns for sequential violations that were model-compliant (termed *drifts,* e.g. 1-2-3-4 | 1-2-4) and those that prompted corrective model updating (termed *switches*, e.g. 1-2-3-4 | 3-2-1). Both studies thus suggest that expected, model-compliant information at critical time points is used to some extent to inform the internal model. Even if everything goes according to plan, it appears intuitive to make use of information that can be

gained to either reaffirm or, more generally, evaluate model assumptions for future profit. Particularly in uncertain or ambiguous contexts, these evaluations can prove useful to estimate higher-level statistics. When driving down an unknown street, you might find yourself checking speed limit signs more frequently, evaluating whether your current set of assumptions (e.g., "I have to go below 60") are still valid. In Bayesian terms, this originates from insufficient *priors* (i.e., not exactly knowing what to expect) increasing the system's reliance on external information (Feldman & Friston, 2010). Therefore, exploitation of model-congruent information for model evaluation conceivably depends on global contextual features to maximise predictive efficiency.

1.4 Predictive efficiency

As highlighted earlier, computational efficiency is one of the fundamental principles of predictive processing: By means of predictive codes sent top-down to cancel out expectable portions of sensory input, the computational load is restricted to bottom-up propagation of error signals. In fact, minimisation of prediction error (or *entropy*) has been proposed to be the one mechanism underlying all brain function (Friston, 2010). While this rather radical proposal does not come without major challenges (see Little & Sommer, 2013, for discussion and a solution to the so-called *dark room dilemma*), the energetically efficient nature of predictive coding is appealing. Consequently, especially when considering global context features as a modulator, it remains an intriguing question how the predictive system adapts to environments with different sets of rules to be learned.

In order for predictive processing to be computationally efficient across a variety of contexts, the system has to a) register higher-order changes and b) flexibly adapt its predictive codes to the changed environment. To start with, processing of such changes in statistical regularities per se has conclusively been demonstrated in both reward (Behrens et al., 2007) and non-reward paradigms (Tobia et al., 2012). As for the nature of these changes, most studies involve some uncertainty

formulation (Bland & Schaefer, 2012; Yu & Dayan, 2005): For example, *volatility* means fluctuation in the probability with which certain outcomes are observed (Behrens et al., 2007) whereas *irreducible uncertainty* refers to uncertainty that remains even after successful learning of local (i.e., lower-order) contingencies (De Berker et al., 2016; Payzan-LeNestour & Bossaerts, 2011). In short, even if a good situational model has been learned, global (i.e., higher-order) changes may still induce uncertainty regarding the applicability of that model.

In light of the challenge this greater uncertainty poses onto the predictive brain, it seems unlikely that an otherwise so highly efficient system would rely on a method of trial and error. Predicting upcoming events can be seen as an investment in that mismatch signals arising from false predictions immediately prompt energetically costly adjustments of the forward model (Clark, 2013). In matters of minimising prediction errors, it would thus be most efficient to strategically adapt the way in which predictions are made - specifically, how far these predictive codes reach into the future. Consider walking down two flights of stairs in the dark: Even if you basically know your way down ("The first flight has 12 steps, the second one has 14"), using that full-length model bears the risk of encountering (potentially hurtful) prediction errors. Instead, recognising the unsafe context, it would be more efficient to make a sequence of partial predictions and evaluate their validity as you go along. More generally, when contextual features indicate that forward models cannot be used reliably, it may be more efficient to employ a short-term strategy of stepwise predictions, evaluating model validity on-line. This way, such strategic adaptation closely relates to context-dependent use of model-compliant information (see *1.3 Model adaptation and evaluation*): Most likely to serve on-line model evaluation are critical points in time where prediction errors *could potentially* occur, i.e. monitoring the end of the first (and, having passed it, the second) flight of stairs in order not to over- or underestimate the number of steps.

Critically, global context modulations have to be distinguished from (cor-)related concepts like trial-by-trial variation in surprise (for details, see discussion of study 1): In reliable, low uncertainty contexts, prediction errors are rare (i.e., surprising) and therefore highly informative with regard to potential model adaptation. In contrast, prediction errors are more frequent - and thus less informative - in highly uncertain environments. Therefore, studies aiming to assess potential strategic adaptation in predictive processing have to disentangle local probability changes from global modulations caused by variations of context.

1.5 Aims of this dissertation

Predictive accounts of brain function are theoretically well-founded and have gained empirical support from a multitude of domains in cognitive neuroscience (see Clark, 2013, for an extensive review). The work presented in this dissertation aimed to address some of the compelling questions that have so far remained unanswered. Broadly speaking, these open questions focussed on the functional profiles of model-congruent and -incongruent event processing as well as their modulation by means of local and global parameters.

More specifically, the aim of study 1 was twofold: First, we aimed to assess the respective neural correlates of qualitatively distinct non-reward prediction error signals using fMRI and behavioural data of a sequential pattern detection task. Second, we manipulated higher-order contextual statistics (i.e., irreducible uncertainty) to investigate potential adaptation of the predictive strategy to these environmental changes. We hoped to demonstrate how the prospective scope of top-down predictions is adjusted to unstable contexts using information at critical sequential positions. Consequently, we proposed the concept of *checkpoints* to capture model-compliant events which inform downstream models under uncertainty. Following up on these findings, study 2 was designed as a replication study exploiting the temporal benefits of EEG. We aimed to characterise the

relationship of canonical prediction errors and checkpoints - namely, commonalities and differences with regard to their functional profiles. As a result, we proposed checkpoint-induced *model evaluation* as an informative process similar to *model adaptation* instigated by prediction errors. Pursuing the question regarding the interplay of local vs global factors, behavioural subgroup analyses were conducted to assess how implicit cue learning may influence information use at critical reference points. This finally motivated study 3 in which we employed a variant of the experimental paradigm that included incidental *adjustment cues* effectively overwriting learned task structures. Combining classic EEG analysis and multivariate pattern classification methods, we followed up on results of studies 1 and 2 to further understand how locally available information influences prediction in changing global contexts.

2. Study 1

Strategic adaptation to non-reward prediction error qualities and irreducible uncertainty in fMRI

Daniel S. Kluger[a,b] and Ricarda I. Schubotz[a,b,c]

[a] Department of Psychology, University of Muenster, Muenster, Germany

[b] Otto-Creutzfeldt-Center for Cognitive and Behavioral Neuroscience, University of Muenster, Muenster, Germany

[c] Department of Neurology, University Hospital Cologne, Cologne, Germany

Correspondence should be addressed to

Daniel S. Kluger
Department of Psychology
University of Muenster
Fliednerstr. 21
D-48149 Muenster
Germany
E-mail: daniel.kluger@uni-muenster.de

Key words: fMRI, expectation, prediction error, sequence learning, top-down control

Published in: Cortex (2017)

2.1 Abstract

Prediction errors are deemed necessary for the updating of internal models of the environment, prompting us to stop or assert current action plans and helping us to adapt to environmental features. The aim of the present study was twofold: First, we sought to determine the neural underpinnings of qualitatively different abstract prediction errors in a serial pattern detection task. Distinct frontoparietal components were found for sequential terminations (inferior frontal gyrus) and extensions (superior frontal sulcus, posterior cingulate cortex, and angular gyrus), respectively. These findings provide a novel approach of distinguishing non-reward prediction error signals with regard to behavioural consequences they entail.

Second, we investigated predictive processing as a function of statistical context (*irreducible uncertainty*). We hypothesised that the prospective scope of model-based expectancies is adapted to the stability of respective contexts in that unstable environments call for more frequent comparisons of expectancies with sensory input, resulting in stepwise predictions. Changes in environmental stability were reflected in activation of the angular gyrus and inferior frontal gyrus for the highly uncertain context at potential points of prediction violation (*checkpoints*). Notably, this effect was not due to local fluctuations in stimulus improbability (*surprise*). Although further behavioural support is needed, data point towards a context-dependent adaptation of predictive strategies. Conceivably, enhanced BOLD responses at sequential checkpoints could reflect stepwise rather than full-length prediction. This strategic adjustment presumably relies on the iterant evaluation of model information retrieved from working memory, as suggested by strengthened functional connectivity of the parahippocampal area during epochs of high uncertainty.

2.2 Introduction

In attempting to make sense of incoming sensory event information in everyday life, we are constantly faced with discrepancies between our internal model of the world and the information we actually obtain (Rao & Ballard, 1999). In predicting future events, the brain efficiently processes predictable portions in perceptual information: only if incoming sensory signals differ from higher-level predictions is the corresponding error signal propagated "upward" to the next highest stage of the processing hierarchy (Friston, 2005; Mumford, 1992). Such prediction errors inarguably differ with regard to their quality and the behavioural consequences they entail: when walking down a familiar flight of stairs, one might occasionally over- or underestimate the remaining number of stairs on the bottom landing, evoking qualitatively different moments of surprise. Furthermore, context will likely influence the way in which we make assumptions about upcoming events: we will compare our expectations with reality more frequently in noisy or unfamiliar contexts, for instance when carrying a fridge through the same staircase in dim lighting.

Research on definitive features of these prediction errors has predominantly been directed towards quantitative rather than qualitative differences between neural error signals. For instance, the concept of *surprise* (Jones, 1979), cast to reflect the improbability of a particular event, has been demonstrated to modulate prediction errors in that more surprising expectancy violations elicit stronger error signals (Egner et al., 2010; Strange et al., 2005). However, aside from a few studies in the cognitive (e.g. den Ouden et al., 2010; Schiffer et al., 2012) and the perceptual domain (O'Reilly et al., 2013), imaging efforts to classify *qualitatively* different types of such prediction errors have so far been restricted to contexts involving a reward component. Here, findings from animal studies (Bayer & Glimcher, 2005) and gambling tasks in humans (Delgado et al., 2003) have demonstrated a functional

distinction of *positive* and *negative* reward prediction errors based on whether an obtained reward value was higher or lower than expected, respectively (Schultz, 1998).

In contrast, such distinctive characteristics of prediction errors are considerably less well understood in absence of external reward (note that predicting or responding correctly may be rewarding in its own right (Holroyd & Coles, 2002), which is not the objective of the present study). In light of the immediate importance of error signals for internal model updating (e.g. Bastos et al., 2012) and adequate action selection, however, the need to understand the qualities of more abstract prediction errors becomes evident: as prediction errors are thought to be the foundation of learning mechanisms (den Ouden et al., 2009), different types of expectancy violations might result in qualitatively distinguishable prediction error signals and - consequently - in distinct modifications of upcoming expectancies and behaviour. Recall that short term predictions are used to facilitate perception when we observe structured regularities such as sequential events. When these regularities end earlier than expected, the appropriate adaptation would be to reject the now invalid predictive model in favour of a more *externally* driven mode of tracking incoming information. In contrast, any regularity we perceive may as well persist longer than expected. Accordingly, such an unexpected extension would then call for a resumption of the *internally* driven mode of model-based prediction.

As much as this reactive, local adaptation of stimulus processing and behaviour should reflect different types of prediction errors, it should likewise depend on global, higher order characteristics of the environment. Exploring the nature of error signals and their effects on model adaptation therefore raises the question of whether predictive strategies remain constant across contexts. Different measures of uncertainty have been discussed to refer to the variability of informational value over time, thus reflecting higher-order statistical features of the environment (Bland & Schaefer, 2012; Yu & Dayan, 2005). While sensitivity to context uncertainty has been linked to activity

in the anterior cingulate cortex (ACC) as a function of reward prediction error computation (Behrens et al., 2007; Scholl et al., 2017; Silvetti et al., 2013), implications of context (in)stability on abstract prediction still remain unclear. Therefore, we manipulated the composition of experimental blocks to vary the amount of *irreducible uncertainty* (De Berker et al., 2016; Payzan-LeNestour & Bossaerts, 2011) that would remain after successful learning. Conceivably, memory-driven internal models of future sensory information might be adapted to the uncertainty of respective contexts: In stable contexts of low uncertainty, prediction errors can be conceived of as more meaningful in that they are highly informative with regard to potential benefits of model adaptation (i.e. learning). In contrast, prediction errors in statistically unstable (i.e. highly uncertain) contexts carry less information contributing to learning-related gains. Predicting upcoming events can be considered an investment in the sense that false predictions inevitably lead to costly adjustments of forward models (Clark, 2013). Therefore, depending on how reliably forward models can be used to predict future events (as conveyed through statistical learning of environment regularities), it may be more efficient to employ a strategy of partial or stepwise short-term predictions rather than predicting consecutive events at full length. For example, as opposed to preparing a complete model of an expected sequence of events, a stepwise prediction might imply an iterative monitoring of sequence continuation at particular sequential positions. Research on sequential actions has suggested *decision points* (Norman, 1981; Reason, 1992) as points in time where the selection of appropriate actions requires accessing information about the broader task context. Translated into the perceptual primacy of our experimental paradigm, we propose *checkpoints* as equally informative sequential positions at which broader, uncertainty-dependent contextual information is exploited to prompt strategic adjustments. Most likely to serve as such checkpoints would be those points in time where the occurrence of stimuli was probabilistically modulated by blockwise manipulation of irreducible uncertainty. This way, statistical learning might be employed to adapt predictive strategies to

particular informational structures in the environment. Support for this hypothesis comes from studies linking the working memory network (most prominently hippocampus and adjacent areas) to both the predictive processing of sequential patterns (Fortin et al., 2002; Lisman & Redish, 2009; Rolls, 2013) as well as the decoding of contextual information (Allen et al., 2014; Davachi & DuBrow, 2015).

We conducted the present fMRI study to assess qualitative differences between prediction error signals and the influence of uncertainty on predictive strategies using an implicit cueing paradigm. Participants were asked to detect short or long ordered digit sequences (5 and 7 items, respectively) within an otherwise pseudorandom stream of single digits. They indicated the onset of a detected sequence by an immediate button press and the sequence ending by button release. The expected sequence length, as implicitly cued by digit colour (see Material and Methods), was occasionally violated by terminations and extensions. Thus, whereas the task was overtly concerned with sequence detection, our analysis was focussed on specific events during or at the end of sequences.

As to our first hypothesis, we expected distinguishable neural correlates to reflect the respective reorientation towards external stimuli (sequential terminations) or towards the internal model derived from working memory (sequential extensions). Particularly, the unexpected need to disregard the currently employed internal model as induced by sequential terminations was hypothesised to engage the inferior frontal gyrus (IFG), an area that has been reported for violations of ordered pattern expectancy across domains (Fiebach & Schubotz, 2006).

Second, we assessed the effects of context uncertainty on predictive strategies by manipulating the proportion of violated sequences over time. Neural activity at checkpoint positions was expected to be elevated in highly uncertain contexts, thus indicating a stepwise prediction mode. Due to its established role in reward-related uncertainty monitoring (specifically *volatility*, Behrens et al., 2007),

we assumed ACC to represent higher-level context information necessary to initiate stepwise predictions in highly uncertain environments. Crucially, these global fMRI effects were controlled for potential confounds by varying levels of stimulus-bound surprise (i.e. the extent to which individual stimuli were locally unexpected). To this end, we employed a parametric regressor of nuisance reflecting a stimulus' respective surprise value, thus allowing us to disentangle higher level context effects from mere differences in improbability.

2.3 Material and methods

Participants

A total of 22 neurologically healthy, right-handed volunteers (13 female, mean age: 24.3 (20 – 30) years) participated in the study. Participants were recruited from the university's volunteer database and had normal or corrected-to-normal vision. Colour blindness was ruled out using Ishihara colour test plates (Ishihara, 1917). Written informed consent was obtained from each participant prior to the start of experimental procedures. Experimental standards complied with the local Ethics Committee of the University of Münster. Participants selectively received payment or course credit as compensation for their participation in the study. Two participants were excluded from further data analysis due to poor behavioural performance and self-reported tiredness during the experiment (see below). Therefore, all reported analyses of functional data are based on a sample of 20 participants (12 female, mean age 24.8 (21 - 30) years).

Stimulus material

The stimuli consisted of pseudorandomly coloured single digits (0 – 9, size 1.5° of visual angle) presented individually for 500 ms in the centre of a light grey computer screen (see Fig. 2-1A). Digits were presented in blocks with a length of approximately 6 min. Numbers of presentations for all colours and digits were equally distributed both within and across blocks. Each block contained *sequential trials* (i.e. digits with a recognisable relation to the preceding one) as well as *random trials* (i.e. digits that were not discernibly related to either the preceding or the following digit). Sequential trials in turn belonged to one of two types of sequences: *ordered sequences* constantly increased the preceding figure by one (e.g. 5 – 6 – 7 – 8 – 9; Fig. 2-1A, left). In order to allow for ordered sequences to start on any figure, the ascending regularity necessarily included the 0 character and was thus continued in a circular fashion after the figure 9 (e.g. 8 – 9 – 0 – 1 – 2). *Colour sequences* were defined as consecutive trials all presented in the same colour (dark red). Importantly, the numerical values during colour sequences were pseudorandom (Fig. 2-1A, right). Therefore, at no point was there a succession of continuously ascending figures presented in the same colour, but only one sequential condition at a time (i.e. ordered or colour). Colour sequences were employed to ensure a high level of attentiveness and were not a pivotal subject to our analyses. Random trials included neither self-repetitions nor immediately adjacent figures and therefore could not be mistaken for ordered sequences.

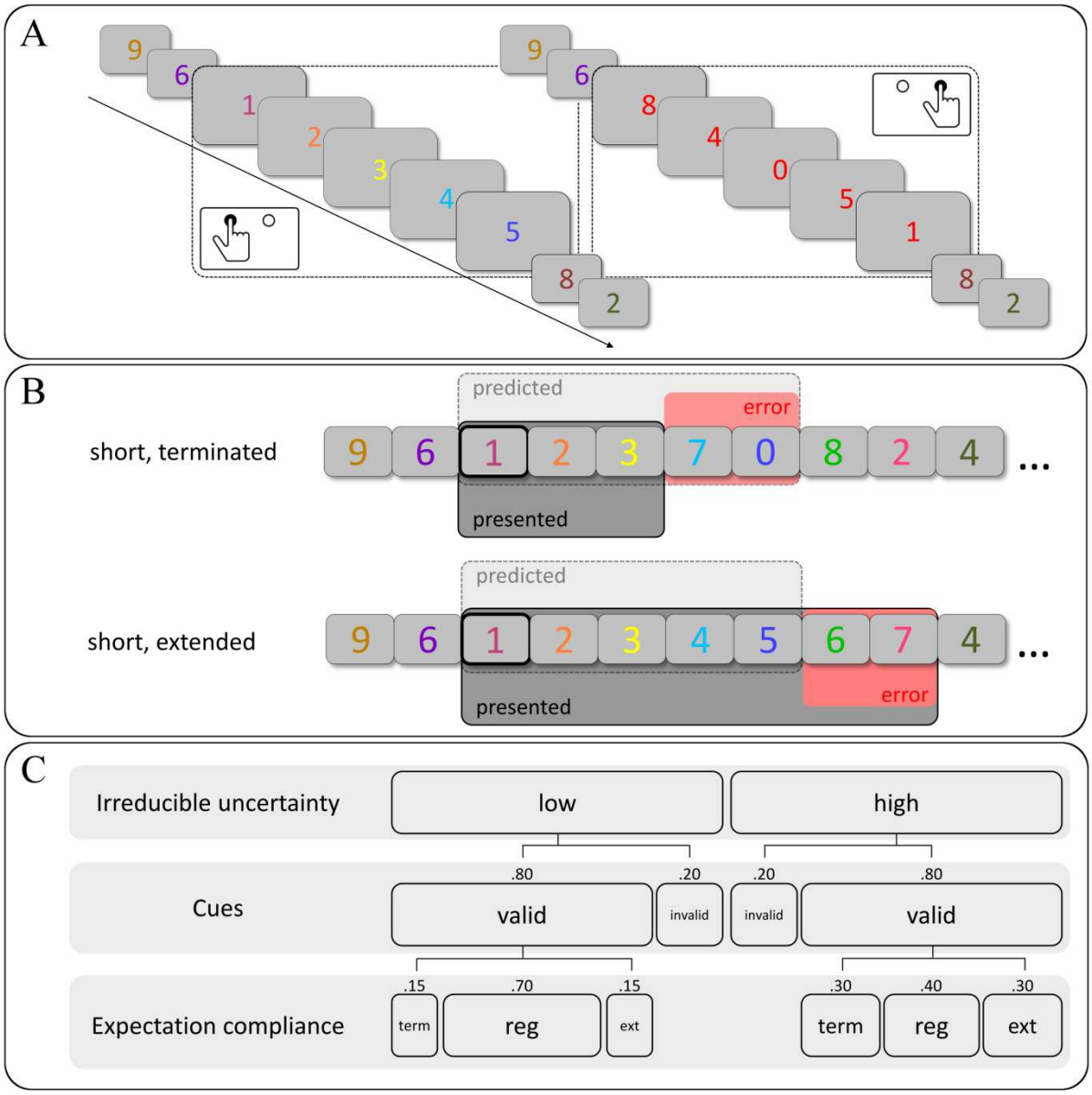

Figure 2-1. (A) Exemplary trial successions and time windows of respective corresponding responses for ordered (left) and colour sequences (right). Sequential trials have been highlighted for illustrative purposes. (B) Cue-based expected sequence length and resulting prediction errors for terminated, and extended short ordered sequences (*expectation compliance*). Based on the cue (in this case, a magenta figure), five regularly ascending figures are expected. Cue trial has been highlighted for illustrative purposes. Top: terminated sequences are shortened by two trials and therefore induce a prediction error (predicted sequence > presented sequence). Bottom: extended sequences are prolonged by two trials and therefore induce a prediction error (predicted sequence < presented sequence). (C) Local transition probabilities for terminated, regular, and extended sequences depending on respective block uncertainty. Note that only the composition of expectation compliance levels varied with uncertainty while cue validity itself remained fixed across blocks.

Undisclosed to the participants, two colours were used as cues to indicate the onset of ordered sequences: one colour marked the first digit of a *short* ordered sequence (regular length of five digits), a second colour marked the first digit of a *long* ordered sequence (regular length of seven digits). Each

participant was assigned two individual cue colours which did not belong to the same hue. Cue validity was fixed at $p = .80$ throughout the experiment with invalid cues being followed by a random figure instead of the next higher one (as would have been suggested by the cue). Neither one of the cue colours nor the red hue used for the colour sequences appeared during random trials. Importantly, the colour cues were employed to trigger predictions with regard to the length of the to-be-observed sequence and were not analysed as events of interest themselves. Implicit cues were used to control for mere attentional effects and to keep participants focussed on the numeric information (instead of the colour information) while tracking the digits for regularities.

In order to induce prediction errors based on the implicit information conveyed by the colour cues, ordered sequences were manipulated in terms of their *expectation compliance.* This factor was introduced to distinguish between *regular*, *terminated*, and *extended* ordered sequences. Given the regular length of five and seven figures, respectively, terminated sequences were shortened by two items (e.g. three instead of five figures for short ordered sequences) while extended sequences were equally prolonged by two items (e.g. seven instead of five figures for short ordered sequences). A graphic display of expected sequence length and expectation compliance is shown in Figure 2-1B.

Finally, the composition of regular, terminated, and extended sequences within a particular block was varied across blocks. This way, the *irreducible uncertainty* of the blocks (i.e. the frequency of change with regard to cue-based expectations) was set to be either *high* or *low*. Blocks of low uncertainty featured local probabilities of $p_{reg} = .70$ for the regular configuration and $p_{term/ext} = .15$ for both terminated and extended sequences. These blocks could therefore be seen as statistically stable regarding cue-based expectations. Local probabilities for terminated and extended sequences were always identical for both sequence lengths. Highly uncertain blocks, in contrast, corresponded to a more unstable statistical structure regarding the expectation based on the cue. Local probabilities

after the cue during highly uncertain blocks were p_{reg} = .40 and $p_{term/ext}$ = .30. It is important to note that at any point during the experiment the regular configuration was the statistically most likely continuation of the sequence. An overview of the statistical structure following the uncertainty factor is provided in Figure 2-1C, please see the Supplementary Material for details on the composition of experimental blocks.

Starting points of all sequences were balanced across digits. Successions of random trials were equally distributed within a range of 5 – 9 digits (M = 7.19 digits) both within and across blocks. To ensure an even distribution of colour presentations, a total of 29 different colours were used. The colours employed in the present study had been validated in a behavioural pilot study (n = 18) to ensure equal visibility of all colours.

Tasks

Participants were asked to indicate detection of ordered sequences as well as colour sequences by corresponding button presses. Therefore, they were instructed to press the left button with their right index finger as soon as they noticed an ordered sequence and to hold the button for the duration of the sequence. Accordingly, release of the left mouse button was to indicate the end of the ordered sequence (i.e. the onset of subsequent random trials). Equivalently, participants were instructed to press and hold the right mouse button with their right middle finger to indicate detection of a colour sequence. Once again, release of the right mouse button marked the end of a colour sequence. A mouse was used in the behavioural sections (i.e. introductory trials, training, and post-measurement) to closely match responses with the two-button response box used during the fMRI session (see below).

Experimental procedures

The study was conducted on two consecutive days. The first day appointment was laid out as a training session in order to allow participants to familiarise themselves with the task and to provide them with implicit knowledge concerning the cues and the underlying statistical structure of the experiment. Importantly, at no point during the training or the fMRI session was it revealed that there was informational content in some of the colours (i.e. the cues) or that the blocks varied in their respective statistical structure (i.e. their level of uncertainty). The second day included the fMRI session as well as a subsequent post-measurement. The experiment was programmed and run using the Presentation 14.9 software (Neurobehavioral Systems, San Francisco, CA, USA).

Day 1 (training session)

After having been informed about the very general scope of the study ("digit processing", supposedly), participants completed a first introduction to the task. Participants were shown a 1 min stream of digits including exactly one presentation of each ordered sequence configuration (i.e. terminated, regular, and extended sequences of both lengths), one invalid cue of each sequence length as well as one colour sequence. Length range and proportion of random trials were matched with the behavioural and the fMRI experiment. Responses were not recorded during the introductory trials and participants were allowed to repeat the introduction until they felt comfortable with the task.

The training consisted of two blocks (one block of high and low uncertainty, respectively) with a total duration of approx. 12 min – (for more details, see section "Stimulus material"). Block order was balanced across participants. Length range and proportion of random trials were matched with the fMRI experiment. After the completion of a block, participants were encouraged to take some time and grant themselves a short break before continuing with the next block.

Day 2 (fMRI session and post-measurement)

Participants once again completed the introductory phase from the previous day in a quiet working environment immediately before entering the scanner. The following fMRI session consisted of four blocks (two blocks of both high and low uncertainty, respectively) with a total duration of approx. 24 min. Contrary to the training, participants were presented with a screen notifying them of a short break for ten seconds after completion of a block. Experimental procedure and task during the fMRI session were otherwise identical to the training session.

Following the functional scanning, participants completed a behavioural post-measurement in order to assess their implicit knowledge of the cue information. To this end, they were presented with one experimental block (duration approx. 5 min) shown on a computer. Length range and proportion of random trials were matched with the training session and the fMRI experiment. Participants were asked to perform the identical task as before (i.e. to indicate sequence detection by button press). Crucially, only half of the ordered sequences were cued by the same colours as during the training and the fMRI session. The other half began with fixed but different colours that had indeed been presented during training and fMRI, but not as cues for the respective participant. Therefore, they contained no implicitly learned information concerning upcoming trials. As with the established cue colours, two previously non-informative colours were assigned to mark the beginning of short and long ordered sequences, respectively.

Finally, participants were interviewed verbally to assess whether they were aware of any regularity at all with regard to the digits' colours. All participants denied having noticed any colour-related regularity.

Behavioural data analysis

Statistical analyses of behavioural responses were performed using R statistical software (R Foundation for Statistical Computing, Vienna, Austria). If not stated otherwise, an α-level of .05 was defined as a statistical threshold.

First, correct and incorrect responses were aggregated separately for training, fMRI session, and post-measurement for each participant. Incorrect responses were further divided into misses (no response over the course of a sequence) and false alarms (response occurring without presentation of sequential trials). Participants' overall performances were assessed via the discrimination index PR (Snodgrass & Corwin, 1988), defined as the difference between hit rate and false alarm rate: correctly reported sequences (i.e. ordered or colour sequences) relative to all sequences was defined as the hit rate. The false alarm rate was defined as falsely reported sequences (again, ordered or colour sequences) relative to all sequences. No specific timeout criterion was defined for the onset of button presses, i.e. responses were registered throughout the whole length of the respective sequence.

Reaction times for button presses (onset latency) and releases (offset latency) were assessed for fMRI session and post-measurement. Latencies were aggregated separately for the levels of *expectation compliance* (terminated, regular, extended) and *uncertainty* (high, low) as well as for established vs new cue colours, respectively. Aggregation was executed for each participant individually. Onset latency was calculated as reaction time relative to the onset of the second trial of any particular ordered sequence: the second trial of the sequence was the earliest possible point to detect a sequential pattern, since the current trial always had to be compared to the preceding one (i.e., to check whether the figure had risen by 1). Offset latency was calculated as reaction time relative to the onset of the first random trial after any particular sequence. Repeated-measures analyses of variance (ANOVA) and paired *t*-tests were used to assess possible differences in offset (depending on expectation compliance

and uncertainty) and onset latency (learned vs new cue colours during post-measurement), respectively. Where appropriate, results of the paired *t*-tests were corrected for multiple comparisons at $p = .05$ using the false discovery rate (fdr) correction by Benjamini & Hochberg (1995).

Functional data analysis

fMRI data acquisition and data preprocessing

Functional and structural imaging data were collected using a 3T Siemens Magnetom Prisma MRI scanner (Siemens, Erlangen, Germany) equipped with a 20-channel head coil. Participants lay supine with their right hand placed on a two-button response box. Index and middle finger were placed on the two response buttons, matching the response contingencies from the training session. Participants' arms were stabilised on form-fitting cushions and foam padding around the head was applied to prevent motion artefacts. Earplugs and noise-cancelling headphones were provided to reduce scanner noise.

During functional imaging, 30 x 4 mm axial slices (1 mm spacing, 64 x 64 voxel matrix, 192 x 192 mm field of view, resulting voxel size 3 x 3 x 5 mm) were acquired parallel to the bi-commissural line (AC-PC) using a single-shot gradient echo-planar imaging (EPI) sequence sensitive to BOLD contrast (TR = 2000 ms, TE = 30 ms, 90° flip angle, ascending recording, 800 repetitions). Prior to the functional session, a high-resolution structural scan was recorded for each participant using a standard Siemens 3D T1-weighted whole brain MPRAGE imaging sequence (1 x 1 x 1 mm voxel size, TR = 2130 ms, TE = 2.28 ms, 256 x 256 mm field of view, 192 sagittal slices).

Data processing was done with the Lipsia software package (Lohmann et al., 2001). Functional data were spike-corrected (using interpolation with adjacent time points) to reduce artefacts within time series. Correction for slice acquisition time (using cubic spline interpolation) and head motion (3 translation, 3 rotational parameters) was applied and functional data were co-registered with the

structural scan (using rigid transformation). Individual structural scans were normalised to the MNI template via general affine transformation and resulting parameters were applied to the functional scans. The resulting normalised functional images were resampled to 3 mm isotropic voxels, high-pass filtered with a 100 s period cutoff and spatially smoothed with an 8 mm full-width half-maximum (FWHM) Gaussian kernel.

fMRI analysis

Event-related BOLD responses were estimated in a general linear model (GLM) approach. The GLM was constructed to test for distinct neural correlates of the different error types as well as for effects of statistical block structure on neural processing at different time points during ordered sequences. Additionally, the parametric effect of surprise was modelled to control for trial-by-trial variation in stimulus improbability. Therefore, the model comprised a total of seven regressors of interest reflecting the 3 x 2 combination of the factors *expectation compliance* (terminated, regular, extended) and *uncertainty* (low, high) plus the surprise parameter. Regressors of nuisance included experimental breaks, colour sequences, and motor responses (button presses and releases) in order to account for variance unrelated to the events of interest.

For terminations, the event onset was time-locked to the first unexpected *random* digit (i.e. the fourth [short sequences] or sixth [long sequences] sequential position, respectively; see Fig. 2-1B). Equivalently, extensions were modelled with the onset time-locked to the first unexpected *sequential* digit (i.e. the sixth [short] or eighth [long] sequential position, respectively). Regular events (termed *checkpoints*) were defined as points in time at which we hypothesised the incoming stimulus to be checked for either a termination (i.e. a check occurring during the ongoing sequence) or an extension (i.e. a check at the regular end) of the ordered sequence. Importantly, checkpoints were only classified as such when the ordered sequence was in fact continued as indicated by the cue, that is, in the regular

configuration (see Fig. 2-2 for an example and Supplementary Fig. 2-S1 for details on event specification). In case the current stimulus did not match the prediction, the event was classified as a prediction error instead of a checkpoint.

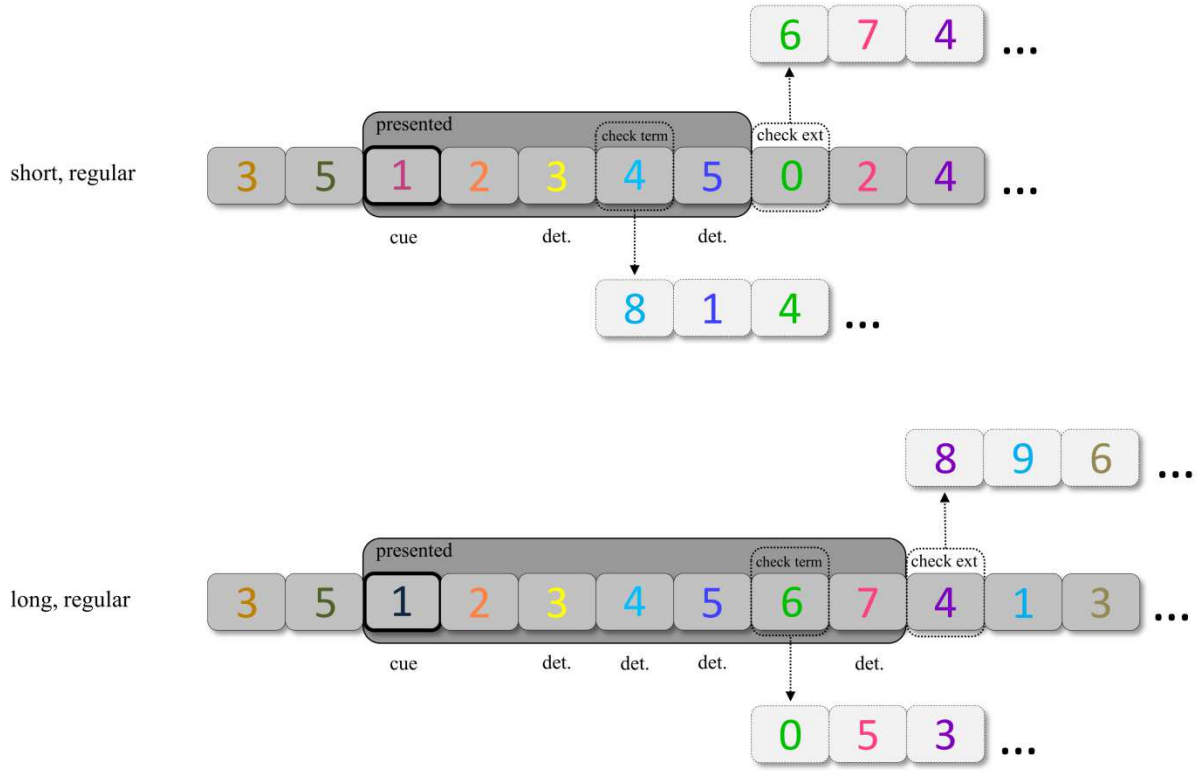

Figure 2-2. Illustration of checkpoints during a regular short and long ordered sequence, respectively. As indicated by the cue (highlighted for the purpose of this graphic), five or seven consecutive digits following the +1 rule are presented. Two checkpoints are hypothesised over the course of each sequence: For instance, during short ordered sequences (top), the fourth item can be used to check whether the sequence is terminated (check term) or not. A hypothetical course of a terminated sequence is outlined (lower trial succession, dashed framing). The sixth item (i.e. the first random figure after the regular short ordered sequence) can be used to check for an extension of the sequence (check ext). A hypothetical course of an extended sequence is outlined (upper trial succession, dashed framing). Within a particular ordered sequence, checkpoints are always preceded by deterministic items (det.).

With respect to difference characteristics, one could justifiably argue that the difference between checkpoints and prediction errors may not at all be qualitative, but rather a quantitative excess of prediction errors with regard to their respective improbabilities. From this point of view, checkpoints could be construed as muted prediction error signals with a lesser degree of surprise. While surprise inarguably modulated cognitive processing in the present task, our key point of suggesting a

functional role for checkpoints was to gain insight into qualitatively different processing of regular events depending on the particular context. By including a parametric surprise regressor, we attempted to discriminate between these qualitative effects and quantitative distinctions that can be ascribed to mere differences in improbability.

The parametric effect of surprise was estimated following the notion of an ideal Bayesian observer (see Harrison et al., 2006). Event-specific surprise $I(x_i)$ was defined as the improbability of event x_i, i.e.

$$I(x_i) = -\ln p(x_i)$$

with

$$p(x_i) = \frac{n_j^i + 1}{\sum_k n_k^i + 1}$$

where n_j^i denotes the total number of occurrences of outcome j (*terminated, regular, extended*) up to the current observation i relative to the sum of all past observations (with k for all possible outcomes).

Contrast specification

The main effect of *expectation compliance* was assessed by the second-level contrast of terminated and extended sequences (TERM > EXT). Common activations of both prediction error types were assessed by the conjunction of the two single contrasts vs checkpoints, respectively (i.e. TERM > CHECK ∩ EXT > CHECK). Note that due to the probabilistic structure of the task, there were considerably more checkpoints than prediction errors contributing to the respective regressors (ratio ~ 2:1). Therefore, for the conjunction of TERM > CHECK and EXT > CHECK, we constructed a parallel GLM identical to the one described above, with the one exception that only half of the

checkpoint events were included in the respective regressors of the second model. While this approach resulted in reduced power of the respective contrasts and their conjunction, its merit lies in a markedly well-balanced, less biased contrast of conditions.

The effects of *uncertainty* on checkpoint and prediction error processing were assessed by the contrasts of high vs low uncertainty checkpoints (CHECK_HIGH > CHECK_LOW) and prediction errors (PE_HIGH > PE_LOW, data not shown), respectively.

For group level analyses, one sample t-tests were calculated using first-level contrast images of all participants. Resulting t-values were converted to z-scores and thresholded at voxel-wise $p < .001$. In a second step, this initial thresholding was combined with a cluster-extent based threshold derived from Monte Carlo Simulation (see Forman et al., 1995). We ran 5000 iterations using the fMRIMonteCluster tool (available at github.com/mbrown/fmrimontecluster), yielding a cluster-extent based threshold of $k > 783$ mm^3 (29 contingent voxels, cluster-level $p < .05$).

Exploratory functional connectivity analyses

Eigenvector centrality mapping

When investigating the interplay of brain regions in forming coactivation networks, functional connectivity measures have become increasingly popular to complement BOLD amplitude effects of participating brain regions. Eigenvector centrality mapping (ECM) has been proposed by Lohmann and colleagues (2010) as a graph-based means to determine the centrality of neural structures within their respective networks. Particularly, eigenvector centrality (Bonacich, 2007) refers to how strongly a certain structure (a *node,* in terms of graph theory) is functionally connected to other highly interconnected nodes. Therefore, both number and quality of connections are factored into the centrality value assigned to a particular voxel. Since the interesting aspect here was precisely to

evaluate the influence of irreducible uncertainty on the neural processing within task-related networks (whose components are themselves highly interconnected), ECM was explicitly well suited for the functional connectivity analysis at hand. Importantly, ECM does not depend on a priori assumptions and, due to computational efficiency, allows for functional connectivity analyses of the entire brain. Finally, ECM does not require parameter adjustments other than the definition of the voxel space and the time period of interest.

In order to assess potential effects of uncertainty on functional connectivity, eigenvector centrality was analysed post-hoc within a whole-brain mask (≈ 60 000 voxels). To avoid connectivity changes caused by the mere duration of scanner time (see Lohmann et al., 2010), ECM analysis was restricted to the first half of the experiment (i.e. one block of each uncertainty level in counterbalanced order) for each participant. Pairwise similarity matrices for time series of any two voxels were computed and subsequently analysed by the ECM algorithm. On the group level, a one-sample *t*-test was used to assess whether the difference between the centrality maps for high and low uncertainty were significantly greater than zero across participants. Resulting t-values were converted to z-scores and thresholded at $p < .001$ (10 contiguous voxels).

Beta series correlation

As a follow-up analysis on the ECM approach, we used a beta series correlation analysis (Rissman et al., 2004) to assess trial-by-trial covariation of activity in our regions of interest. In short, trial-specific beta series within ROI were correlated for each condition on the group level. For a step-by-step description of the analysis, the reader is referred to the Supplementary Material.

2.4 Results

Behavioural results

fMRI session

Participants showed an overall high level of performance with a mean PR score of $\mathcal{M}_{PR}$ = 0.89 (*SD* = 0.06) during the fMRI session, indicating good attentiveness throughout the experiment. Out of 176 detectable sequences (not counting invalid cues), participants correctly responded to 165.73 events on average (*SD* = 12.91). Mean PR scores (Fig. 2-3A) did not differ significantly between experimental blocks ($F(3, 76) = 1.23$, $p = .30$), nor as function of uncertainty ($t(21) = 0.21$, $p = .84$). As stated above, two participants were excluded from further analyses due to their behavioural performance (standardised PR scores of z_{PR} = -1.79 and z_{PR} = -2.47, respectively).

The repeated-measures ANOVA yielded a significant main effect of expectation compliance on offset latency ($F(2, 38) = 45.75$, $p < .001$, Greenhouse-Geisser-corrected). Post-hoc pairwise *t*-tests revealed participants' button releases to be significantly slower after terminated (M = 783.72 ms, SD = 77.94 ms) than after regular (M = 713.92 ms, SD = 76.91 ms, fdr-adjusted $p < .01$) as well as after extended sequences (M = 549.94 ms, SD = 121.73 ms, fdr-adjusted $p < .001$). The difference between extended and regular sequences was significant as well (fdr-adjusted $p < .001$). This pattern of offset latency differences appears intuitive in the sense that premature terminations unexpectedly violated the prediction of continued sequential input, thus inducing a delayed response compared to the regular condition. Critically, neither the main effect of uncertainty ($F(1, 19) = 0.16$, $p = .70$) nor the interaction term of *uncertainty X expectation compliance* ($F(2, 38) = 0.16$, $p = .85$) reached statistical significance, suggesting that participants were able to discriminate regular from manipulated sequences regardless of the respective level of uncertainty. The number of misses ($t(19) = -0.58$, $p =$

.57) and false alarms ($t(19) = 0.24$, $p = .81$) did not differ significantly between high and low uncertainty blocks (see Fig. 2-3A).

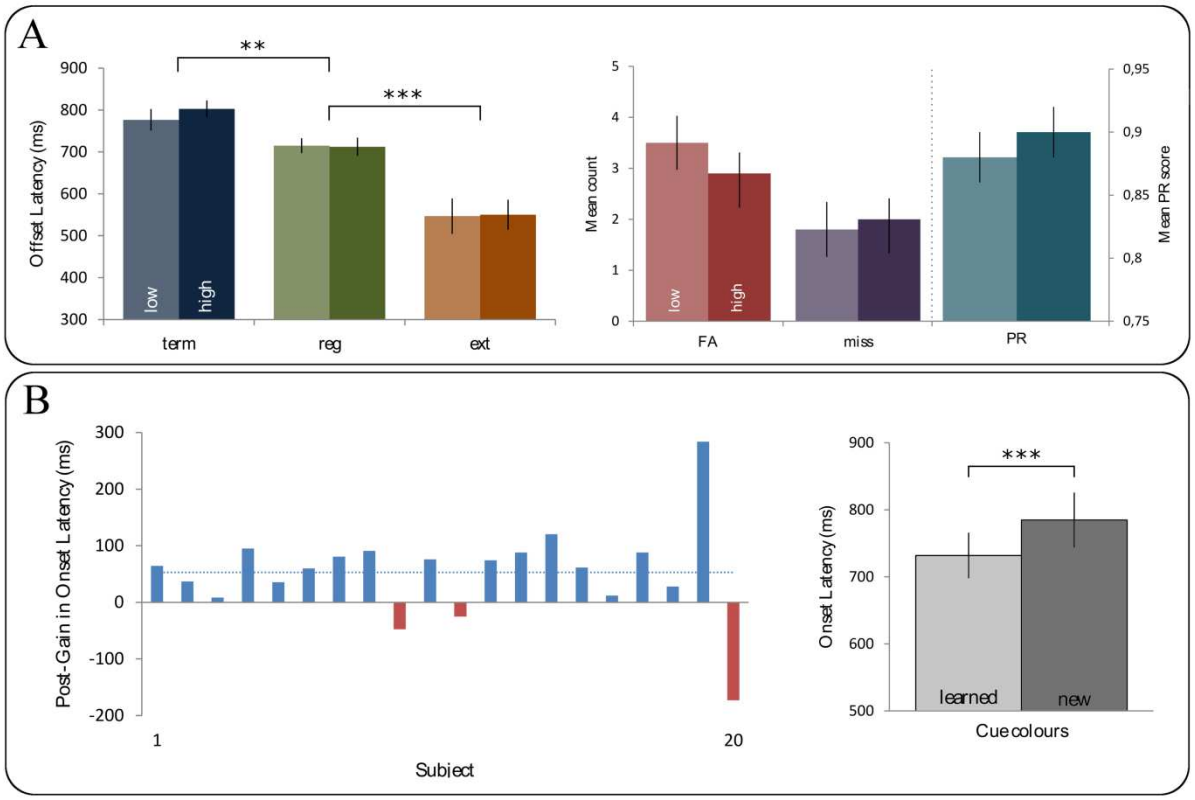

Figure 2-3. (A) Left panel: Effects of expectation compliance (terminated, regular, extended) and irreducible uncertainty (low, high) on mean offset latency. Right panel: Mean count of false alarms and misses as well as the mean PR score as a function of uncertainty. (B) Left panel: Differences in onset latency between new and learned cue colours during post-measurement for each subject. Blue bars indicate faster onset for learned cue colours, dotted line depicts mean gain across subjects. Right panel: Overall mean onset latency for learned and new cue colours, respectively. Error bars show standard error of the mean (SEM). ** = $p < .01$, *** = $p < .001$

Post-measurement

Participants performed equally well during the post-measurement ($M_{PR} = .90$, $SD = 0.06$) as they had during the fMRI session. Out of 40 detectable sequences (not counting invalid cues), participants correctly responded to 38.77 (SD = 2.02) sequences on average.

The post-measurement was conducted in order to assess accessibility of the signalling information provided by the cues. If participants had learned the association of cue colours and prospective ordered sequences over the course of the training and the fMRI session, they were expected to react faster to sequences beginning with established cue colours than to those starting with new colours during the post-measurement. Indeed, the corresponding *t*-test confirmed a significant difference between learned and new cue colours ($t(19) = -2.78$, $p = .006$, one-tailed). Participants exhibited a shorter reaction time (defined as onset latency relative to the second sequential stimulus, see above) to learned cue colours ($M = 731.76$ ms, $SD = 152.68$) than to new cue colours just introduced during the post-measurement ($M = 784.78$ ms, $SD = 184.04$; see Fig. 2-3B).

fMRI results

Supporting our first hypothesis, group-level activations discernibly related to sequential terminations (TERM > EXT) were found in left inferior frontal gyrus (IFG). In contrast, sequential extensions (EXT > TERM) were found to be distinctly reflected in activations across an extensive network comprising posterior cingulate cortex (PCC), right superior frontal sulcus (SFS), and right angular gyrus (Table 2-1).

Table 2-1. Activation peaks, z-values, and anatomical locations for the main effect of expectation compliance (TERM > EXT) and the conjunction of the two prediction error types vs checkpoints, respectively (TERM > CHECK ∩ EXT > CHECK).

	local maxima				
	MNI				
area	x	y	z	z-value	volume (mm^3)
TERM > EXT					
left IFG / BA44	-45	7	26	3.73	3402
right PCG	36	-29	71	4.59	2916
EXT > TERM					
right anterior SFS	27	58	23	-3.94	4212
right STG	48	-5	-4	-4.13	2349
right PCC/BA31	15	-32	41	-3.90	2025
right BA40	60	-35	44	-4.85	11394
left BA7	-21	-62	71	-4.52	1512
left STG	-51	1	-7	-4.46	5103
(TERM > CHECK) ∩ (EXT > CHECK)					
right putamen	30	-5	-1	4.02	891

IFG = inferior frontal gyrus, BA = Brodmann Area, PCG = postcentral gyrus, SFS = superior frontal sulcus, STG = superior temporal gyrus, PCC = posterior cingulate cortex.

As expected, uncertainty had a significant effect on checkpoints processing (CHECK HIGH > CHECK LOW). Contrary to our hypothesis, however, enhanced activation at sequential checkpoints during blocks of high uncertainty was not found in ACC, but in bilateral IFG and the corresponding projection area in parietal cortex, right angular gyrus / temporoparietal junction (rANG/TPJ, Fig. 2-4B; see Table 2-2 for coordinates).

Table 2-2. Activation peaks, z-values, and anatomical locations for the effect of irreducible uncertainty on checkpoint processing (CHECK_HIGH > CHECK_LOW) and the parametric effect of surprise.

	local maxima				
	MNI				
area	x	y	z	z-value	volume (mm^3)
CHECK_HIGH > CHECK_LOW					
right IFG / BA44	33	16	26	4.14	1377
right TPJ / angular gyrus	36	-65	41	3.94	1458
SURPRISE					
right dorsal insula	30	28	2	4.60	2781
right ACC	0	13	29	3.74	1431
right BA40	60	-44	26	3.77	1431
right BA7	15	-47	65	3.67	837
left dorsal insula	-42	10	-1	4.46	6183
left BA4/BA6	-45	-8	56	3.97	891

IFG = inferior frontal gyrus, BA = Brodmann Area, TPJ = temporoparietal junction, ACC = anterior cingulate cortex.

The conjunction of terminations and extensions relative to checkpoints (i.e. unexpected relative to expected events, TERM > CHECK ∩ EXT > CHECK) yielded significant activation in the right putamen (Fig. 2-4C; see Table 2-2 for peak coordinates). Conceptually, this joint activation reflects the shared portion of prediction error processing (i.e. an unexpected violation of the internal model) in distinction to the qualitative differences in how the internal model is violated (i.e. having to prematurely neglect vs unexpectedly resume the sequence model, see above).

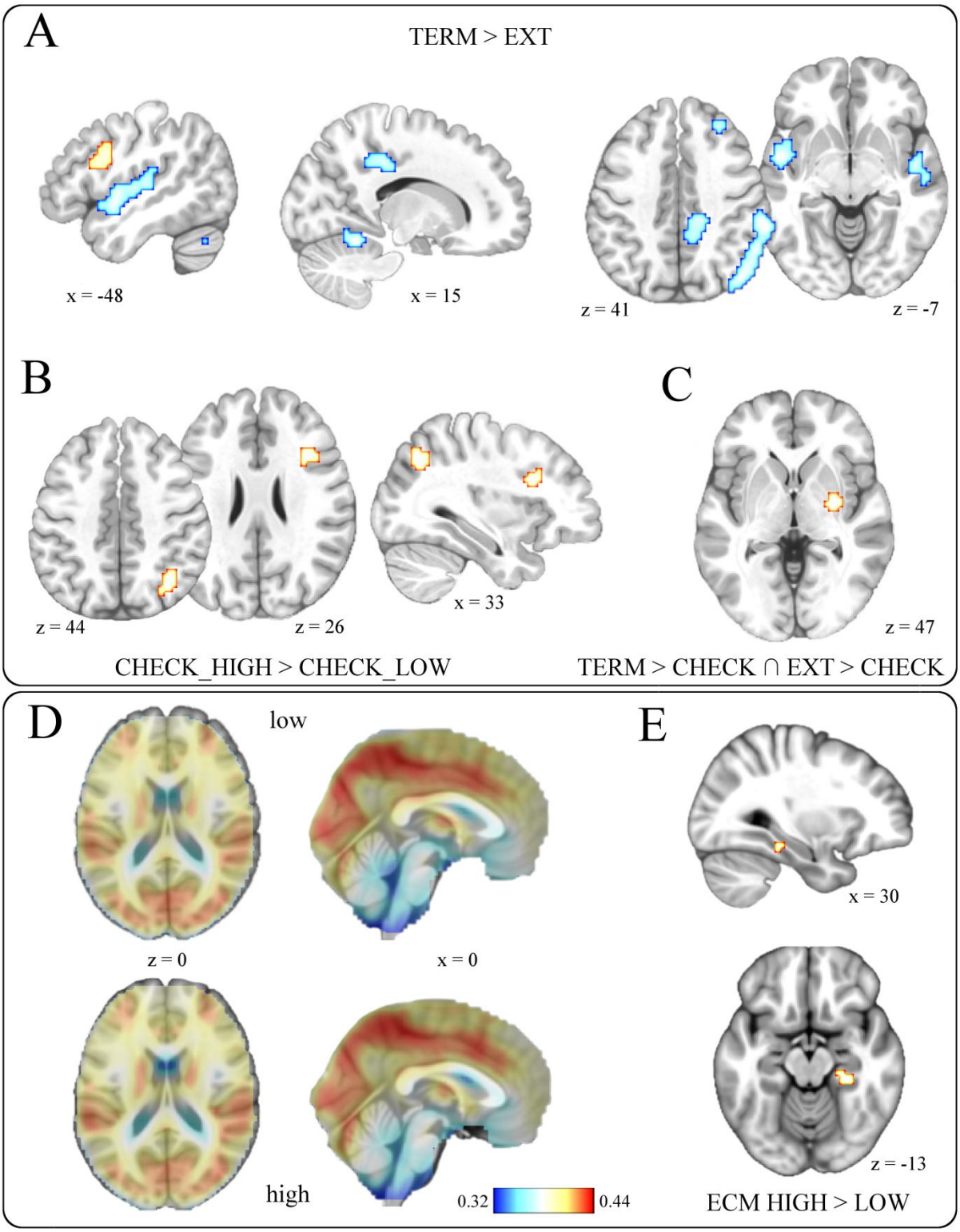

Figure 2-4. (A) Areas positively correlated with sequential terminations (red) and extensions (blue). (B) Areas reflecting neural processing at sequential checkpoints for high (vs low) uncertainty. (C) Common activations of terminations and extensions relative to checkpoints. (D) Group averages of eigenvector centrality for low and high uncertainty blocks. (E) Significantly higher centrality within the parahippocampal region for high > low uncertainty as revealed by a pairwise t-test.

The surprise parameter modulated neural responses in a widespread network (Table 2-2) including bilateral anterior-dorsal insular cortex, right anterior cingulate cortex (ACC), and right fusiform gyrus (see Supplementary Fig. 2-S2).

Exploratory eigenvector centrality mapping

Based on the finding that irreducible uncertainty modulated neural processing of sequential checkpoints, we used eigenvector centrality mapping (ECM) as an exploratory tool to further analyse uncertainty effects. Adding to the reported modulations in BOLD amplitude, ECM as a measure of functional connectivity was employed to detect uncertainty-related differences on the neural network level.

Both high and low uncertainty conditions equally displayed a network of highly interconnected nodes including cingulate cortex, precuneus, basal ganglia, inferior frontal gyrus, and visual areas (Fig. 2-4D). Intriguingly, significantly higher eigenvector centrality was observed in the right parahippocampal region (MNI coordinates 27, -35, -13) for the high uncertainty condition, indicating stronger functional connections between parahippocampal areas and the other central nodes (see above) during time periods of statistical instability (Fig. 2-4E).

Beta series correlation

Significant trial-by-trial correlations between our regions of interest were found to be exclusively positive. Uncertainty-induced changes in network connectivity appear to differ between conditions: For terminations and extensions, there were several significant links between areas under high uncertainty that were absent under low uncertainty (see Fig. 2-5). This does not seem to be the case for checkpoints for which all network components were significantly correlated under low uncertainty. PHC beta series under high uncertainty were found to be significantly correlated with those of virtually all other network components for terminations and checkpoints, but not for

extensions. Moreover, whereas the increase in significant PHC connections for high > low uncertainty is most evident for terminations (with no significant correlation of PHC under low uncertainty), PHC was highly interconnected even under low uncertainty at checkpoints and thus showed no increase under high uncertainty. Finally, PHC was not significantly correlated with any other network component at extensions (under neither uncertainty level).

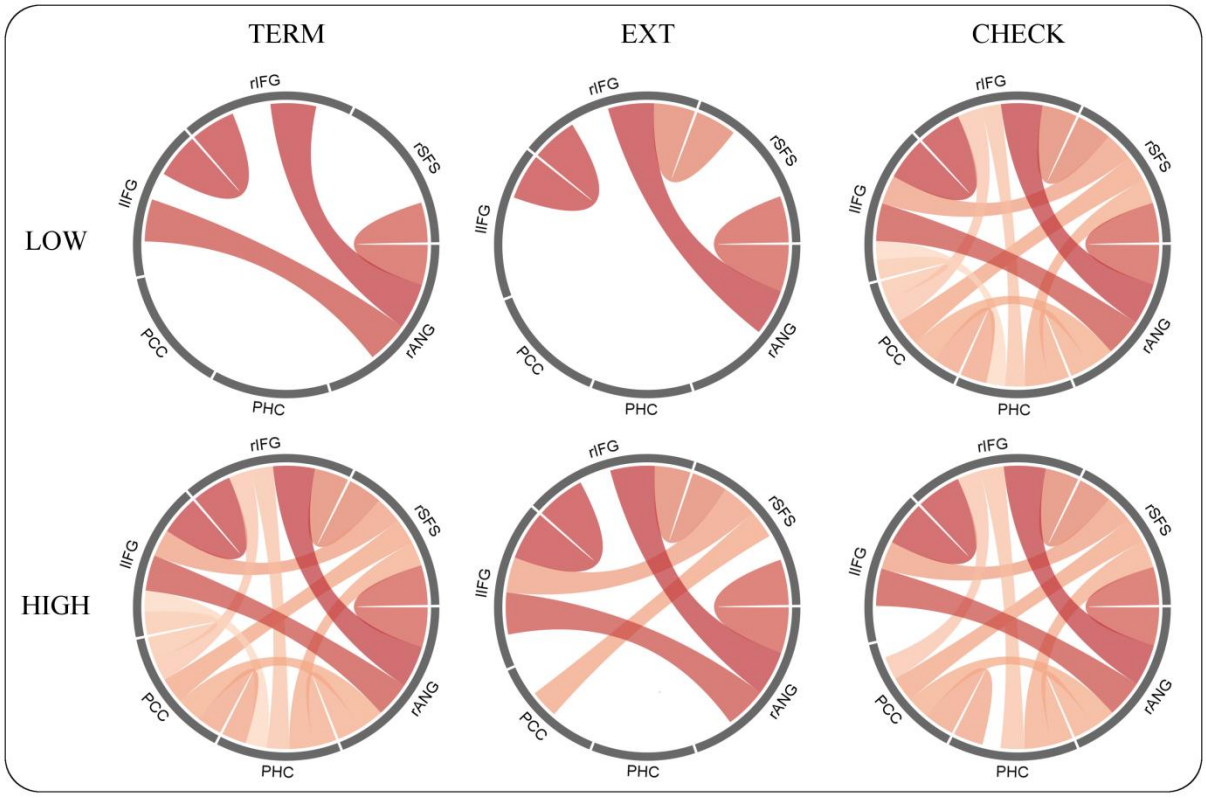

Figure 2-5. Connectivity patterns within the network identified in the fMRI contrasts. Significant correlations of trial-specific beta series for terminations (TERM), extensions (EXT), and checkpoints (CHECK) are shown for both low and high uncertainty. Links coloured in red depict positive correlations, with darker tone corresponding to higher r. lIFG = left inferior frontal gyrus, rIFG = right inferior frontal gyrus, rSFS = right superior frontal sulcus, rANG = right angular gyrus, PHC = parahippocampal cortex, PCC = posterior cingulate cortex.

Correlation of behavioural and functional data

We hypothesised predictive processing to be adapted to statistical properties of environments (in this case, irreducible uncertainty). Conceivably, this strategic adaptation could manifest on the behavioural level - successful adaption to highly uncertain contexts should facilitate resolution of conflict under uncertainty and thus result in response time advantages at corresponding events. As

we propose such adaptation to originate from more pronounced processing at checkpoints under high uncertainty, the uncertainty -related behavioural measures reported so far may very well have been too non-specific to reflect particular effects on reaction times. To closely link behavioural measures to the reported BOLD effects, we correlated subject-level regression weights from the CHECK_HIGH > CHECK_LOW contrast (i.e. rIFG and rANG, see above) with individual differences in offset latency between high and low uncertainty blocks. Importantly, since checkpoints were exclusively sampled from regular sequences for the fMRI contrasts, equally did only button releases at the end of regular sequences contribute to the reported differences in the correlation analysis. Both rANG ($r(18) = .21$, $p = .19$, one-sided) and rIFG ($r(18) = .12$, $p = .31$, one-sided) were found to show a small but non-significant correlation with faster reaction times in highly uncertain blocks (Fig. 2-6A).

In a second brain-behaviour correlation analysis, we assessed whether individual differences in successful learning of the cue-length-contingency (operationalised as quicker reactions to learned vs new cue colours in the post measurement, see Fig. 2-3B) were differentially related to activity in rIFG and rANG (Fig. 2-6B). The difference between the two correlations ($r(18) = .35$, $p = .07$ and $r(18) = .12$, $p = .31$, one-sided) was found to be non-significant ($z = 1.01$, $p = .16$, one-tailed).

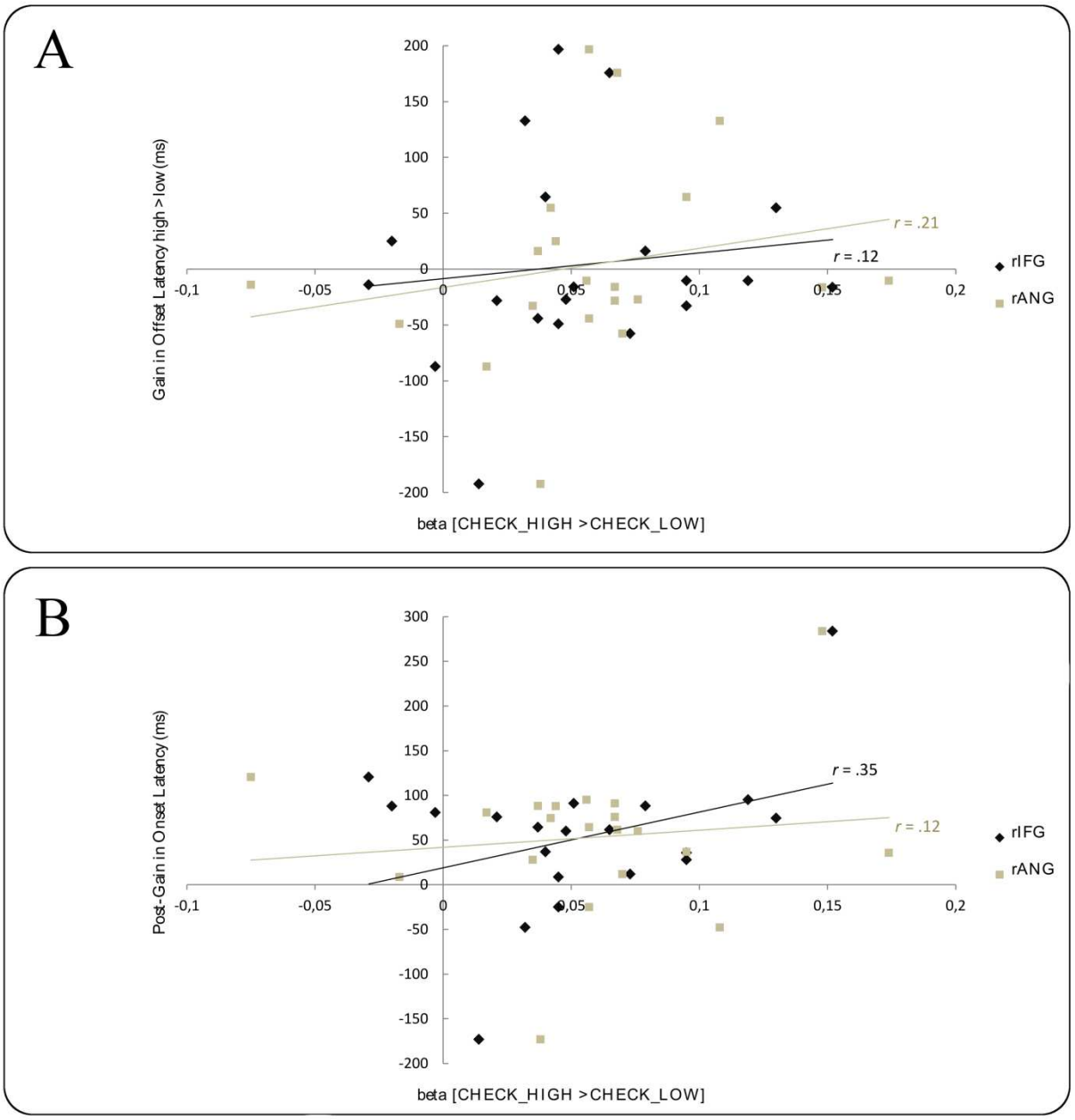

Figure 2-6. (A) Correlation of beta weights extracted from the contrast of high vs low uncertainty checkpoints and the difference in offset latency following regular sequences in high vs low uncertainty environments. Differences in offset latency are depicted as 'gain under high uncertainty, i.e. faster reaction times for button releases at the end of regular sequences under high (vs low) uncertainty. (B) Correlation of beta weights extracted from the contrast of high vs low uncertainty checkpoints and the difference in onset latency for learned vs new cue colours during the post test. Differences in onset latency are depicted as 'gain during post measurement', i.e. faster reaction times for button presses at the beginning of ordered sequences following learned (vs new) cue colours.

2.5 Discussion

The present fMRI study was conducted to investigate the strategic adaptation of predictive processing to different non-reward prediction error qualities and to the prediction's contextual uncertainty. Distinguishable activation patterns were elicited by prediction errors depending on the respective error type, lending support to a differential concept of non-reward prediction errors. In particular, we found that unexpected terminations and extensions of predicted stimulus regularities elicited increased activity in distinct brain networks. Moreover, increased activity at checkpoints for high vs low irreducible uncertainty suggests that context stability affects predictive strategies. Notably, this effect was controlled for the quantitatively variable surprise level (i.e. respective improbability) of the sampled event types. Potentially, stable (i.e. low uncertainty) contexts allowed a full-length prediction of sequential input based on the internal model, whereas highly uncertain contexts induced iterant comparisons of sensory information with the internal model. Neural processing at checkpoints was more pronounced in these unstable contexts, possibly pointing towards a stepwise (rather than full model length) prediction. Although further research is needed to link neural effects to changes on the behavioural level, these results provide novel insight into potentially adaptive prediction strategies and their respective neural underpinnings.

Qualitative differences of prediction errors

Predictions of digit sequences could be violated by termination or extension of the expected sequence length. While both types of prediction error (relative to checkpoints) commonly elicited enhanced activity in right-lateralised putamen, their direct contrast yielded distinct frontal/ frontoparietal activity specific to the respective prediction error type. While we tested a specific hypothesis for this contrast (IFG for terminated sequences), we discuss further findings to suggest testable hypotheses for future studies.

Terminated sequences correlated with neural activity within IFG (BA44, see below) whereas extensions were reflected in effects along the anterior portion of right SFS (lateral BA8/9). Evidence from the action observation literature has implicated the SFS in the processing of event boundaries (Schubotz et al., 2012), i.e., behaviourally relevant transition points in event perception. In line with the authors' interpretation of boundary-related SFS activity as a correlate of updating attention to the next stimulus, prediction error signals caused by sequential extensions presumably reflect the violation of the (expected) sequence ending: at the point of a sequential extension, participants were presented an unpredicted sequential digit when in fact expecting a random digit denoting the end of the ordered sequence (vice versa for terminated sequences). In other words, present correlates of extended sequences supposedly express a prediction error signal that ultimately results in memory-directed reorientation of attention, i.e., the resumption of the internal model.

This interpretation is substantially supported by coactivation of the angular gyrus, another component frequently associated with attentional reorientation towards salient or informative stimuli (Gottlieb, 2007; Kincade et al., 2005; Rushworth et al., 2001). The angular gyrus, specifically the dorsal portion we report for sequential extensions, has been shown to be connected to superior frontal areas via the occipitofrontal fascicle (Makris et al., 2007; Nelson et al., 2010). Functionally, one suggested role for the angular gyrus in attentional updating has been the integration of current stimuli with recent task history (Taylor et al., 2011) - a highly relevant operation for extended but not for terminated sequences. In a similar vein, O'Connor and colleagues (2010) reported expectancy violations in an attentional cueing task to be reflected in supramarginal and angular gyrus. Their differential findings of prediction errors following old vs new items led the authors to assume that IPL lesions should affect cognitive control mechanisms especially when unexpected familiar items violate a strong expectation of novel stimuli. Our results further substantiate these suggestions, as the

characteristic quality of extended sequences was precisely the observation of history-conform sequential digits when participants expected novel (i.e. random) digits.

In contrast to sequential extensions, *one specific* digit (i.e. the preceding number raised by one) was expected at the point of violation during terminated sequences. Therefore, sequential terminations can be considered violating a more specific prediction than was the case for sequential extensions. Our finding of IFG activity increase for sequential terminations adds to consistent reports of BA44 reflecting violations of expected regularities in language syntax (Friederici & Kotz, 2003), musical structure (Maess et al., 2001), actions (Wurm & Schubotz, 2012), and abstract stimuli (Huettel et al., 2002). Contrary to an unexpected *continuation* of observed regularities (see above), these studies and our own results commonly point to a role for IFG in processing premature rule violations when regular input of certain specificity is expected.

Predictive processing as a function of statistical context

The second main aim of the present study was to assess predictive strategies in varying statistical contexts. Processing of changes in statistical regularities per se has been demonstrated for both reward (Behrens et al., 2007) and non-reward paradigms (Tobia et al., 2012). However, it remains unclear whether or not predictions of abstract upcoming input strategically change with context (note that here, "strategically" does not imply a conscious effort). To dissociate predictions in a stable vs unstable context, the reasoning was as follows: while an implicit cue might trigger the prediction of sequential input at full length in a stable context, an unstable context might lead to a stepwise processing confirming the prediction is still valid. Recall the staircase analogy from before: in an unstable context (e.g. in dim lighting), one might be well advised to verify the initial prediction ("This flight has 13 steps") at some critical point in order not to encounter a (potentially precarious) prediction error.

Following this rationale, we investigated effects of irreducible uncertainty on the neural processing of possible checkpoints during prediction. Checkpoints were defined as regular events at those sequential positions where prediction errors occurred in terminated or extended sequences. Note that, by definition, violations of the cue – length contingency did not occur in regular sequences from which checkpoints were sampled. Right IFG showed increased BOLD activity at checkpoints within the high (vs low) uncertainty condition. Given the central involvement of prefrontal cortex (PFC) in flexible interactions with the environment, a process oftentimes termed *cognitive control* (Corbetta & Shulman, 2002; Petrides, 2000), enhanced prefrontal responses at *potential* violation sites might indeed be interpreted as an updating mechanism accounting for changes of statistical regularities. Previous work has shown that experimental context can be decoded from prefrontal neurons (Waskom et al., 2014), suggesting that cognitive control and decision-making benefit from representations of current context encoded in PFC. Conceivably, the use of specific vs higher-order information for context encoding in PFC depends on the respective network in which frontal sites are coactivated. Recall that left-lateralised IFG activation was found for violations of number-specific expectations: Consolidating IFG effects for both sequential terminations and highly uncertain checkpoints, the common role for prefrontal sites may be a close monitoring of structured incoming information. Depending on whether these monitoring processes lead to the detection of a specific prediction error (as in the TERM > EXT contrast) or provide vital information about the current task context (i.e. CHECK_HIGH > CHECK_LOW), respective network partners are coactivated accordingly. Support for this interpretation comes from our finding that checkpoint processing under high uncertainty was also found to be reflected within the right angular gyrus / TPJ. As a direct projection site of IFG, TPJ has been established as part of a ventral network engaged in attentional control (Corbetta & Shulman, 2002; see Cabeza et al., 2008 for review). Specifically, this network initiates a bottom-up reorientation driven by behaviourally relevant but unattended stimuli.

Fittingly, demanding the actor to enter a special attentional mode is one characteristic of junctures from the field of sequential action selection (*decision points*; Reason, 1992). This way, increased allocation of attentional resources may be one plausible interpretation of rANG / TPJ effects for high uncertainty checkpoints. However, an intriguing hypothesis redefining TPJ function was put forward by Geng and Vossel (2013): the authors propose contextual updating as the main role of TPJ in cognitive processing, meaning the updating of internal models of context based on new sensory information (Seghier, 2013). It is important to recall that the contrast of interest (*CHECK_HIGH > CHECK_LOW*) only contained events of regular sequences where a violation of prediction (termination or extension, respectively) was probable but did in fact not occur. Since there was no external signal initiating a change in predictive strategies, enhanced processing of checkpoints for highly uncertain contexts was solely based on the internal sequence model developed through previous experience. Consequently, contextual updating as well as corresponding strategic adjustments (i.e. the use of incremental predictions in unstable environments) do not seem to require a bottom-up trigger signal but can instead be prompted by model-based expectancies alone. This suggests an extension of the understanding of TPJ functioning for paradigms where contextual updating relies on top-down generated internal models of context.

Due to the higher average surprise value for checkpoints in high vs low uncertainty blocks, an increase in surprise could potentially present an intuitive explanation for the reported results. Crucially, however, since the surprise parameter was modelled separately within the GLM, our findings cannot be attributed to stimulus-bound surprise but instead reflect higher-order cognitive processes exceeding trial-by-trial variation in informational value.

In sum, elevated BOLD responses within rIFG and rANG at sequential checkpoints potentially point to an adaptive quality of abstract predictive processing that to our knowledge has not been

demonstrated. Depending on the stability of statistical contexts, the top-down predictive strategy may vary with regard to how far expectations reach into the future. Supporting this interpretation on the behavioural level, the extent to which participants learned the colour-length association (i.e. their gain in response speed during post-measurement) was more strongly correlated with the BOLD uncertainty effects in the frontal (rIFG) than in the parietal (rANG) component of the checkpoint network (although the difference in correlation did not reach significance). In other words, especially for rIFG, the more successful the contingency between cue colour and sequence length was learned, the more pronounced was the neural activity at checkpoints under high uncertainty (see Fig. 2-6).

Motivated by the reported BOLD amplitude effects of uncertainty, exploratory functional connectivity analysis (ECM) revealed eigenvector centrality of the right middle temporal gyrus (MTG) / parahippocampal region (PHC) to increase during epochs of high uncertainty. This reflects a strengthened connectivity between the parahippocampal region and those highly interconnected frontoparietal circuits that are central for the currently employed task. More detailed analyses of trial-specific beta series within our regions of interest (see Supplementary Material) suggest differential connectivity patterns of PHC depending on the outcome of a sequence and the respective level of uncertainty: While PHC was found to be highly interconnected at checkpoints (regardless of uncertainty), the same level of connectivity was found only for high uncertainty terminations. Lastly, PHC activity at extensions did not covary significantly with any other network component. One possible hypothesis would be that representations of terminations and checkpoints could be more similar than, say, terminations and extensions; an intriguing starting point for multivariate analyses of representational similarity (RSA; Kriegeskorte et al., 2008). Another possibility would be to look at network configurations at different points in time and formulate hypotheses about graph measures (e.g. network density) to learn more about the respective roles of key regions implicated in predictive

processing under uncertainty. These results motivate intriguing hypotheses and objectives for future research targeting network configuration as a function of uncertainty (see Limitations and Future Directions).

The medial temporal lobe including the hippocampal formation has been established as a crucial structure for prospective processing and pattern completion during perception (Turk-Browne et al., 2010), sequence learning (Schapiro et al., 2014; Schendan et al., 2003), and especially at points of ambiguity (Bornstein & Daw, 2012; Kumaran & Maguire, 2006; Ross et al., 2009). When multiple competing predictions with regard to the next sequential item arise under uncertainty, the MTG network is hypothesised to resolve this ambiguity, resulting in enhanced hippocampal responses (for review, see Davachi & DuBrow, 2015). Critically, the present high uncertainty condition was hypothesised to intensify neural processing at ambiguous checkpoints by means of a higher proportion of sequential deviants. Therefore, the finding of strengthened functional connectivity between the parahippocampal region and central task-related structures in a highly uncertain environment further corroborates the understanding of MTG as a contingency-sensitive circuitry engaged in encoding and extracting statistical information. Our results moreover suggest that this information could then be employed to facilitate a context-appropriate change in predictive processing, namely a stepwise prediction strategy for highly uncertain contexts. This way, instability in the environment could supposedly be compensated by more frequently comparing model-based expectations with actual sensory input. Combining the contextual updating hypothesis and MTG involvement in sequential ambiguity, enhanced overall connectivity of MTG / parahippocampus under high uncertainty could thus be interpreted as evaluating the validity of top-down, model-based predictions in light of incoming sensory information (Kumaran & Maguire, 2006; Lisman, 1999). Presumably, a stepwise prediction is not necessary in stable contexts, possibly due to frequent

validation of the internal model (i.e. the high proportion of regular sequences) resulting in high confidence in the initial full-length prediction. As for computational efficiency and economic processing, stepwise prediction would appear to be a strategic adaptation to unstable contexts in order to avoid the cost of prediction errors. One intuitive neural implementation of these stepwise predictions would be through enhanced communication between MTG and frontoparietal networks, thus enabling recourse to model information in working memory. Indeed, previous studies have demonstrated the anatomical connectivity between angular gyrus/TPJ, associated frontal areas (IFG / lateral PFC), and MTG (Clower et al., 2001; Makris et al., 2005; Petrides & Pandya, 1999; Vincent et al., 2006). Within the scope of the contextual updating hypothesis, these connections are thought to integrate internal representations of current context information with the appropriate sensorimotor transformation necessary to respond adequately (Geng & Vossel, 2013). Strengthened MTG involvement in contexts where regular updating is beneficial suggests that the internal model is iteratively checked based on cue information retrieved from working memory.

Limitations and Future Directions

The present study, conceptualised as a first step into investigating adaptive qualities in predictive processing, does not come without limitations. As mentioned above, one critical objective for future efforts is to replicate the differential effects for terminations and extensions in an experimental design that ensures equal response requirements for both PE types. Even though the direct contrast (TERM > EXT) does not contain components primarily associated with motor function, identical behavioural correlates would allow further analyses of response patterns, e.g. assessing modulatory effects of surprise on offset latency. Furthermore, subsequent efforts could aim to uncouple terminations and extensions from fixed sequential positions, thus eliminating potential influence of timing or sequence length. Future directions also include the use of multivariate approaches, as single

run measurement of fMRI data limited the applicability of multivariate analyses for our present study. Follow-up studies could then use tools such as representational similarity analysis to decode neural representations of the internal model at different points in time (see Schuck et al., 2015 for preparatory mPFC encoding prior to strategy change). Such multivariate approaches could also be combined with EEG data (currently in preparation) to provide novel information on the time course of representations as a function of statistical context.

2.6 Conclusion

Different classes of abstract prediction errors were reflected in distinct brain activation patterns, predominantly within separate frontoparietal networks. Depending on whether the respective prediction error called for reorienting towards external stimuli or staying with the internal model, cognitive processing was adjusted accordingly. High irreducible uncertainty resulted in more pronounced processing of sequential checkpoints we preliminarily interpret as iterative comparisons of sensory information and the internal model. Although further research is needed, our findings suggest that this stepwise predictive strategy may be conducted through enhanced connectivity between frontoparietal circuits and the (para-)hippocampal area.

2.7 Acknowledgements

We would like to thank Christiane Ahlheim, Laura Quante, Ima Trempler, and Leon Windscheid for their valuable comments on earlier drafts of this manuscript, as well as Sarah Fromme, Irina Kaltwasser, and Monika Mertens for their help during data collection. Finally, we are thankful for the insightful and productive comments by our reviewers.

2.8 Supplementary material

Stimulus material

Both high and low uncertainty blocks included five invalid cues for each sequence length. Furthermore, each block included four colour sequences which served as a motor control condition (fixed length of five figures). Participants were informed that colour sequences would occur much less frequently than ordered sequences. Each block included 20 ordered sequences for either sequence length, the composition of which depended on the irreducible uncertainty level of the respective block: blocks of low uncertainty comprised 14 regular, three terminated, and three extended sequences of each sequence length. Highly uncertain blocks contained eight regular, six terminated, and six extended sequences of each sequence length. The proportion of trials which were part of any sequence (as opposed to being part of a random trial succession) was fixed at $p_{event} = .45$ across all blocks (including the post measurement, see below).

Since only one block of trials was presented during the post-measurement, the composition of this block was designed to match an approximate average of the two established levels of uncertainty (high, low). The post-measurement comprised ten regular, four terminated, and four extended ordered sequences of each sequence length, therefore corresponding to a medium level of uncertainty ($p_{reg} = .55$, $p_{term/ext} = .22$). In addition, four invalid cues of each sequence length as well as four colour sequences were presented.

Detailed task instructions

Participants were informed that the experiment's objective was to detect ordered digit sequences. They were instructed to track a continuous stream of coloured single digits in the middle of the screen and to perform two tasks: First, to indicate a regularly ascending sequence (such as 3 – 4 – 5 – 6 – 7) by pressing the left mouse button and to keep the button pressed until the ordered sequence ended. Second (and not as frequently), to indicate a sequence of digits continually presented in red by pressing the right mouse button and to keep the button pressed until a digit was presented in a different colour again. Participants were then made aware of the inclusion of the 0 character, so that successions such as 8 – 9 – 0 – 1 – 2 were legitimate ordered sequences. Finally, participants were asked to respond both as quickly and as correctly as possible when they detected either an ordered sequence or a colour sequence. The two participants with the best task performance were promised gift certificates for a popular online marketplace as an incentive to perform to the best of their ability.

Covariation in univariate contrasts

Methods

To gain further insight into functional connectivity patterns between the key regions yielded from our fMRI contrasts, we extracted regression weights from the TERM > EXT as well as from the CHECK_HIGH > CHECK_LOW contrast. In order to avoid circularity (see Kriegeskorte et al., 2009) we used the *leave-one-out* approach: We created functional regions of interest (ROI) by calculating the group-level contrast considering *n*-1 participants. Group-level contrasts were corrected for multiple comparisons as described in the Methods section of our manuscript. Regression weights were then extracted on the subject level for the remaining participant. This procedure was repeated

n times and ensured statistical independence between ROI generation and subsequent analyses of regression weights extracted from that ROI.

In accordance with our interpretation of the eigenvector centrality analysis findings (see the Results section of the manuscript), we attempted to assess the role of the parahippocampal region (PHC) in more detail: The hypothesis-free ECM analysis had revealed higher eigenvector centrality for PHC under high uncertainty (i.e. a long-term effect of uncertainty on PHC connectivity to other highly connected brain areas). In short, we interpreted this increase in functional connectivity as support for the role PHC has been suggested to play in resolving conflict under uncertainty (for more details, see the Discussion section). Conceivably, correlations of activity within PHC and "conflict-specific" brain areas, as defined e.g. by our TERM > EXT contrast, could provide additional support for that interpretation.

Results

Supplementary Fig. 2-S3 shows the respective correlation of regression weights extracted from PHC with those of brain areas reported for the TERM > EXT and the CHECK_HIGH > CHECK_LOW contrast. The former is the contrast of interest in the present analysis, given that terminations and extensions are the two types of prediction errors. Of the brain areas whose activation was found to be significant in the fMRI contrasts (see Results) we report correlations for the left inferior frontal gyrus (lIFG, found for TERM > EXT), right superior frontal sulcus (rSFS), and posterior cingulate cortex (PCC, both found for EXT > TERM). As a control condition, we also correlated PHC activity with that of right inferior frontal gyrus (rIFG) and right angular gyrus (rANG), both reported for the CHECK_HIGH > CHECK_LOW contrast.

Significant positive correlations were found between activity within PHC and frontal components of the networks reported for terminations and extensions: PHC was found to be functionally connected to lIFG ($r(18) = .57$, $p < .01$) as well as to rSFS ($r(18) = .54$, $p < .05$). In contrast, neither PCC nor the two areas from right-lateralised CHECK_HIGH > CHECK_LOW network (rIFG and rANG) showed a significant correlation with respective PHC activity (all $p > .13$).

Discussion

The more involved PE-specific areas were in the direct contrast of terminations and extensions, the more PHC was found to be coactivated with the respective regions of interest. Although this analysis does not account for trialwise variability in the BOLD signal, it points toward a functional connection between PHC and frontal network components reflecting the qualitative difference in prediction errors (i.e. lIFG for terminations and rSFS for extensions). It was our hope that the present results would function as a first step in disclosing the relationship between PHC and other key regions under the modulating influence of uncertainty. To this end, we conducted a beta series correlation analysis to consolidate results from the ECM analysis, univariate fMRI contrasts, and the rather coarse coactivation pattern found therein.

Beta series correlation

Trial-by-trial connectivity patterns of the key regions identified in our fMRI contrasts was assessed using the *beta series correlation* approach (Rissman et al., 2004). BOLD responses were estimated separately for each trial of a given experimental condition (e.g. terminations under low uncertainty, see Fig. 2-S1 for design specifics), yielding a set of trialwise beta weights for each voxel (or an array of voxels). We defined functional regions of interest (ROI) using the leave-one-out procedure described above. Following the concept of functional connectivity as the temporal correlation between event-related neural activations (Friston et al., 1993), these beta series (averaged across voxels within each ROI) were then correlated between our predefined ROIs to assess the degree of their respective connectivity per condition. Thus, we ended up with n (20 subjects) x k (6 conditions) x m (6 ROIs) symmetric correlation matrices. We first applied an arc-hyperbolic tangent transform (Fisher, 1921) to all m x m entries of the k correlation matrices. Given the correlation coefficient's restricted range of [-1, 1], this transformation makes the coefficients' distribution approximate a normal distribution. In accordance with the procedure by Rissman and colleagues, the transformed correlation coefficients were then divided by their known standard deviation $1/\sqrt{N-3}$ to yield z scores (N being the number of data points contributing to the respective coefficient). Mean z-transformed correlation matrices were calculated (averaged across participants) for each condition. Significance was determined by assigning fdr-corrected p-values to the upper triangle of all k z-transformed mean correlation matrices.

Behavioural analysis including the surprise regressor

In order to approximate our fMRI analysis on the behavioural level, we controlled the reported effects of expectation compliance on offset latency for the surprise values of the respective events. To this end, we set up a linear model (Gaussian distribution, identity link function) with the factors expectation compliance, uncertainty, and surprise for each participant and subsequently tested the resulting regression coefficient of surprise for significance. The corresponding *t*-test revealed the effect of surprise on mean offset latency to be not significant ($t(19) = -0.78$, $p = .44$, two-tailed). Please note that regular events were equally sampled during and at the end of regular sequences for the fMRI analysis (see Supplementary Fig. 2-S1). Since only checkpoints at the end of (but not during) regular sequences required a response, comparability of fMRI and behavioural analyses is restricted by varying event selection (see Limitations).

Supplementary Figures

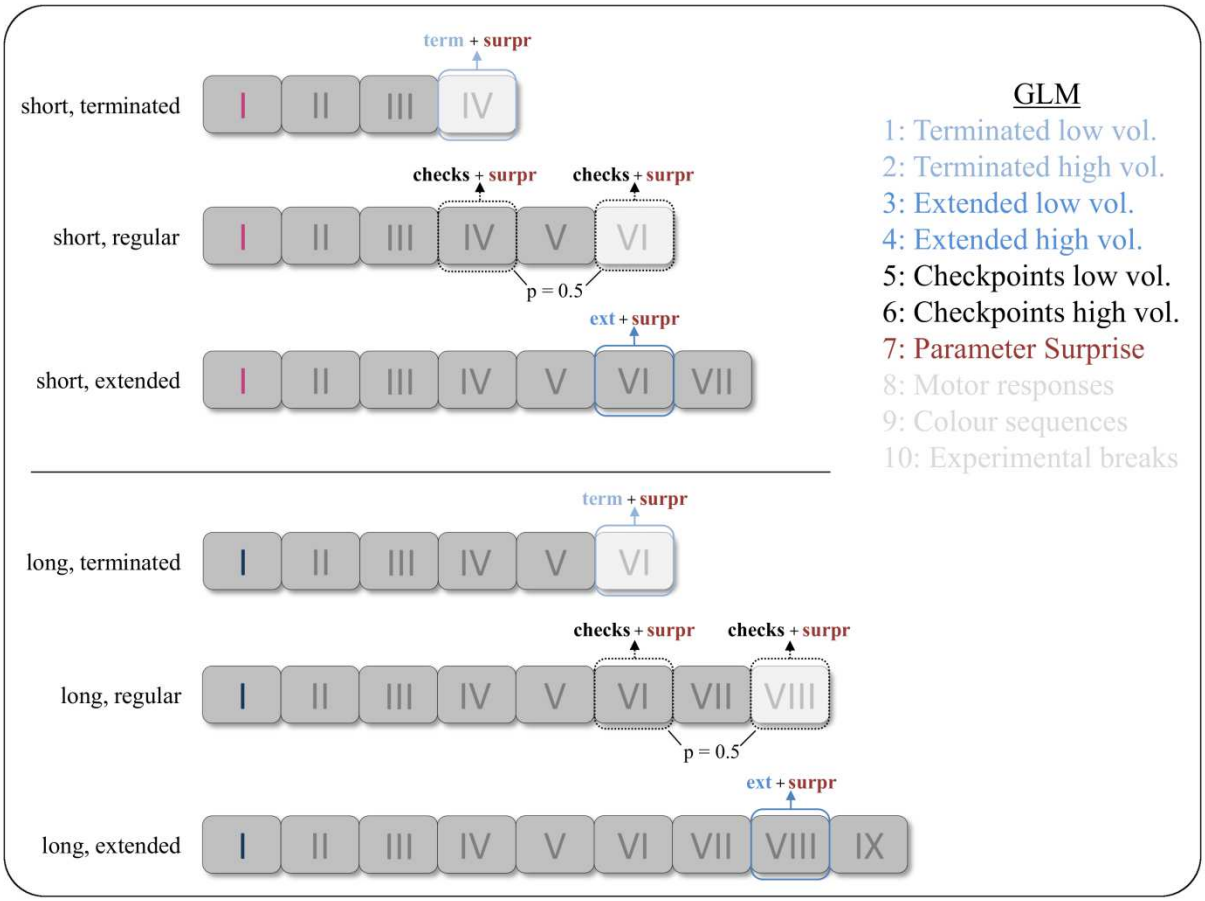

Figure 2-S1. Illustration of event selection for each of the GLM regressors and both sequence lengths. Event selection was identical for both high and low uncertainty. Sequential positions are shown in Roman numerals, light grey numerals depict would-be sequential positions at which factually a random digit was shown. Since the two checkpoint positions of regular sequences occurred within the same TR, only one of two possible events was alternately sampled from regular sequences.

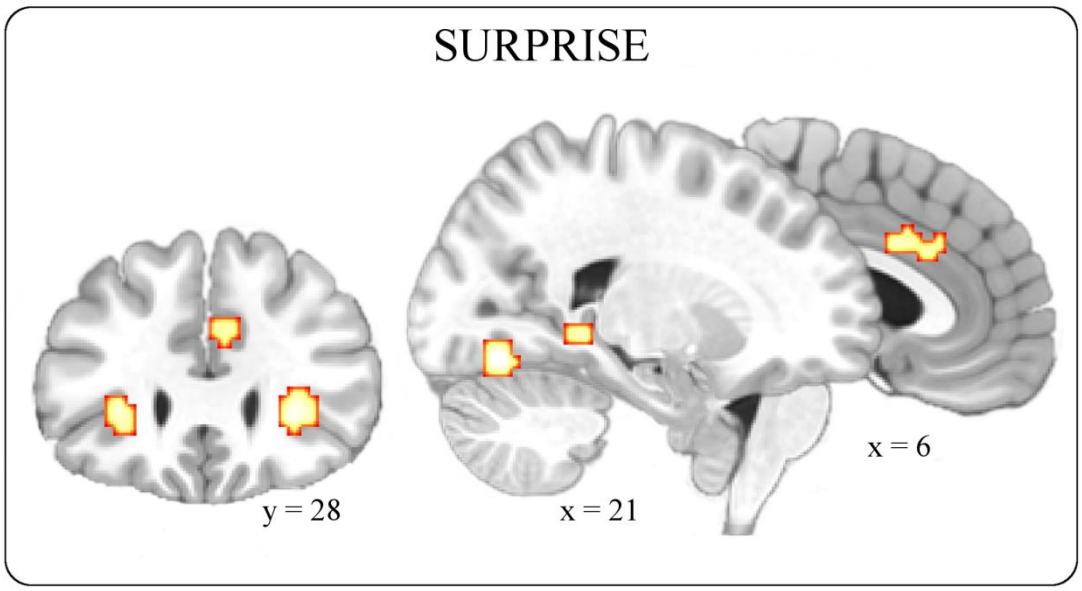

Figure 2-S2. Areas positively correlated with the surprise parameter.

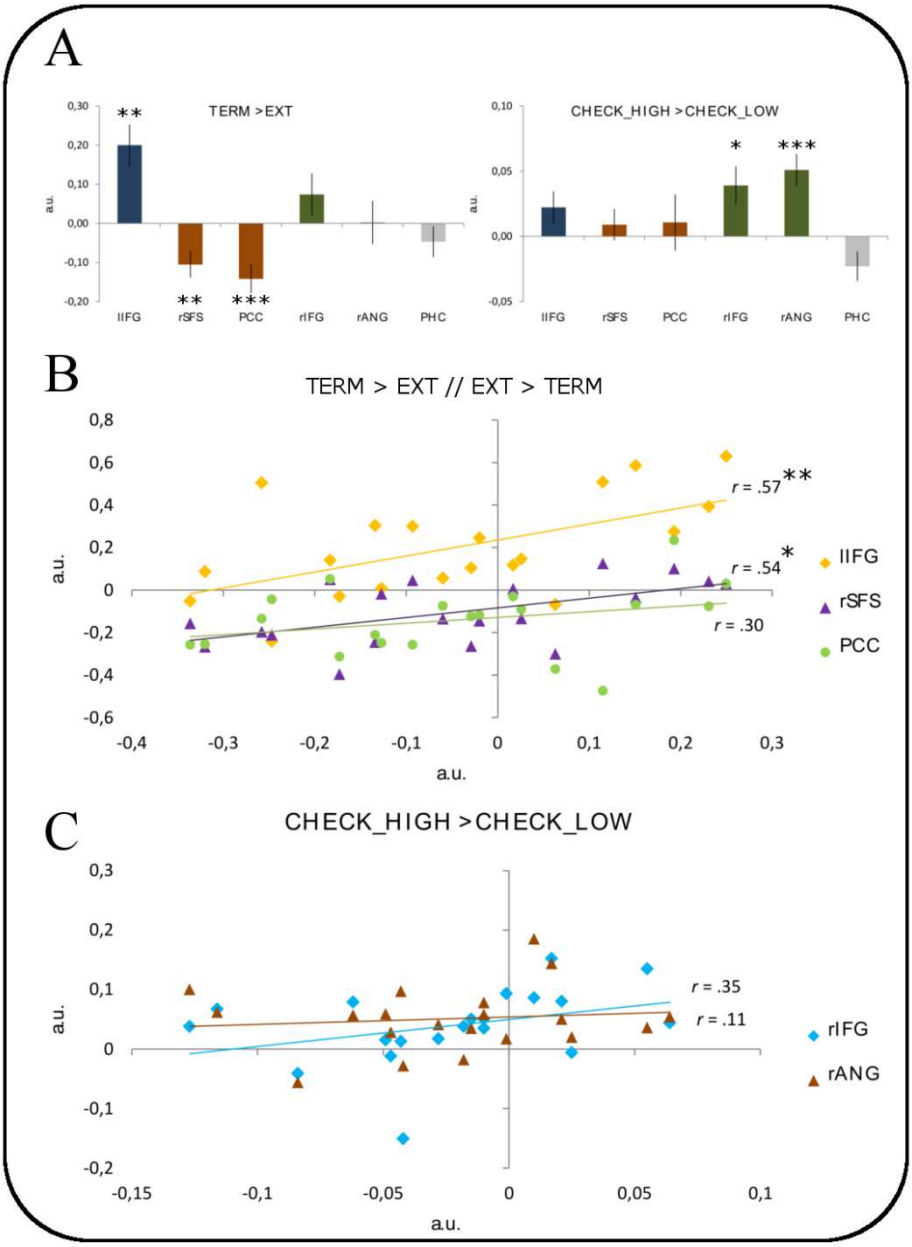

Figure 2-S3. (A) Regression weights from the TERM > EXT (left panel) and the CHECK_HIGH > CHECK_LOW contrast (right panel). (B) Correlation of regression weights from key areas of the TERM > EXT contrast with those of PHC. (C) Correlation of regression weights from key areas of the CHECK_HIGH > CHECK_LOW contrast with those of PCC. lIFG = left inferior frontal gyrus, rSFS = right superior frontal sulcus, PCC = posterior cingulate cortex, rIFG = right inferior frontal gyrus, rANG = right angular gyrus. * = p < .05, ** = p < .01, *** = p < .001.

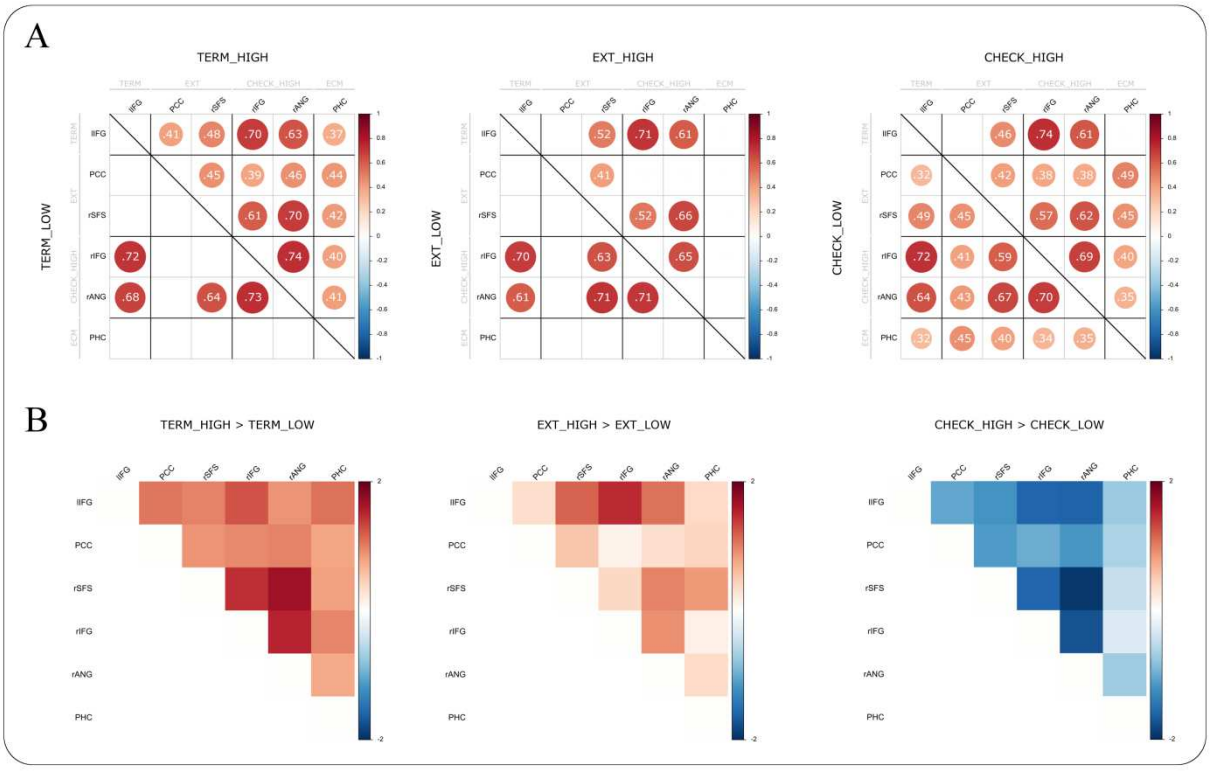

Figure 2-S4. (A). Numeric values of the significant mean correlations between ROI from the beta series correlation analysis (re-converted from Fisher-z scores). Correlation matrices show terminations, extensions, and checkpoints under low (lower triangle) and high irreducible uncertainty (upper triangle). (B) Mean differences of z values under high vs low uncertainty (e.g. TERM HIGH > TERM_LOW). lIFG = left inferior frontal gyrus, rIFG = right inferior frontal gyrus, rSFS = right superior frontal sulcus, rANG = right angular gyrus, PHC = parahippocampal cortex, PCC = posterior cingulate cortex.

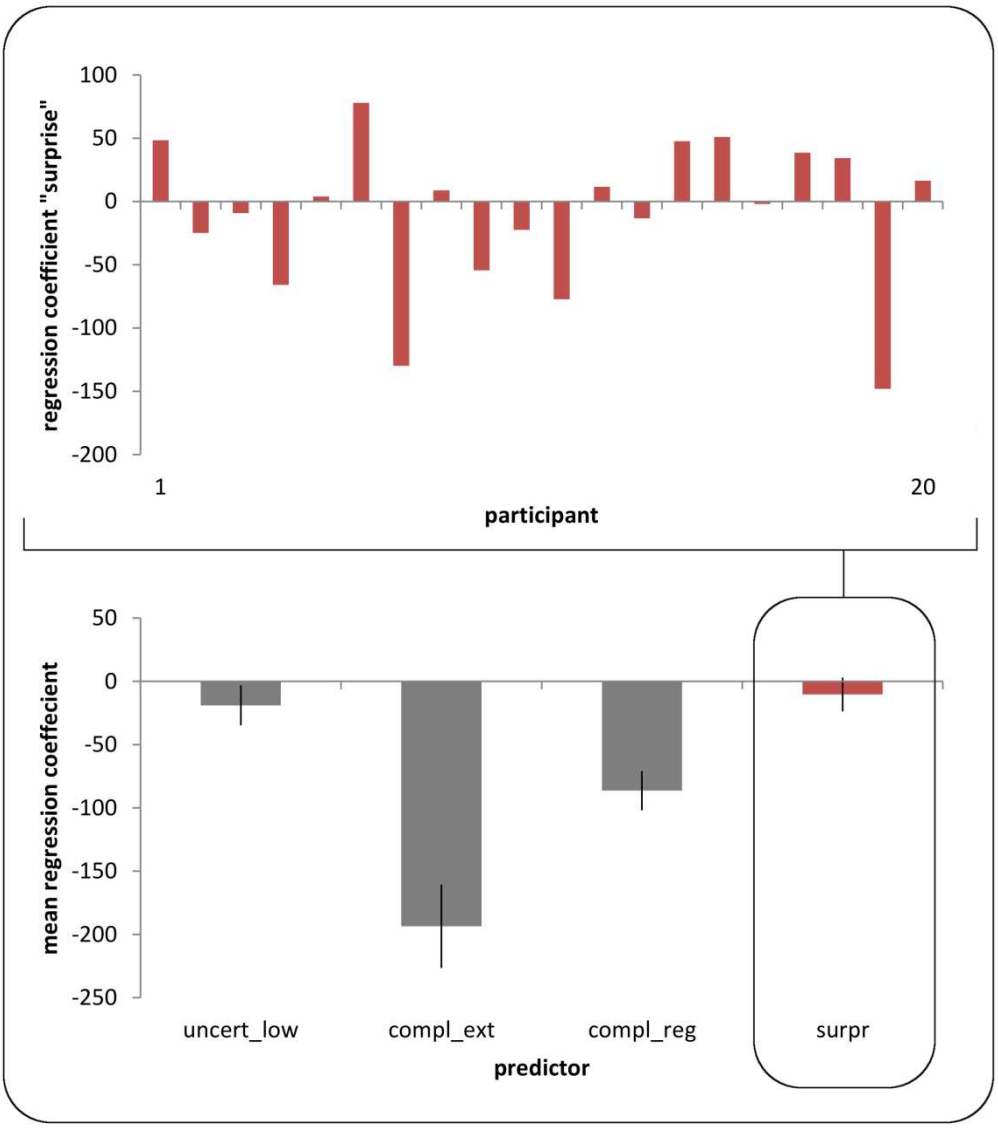

Figure 2-S5. Mean regression coefficients from the GLM including the surprise regressor (lower panel) and individual surprise coefficients (upper panel).

3. Study 2

Being right matters: Model-compliant events in predictive processing

Daniel S. Kluger[a,b], Laura Quante[a,b], Axel Kohler[c], and Ricarda I. Schubotz[a,b,d]

[a] Department of Psychology, University of Muenster, Muenster, Germany

[b] Otto-Creutzfeldt-Center for Cognitive and Behavioral Neuroscience, University of Muenster, Muenster, Germany

[c] Goethe Research Academy for Early Career Researchers, University of Frankfurt, Germany

[d] Department of Neurology, University Hospital Cologne, Cologne, Germany

Correspondence should be addressed to

Daniel S. Kluger
Department of Psychology
University of Muenster
Fliednerstr. 21
D-48149 Muenster
Germany
E-mail: daniel.kluger@uni-muenster.de

Key words: EEG, expectation, prediction error, sequence learning, top-down control

Submitted to: PLoS One (2018)

3.1 Abstract

While prediction errors (PEs) have been established to drive learning through adaptation of internal models, the role of model-compliant events in predictive processing is less clear. Checkpoints (CPs) were recently introduced as points in time where expected sensory input resolved ambiguity regarding the validity of the internal model. Conceivably, these events serve as on-line reference points for model evaluation, particularly in uncertain contexts.

Using electroencephalography (EEG), the aim of the present study was to characterise the functional relationship of CPs and PEs in a serial pattern detection task. Specifically, we first hypothesised a joint P3b component of both event classes to index recourse to the internal model (compared to non-informative standards, STD). Second, we assumed the mismatch signal of PEs to be reflected in an N400 component when compared to CPs. Event-related findings supported these hypotheses. We suggest that while model *adaptation* is instigated by PEs, CPs are similarly used for model *evaluation*. Intriguingly, behavioural subgroup analyses showed that the exploitation of potentially informative reference points may depend on initial cue learning: Strict reliance on cue-based predictions may result in less attentive processing of these reference points, thus impeding upregulation of response gain that would prompt flexible model adaptation.

Overall, present results highlight the role of CPs as model-compliant, informative reference points and stimulate important research questions about their processing as function of learning und uncertainty.

3.2 Introduction

Predicting upcoming events constitutes one of the fundamental qualities of brain function. Based on internal models shaped by previous experience, top-down predictions are compared to bottom-up sensory signals (Rao & Ballard, 1999). Redundant components of perceptual information are disregarded whereas surprising expectancy violations are propagated upward in the processing hierarchy (Friston, 2005; Mumford, 1992). Model adaptation in consequence of such prediction errors (PEs) has been proposed to be the foundation of associative learning mechanisms (Bastos et al., 2012; Rescorla & Wagner, 1972), as unexpected events are particularly informative with regard to their current context. Importantly, probabilistically occurring *expected* events have also been suggested to inform the internal model (Kühn & Schubotz, 2012): While PEs instigate model adaptation, expected events verify model-based predictions. These verifications are particularly informative when we face uncertain environments. A recent fMRI study (Kluger & Schubotz, 2017) found that in uncertain environments, so-called "checkpoints" (CPs) emerged as points in time where distinctive processing of expected events pointed to a context-sensitive adaptation in predictive processing. While the entire stimulus sequence could be predicted reliably in stable environments, unstable environments prompted stepwise predictions. This way, CPs were used to verify the internal model in order to predict the next section accordingly. Thus, while model *adaptation* is induced by PEs, context-dependent model *evaluation* does not seem to require expectancy violations. Instead, selected time points carry information about the on-line validity of the internal model, raising the intriguing question of how CPs and PEs functionally relate to one another.

For the present study, we employed the paradigm from Kluger and Schubotz (2017) in an electroencephalography (EEG) experiment. Exploiting the temporal benefits of EEG, we aimed to further understand the functional relationship of CPs and PEs as well as their respective evolution

over time. Specifically, we aimed to show how functional commonalities of and central distinctions between the two event types translate to electrophysiological signals.

Participants performed a serial pattern detection task in which they were asked to press and hold a response button whenever they detected a short or long ordered digit sequence (e.g. 1-2-3-4-5, length of either 5 or 7 items) within an otherwise pseudorandom stream of coloured single digits. Expectable sequence length was cued by digit colour and occasionally violated by premature terminations or unexpected extensions. In addition to these two types of prediction errors, checkpoints were defined as sequential positions where PE could potentially occur, *but did not*. Thus, although checkpoints were exclusively sampled from regular events consistent with the previous cue, their occurrence was probabilistically modulated by blockwise manipulation of irreducible uncertainty (De Berker et al., 2016; Payzan-LeNestour & Bossaerts, 2011). Going back to our research question, both CPs and PEs provide central information for model evaluation or adaptation, respectively, whereas deterministic standard trials (STD) did neither. Consequently, we first hypothesised a joint event-related (ERP) component of CPs and PEs (compared to STD) reflecting recourse to the internal model. The P3b component has been conclusively shown to co-vary with subjective improbability or unexpectedness of a stimulus (Kutas & Federmeier, 2011; Mars et al., 2008; Seer et al., 2016). Such highly informative events supposedly initiate contextual updating (Kimura et al., 2010; Verleger et al., 2016) or memory-based revision of mental representations (Polich, 2007). Importantly, the P3b is elicited by behaviourally relevant rather than merely deviant stimuli in order to facilitate motor responses (Katayama & Polich, 1998; Nieuwenhuis et al., 2011), making it a promising candidate for a joint physiological component of CPs and PEs.

Aside from the aforementioned conceptual commonalities of CPs and PEs, one critical distinction remains, namely the mismatch signal that is intrinsic to PEs. We hypothesised an enhanced N400 component for PEs to reflect this mismatch signal in contrast to CPs. The N400 amplitude is known

to scale with event surprise in language (for review, see Kutas & Federmeier, 2011) and arithmetic tasks (Niedeggen et al., 1999; Szűcs & Csépe, 2005), presumably marking modality-independent integration of incongruous information.

Complementing ERP analyses, we assessed topographic microstates (Lehmann et al., 1987) for a multivariate, assumption-free comparison of the temporal dynamics underlying CP and PE processing. This way, we aimed to characterise the two event classes using similarities and differences in the onset, duration, and strength of their respective network activation.

3.3 Material and methods

Participants

A total of 32 neurologically healthy, right-handed volunteers (26 female) at the age of 23.4 ± 2.5 years (M ± SD) participated in the study for payment or course credit. Participants were recruited from the university's volunteer database and had (corrected-to-) normal vision. Written informed consent was obtained from all participants prior to the start of experimental procedures. Experimental standards complied with the local Ethics Committee of the University of Münster. One participant was excluded from further data analysis due to poor behavioural performance during the experiment (see below); a second participant was excluded due to technical difficulties during the EEG session. Therefore, all reported analyses are based on a sample of 30 participants (25 female, age 23.2 ± 2.5 years).

Stimulus material

Task and stimulus material of the present study were adopted from a previous fMRI study conducted in our lab (Kluger & Schubotz, 2017). In short, participants were shown pseudorandomly coloured single digits presented for 500 ms in the centre of a light grey computer screen (see Fig. 3-1A).

Presentation frequencies for all colours and digits were equally distributed both within and across blocks of approximately 6 minutes. Each block contained *ordered sequences* increasing the previous digit by one (e.g. 1 – 2 – 3 – 4 – 5; Fig. 3-1A, left) embedded in *random trials* with no discernible relation between consecutive digits. In order to balance sequential starting points across digits, the ascending regularity necessarily included the 0 character and continued in a circular fashion after the figure 9 (e.g. 8 – 9 – 0 – 1 – 2).

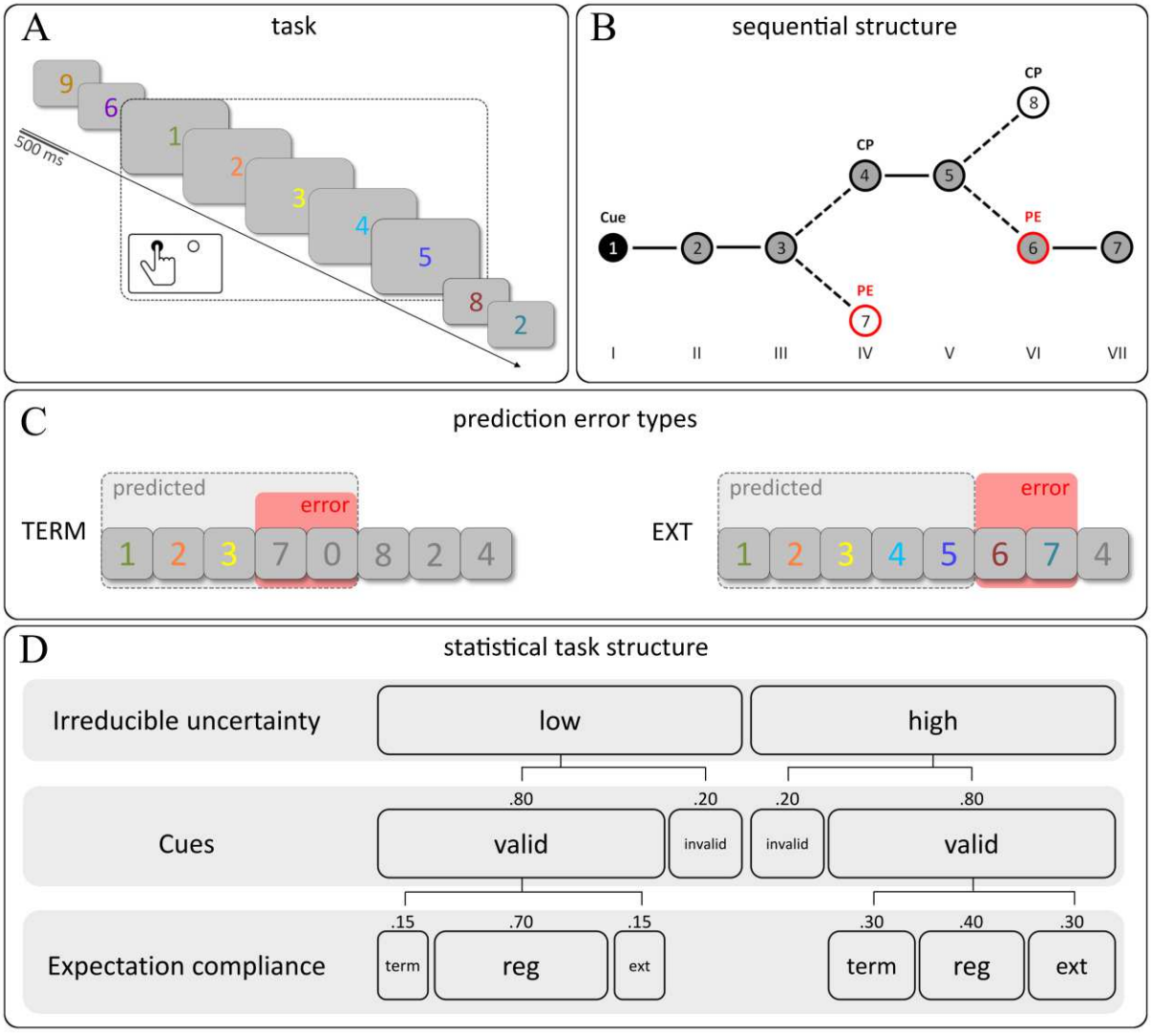

Figure 3-1. (A) Exemplary trial succession and time frame of the corresponding response for ordered sequences. Sequential trials have been highlighted for illustrative purposes. (B) Schematic structure of a short ordered sequence showing the positions of checkpoints (CP) and prediction errors (PE, red). At the fourth position, the sequence could either be terminated (PE) or continued as expected (CP). Similarly, the sixth position contained either the regular end (CP) or an unexpected extension of the sequence (PE). (C) Cue-based expected sequence length and resulting prediction errors for terminated and extended short ordered sequences (*expectation compliance*). (D) Local transition probabilities for terminated, regular, and extended sequences depending on the respective level of irreducible uncertainty.

Undisclosed to the participants, two colours were exclusively used as cues (fixed validity $p = .80$) to indicate the start of ordered sequences: one colour marked the first digit of a *short* ordered sequence (regular length of five digits), a second colour marked the first digit of a *long* ordered sequence (regular length of seven digits). Each participant was assigned two individual cue colours from distinct hues.

Prediction errors were induced by manipulation of the sequences' *expectation compliance*. While the cues indicated the length of *regular* ordered sequences (e.g. seven digits for long ordered sequences), *terminated* sequences were shortened by two items. Conversely, *extended* sequences were prolonged by two items (Fig. 3-1B). Finally, the composition of regular, terminated, and extended sequences within a particular block was varied across blocks. This way, the *irreducible uncertainty* of a block was set to be either *low* or *high* (Fig. 3-1C). Low uncertainty blocks could be seen as statistically stable regarding cue-based expectations whereas highly uncertain blocks formed a more unstable statistical structure. The experiment was programmed and run using the Presentation 14.9 software (Neurobehavioral Systems, San Francisco, CA, USA).

Task

Participants were instructed to press and hold the left button of a response box with their right index finger as soon as they noticed an ordered sequence. Release of the button was to indicate the end of the ordered sequence.

Experimental procedures

The study was conducted on two consecutive days. On the first day, participants completed a training session to familiarise themselves with the task and to provide them with implicit knowledge of the cues and the underlying statistical structure of the experiment. The training consisted of two blocks (one block of low and high uncertainty, respectively) with a total duration of approximately 12

minutes. Importantly, at no point during the training or the EEG session was it revealed that there was informational content in some of the colours (i.e. the cues) or that the blocks varied in their respective statistical structure (i.e. their level of uncertainty).

The second day included the EEG session as well as a subsequent post-measurement. The EEG session consisted of eight blocks (four blocks of each uncertainty level) with a total duration of approximately 48 minutes. Participants were sitting comfortably on a chair in a darkened, sound-dampened and electrically shielded EEG booth. They were instructed to avoid blinking the best they could, most importantly during button presses. Experimental procedure and task during the EEG session were otherwise identical to the training session.

Following the EEG session, participants completed a behavioural post-measurement in order to assess their implicit knowledge of the cue information. To this end, they were shown one final experimental block (duration approx. 5 min) on a computer outside the EEG booth, performing the identical task as before. Crucially, only half of the ordered sequences were cued by the colours learned during the training and the EEG session. The other half began with fixed but different colours that had indeed been presented during training and EEG, but not as cues for the respective participant. Therefore, these colours were non-informative in that they contained no implicitly learned information concerning upcoming trials. In a verbal interview following the post-measurement, all participants denied having noticed any colour-related regularity.

Behavioural data analysis

Statistical analyses of behavioural responses were performed in R (R Foundation for Statistical Computing, Vienna, Austria). First, correct and incorrect responses were aggregated separately for training, EEG session, and post-measurement for each participant. Incorrect responses were further

divided into misses (no response over the course of a sequence) and false alarms (response occurring without presentation of sequential trials). Participants' overall performances were assessed via the discrimination index P_R (Snodgrass & Corwin, 1988).

Reaction times for button presses and releases were assessed for the EEG session and post-measurement. Onset latency was calculated as reaction time relative to the onset of the second trial of any particular ordered sequence (i.e. the earliest possible point to detect a sequential pattern). Offset latency was calculated as reaction time relative to the onset of the first random trial after a particular sequence. Reaction times occurring either before the cue trial (i.e. earlier than -500 ms) or more than 2000 ms after the end of the sequence were excluded. Repeated-measures analyses of variance (ANOVA) and paired *t*-tests were used to assess possible differences in offset (as a function of expectation compliance and uncertainty) and onset latency (learned vs new cue colours during post-measurement). Where appropriate, results of paired *t*-tests were corrected for multiple comparisons at $p = .05$ using the false discovery rate (fdr) correction by Benjamini & Hochberg (1995).

EEG data analysis

EEG data acquisition and data preprocessing

Scalp EEG was recorded from 62 Ag/AgCl-electrodes mounted in a BrainCap TMS electrode cap (Brain Products, Gilching, Germany) using the BrainVision Recorder software (Brain Products, Gilching, Germany). All scalp channels were measured against a ground electrode at position FPz and referenced against FCz during recording. Two additional electrooculogram (EOG) electrodes were applied above and below the right eye for the detection of horizontal and vertical eye movements. All impedances were kept below 10 kOhm. EEG was recorded at a sampling rate of 1 kHz with recording filters set to 0.1 – 1000 Hz bandpass.

EEG preprocessing was conducted in EEGLAB (Delorme & Makeig, 2004). Data segments containing experimental breaks were discarded prior to independent component analysis (ICA). Resulting components distinctly reflecting eye movements were subsequently rejected manually (mean = 2.83 components) using the SASICA toolbox (Chaumon et al., 2015). Data were then filtered with a 0.1 Hz low cut and 30 Hz high cut filter and recalculated to common average reference. Based on participants' overall pattern of reaction times at the end of sequences, a time frame of [-100, 600] ms was defined for the analysis of event-related potentials (ERP) and multivariate segmentation. Epochs containing artefacts were discarded by semiautomatic inspection with an allowed maximum/minimum amplitude of ± 200 μV and voltage steps no higher than 50 μV per sampling. Channels with a high proportion of outliers (kurtosis criterion: $z > 6$) were replaced by a linear interpolation of their neighbour electrodes (M = 1.8 interpolated channels).

Event-related potentials

Averages of the epochs representing our events of interest were calculated separately for each participant. For terminations, the event onset was time-locked to the first unexpected *random* digit, whereas extensions were defined with the onset time-locked to the first unexpected *sequential* digit. *Checkpoints* were defined as points in time at which we hypothesised an incoming regular stimulus to be checked for either a termination (i.e. a check occurring during the ongoing sequence) or an extension (i.e. a check at the regular end) of the ordered sequence. Due to their unambiguous characteristic and temporal distinctiveness, digits at the fourth position of every long extended sequence were defined as *sequential standard trials*. Likewise, digits at the fourth position of every random sequence were sampled as *random standard trials*. Finally, grand averages across participants were calculated for all events of interest.

Using the Mass Univariate ERP Toolbox (Groppe et al., 2011), we employed a two-stage approach to assess reliable ERP differences between conditions: First, we restricted our analyses to specific time frames and electrodes for stringent testing of our hypotheses. In a second step, we performed a whole-brain analysis including all time points to increase sensitivity. In each case, ERPs from the respective conditions were submitted to a repeated measures cluster mass permutation test (Bullmore et al., 1999) using a family-wise significance level of $\alpha = .05$. Repeated measures *t*-tests were performed for each comparison using the original data and 5000 random within-participant permutations of the data. For each permutation, all *t*-scores corresponding to uncorrected p-values of $p = .05$ or less were formed into clusters. The sum of the *t*-scores in each cluster defined the 'mass' of that cluster and the most extreme cluster mass in each of the 5001 sets of tests was used to estimate the distribution of the null hypothesis.

Multivariate segmentation

We used the Cartool software package (available via www.sites.google.com/site/cartoolcommunity) for a segmentation of event-related EEG data sets into topographic maps. This procedure was first introduced by Lehmann and colleagues (1987) to describe what they termed *functional microstates*, meaning brief time periods of stability in otherwise fluctuating field configurations. More generally, this segmentation allows the assessment of spatial field characteristics and their temporal dynamics (see Brunet et al., 2011). As these topographic ERP analyses consider information from all electrodes (i.e. the field configuration as a whole), they offer a multivariate approach to investigating effects between time periods or experimental conditions without a priori selection.

The methodology behind topographic ERP analyses has been described in great detail (see Murray et al., 2008 for an excellent step-by-step tutorial) and is briefly outlined here. Based on the so-called AAHC (*atomize and agglomerate hierarchical clustering*) algorithm, Cartool first iteratively

generated clusters of ERP topographies to identify template maps separately for each experimental condition. A cross-validation criterion was then used to determine which number of template maps optimally described the group-averaged ERPs (i.e. how many template maps were needed to maximise explained variance in the data). Finally, the optimal number of cluster maps per experimental condition was fitted back to the original data, allowing us to compare onset and duration of the same template maps across conditions.

Single-trial analyses: event-specific surprise

In addition to global context effects of uncertainty, single-trial behavioural and physiological correlates of CPs and PEs conceivably depend on how much information is carried by respective stimuli. For single-trial analyses of reaction times and ERPs, we modelled event-specific surprise following the notion of an ideal Bayesian observer (see Harrison et al., 2006). Surprise $I(x_i)$ was defined as the improbability of event x_i, i.e.

$$I(x_i) = -\ln p(x_i)$$

with

$$p(x_i) = \frac{n_j^i + 1}{\sum_k n_k^i + 1}$$

where n_j^i denotes the total number of occurrences of outcome j (*terminated, regular, extended*) up to the current observation i relative to the sum of all past observations (with k for all possible outcomes).

3.4 Results

Behavioural results

EEG session

All participants showed an overall high level of performance with a mean P_R score of $\mathcal{M}_{PR}$ = 0.90 (SD = 0.06) during the EEG session, indicating good attentiveness throughout the experiment. Mean P_R scores did not differ significantly between experimental blocks ($F(7, 248) = 0.03$, $p = .999$) or as a function of block uncertainty ($t(29) = 1.58$, $p = .139$, see Fig. 3-2A).

The repeated-measures ANOVA yielded a significant main effect of expectation compliance on offset latency ($F(2, 58) = 51.54$, $p < .001$). Post-hoc pairwise t-tests revealed participants' button releases to be significantly slower after terminated (M = 619.48 ms, SD = 96.55 ms) than after regular (M = 532.81 ms, SD = 99.74 ms, fdr-adjusted $p = .003$) as well as after extended sequences (M = 346.68 ms, SD = 219.35 ms, fdr-adjusted $p < .001$). The difference between extended and regular sequences was significant as well (fdr-adjusted $p < .001$, see Fig. 3-2B). This pattern of offset latencies fully replicated the findings from our previous fMRI study (see Kluger & Schubotz, 2017). Neither the main effect of uncertainty ($F(1, 29) = 0.05$, $p = .821$) nor the interaction term of uncertainty X expectation compliance ($F(2, 58) = 1.72$, $p = .187$) reached statistical significance, suggesting that participants were able to discriminate regular from manipulated sequences regardless of the respective uncertainty level. The number of misses ($t(29) = -1.89$, $p = .068$) and false alarms ($t(29) = 0.10$, $p = .923$) did not differ significantly between high and low uncertainty blocks (see Fig. 3-2A).

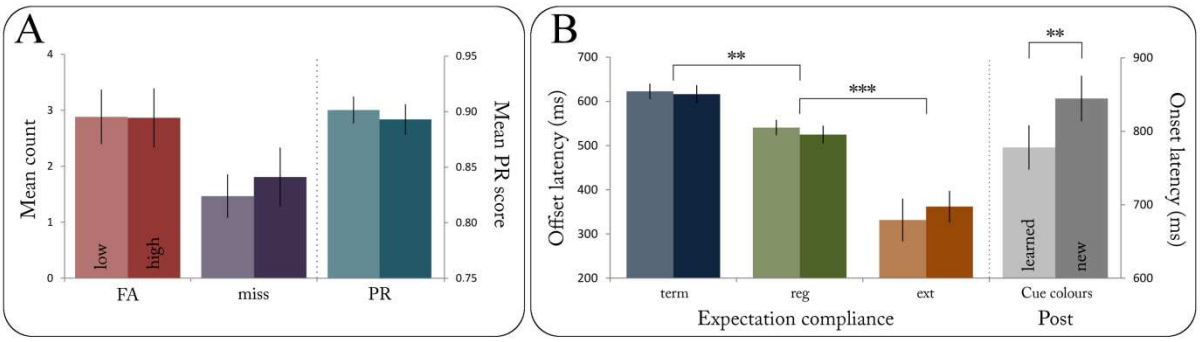

Figure 3-2. (A) Mean count of false alarms (FA) and misses per block as well as mean PR score as a function of uncertainty. (B) Mean offset latencies for terminated, regular, and extended sequences as well as mean onset latencies for learned and new cue colours during post-measurement. ** = $p < .01$, *** = $p < .001$

Post-measurement

Participants performed equally well during the post-measurement (M_{PR} = .90, SD = 0.05) as they had during the EEG session. The post-measurement was conducted in order to assess cue learning: If participants had learned the association of cue colours and prospective ordered sequences over the course of the training and the EEG session, they could be expected to respond more quickly to sequences beginning with established cue colours than to those starting with new colours during the post-measurement. Indeed, the corresponding *t*-test confirmed a significant difference between learned and new cue colours ($t(29) = -2.47$, $p = .01$, one-tailed): Participants exhibited a shorter reaction time to learned cue colours (M = 788.02 ms, SD = 168.00 ms) than to new cue colours just introduced during the post-measurement (M = 844.59 ms, SD = 172.31 ms; see Fig. 3-2B).

Performance-based subgroup analyses

Based on participants' reaction times at onset during the post-measurement, the sample was median-split into one group that had shown a gain in response speed following the learned cue colours (*gain*) and another group that had not (*no gain*, see Fig. 3-3A). The gain group showed a significantly higher difference between reactions to new and learned cue colours (M = 178.54 ms, SD = 106.62 ms) than did the no gain group (M = -45.40 ms, SD = 82.24 ms, $t(26.64) = 6.29$, $p < .001$, one-sided).

The rationale behind comparing the two subgroups' behavioural performance was that a stronger association of cue colour and sequence length (as reflected by a pronounced gain in response speed during post-measurement) should entail distinct response patterns at the end of regular and manipulated sequences. We repeated the offset latency ANOVA separately for gain and no gain groups and found the overall main effect of expectation compliance to be present in both groups (gain: $F(2, 28) = 30.15$, $p < .001$; no gain: $F(2, 28) = 28.23$, $p < .001$; see Fig. 3-3B). Notably, only the gain group showed a significant interaction of expectation compliance and irreducible uncertainty ($F(2, 28) = 7.98$, $p = .002$): Button releases at the end of extended sequences occurred significantly earlier when uncertainty was low (M = 221.16 ms, SD = 328.87 ms) than when it was high (M = 333.75 ms, SD = 252.23 ms, $t(14) = 2.90$, $p = .012$).

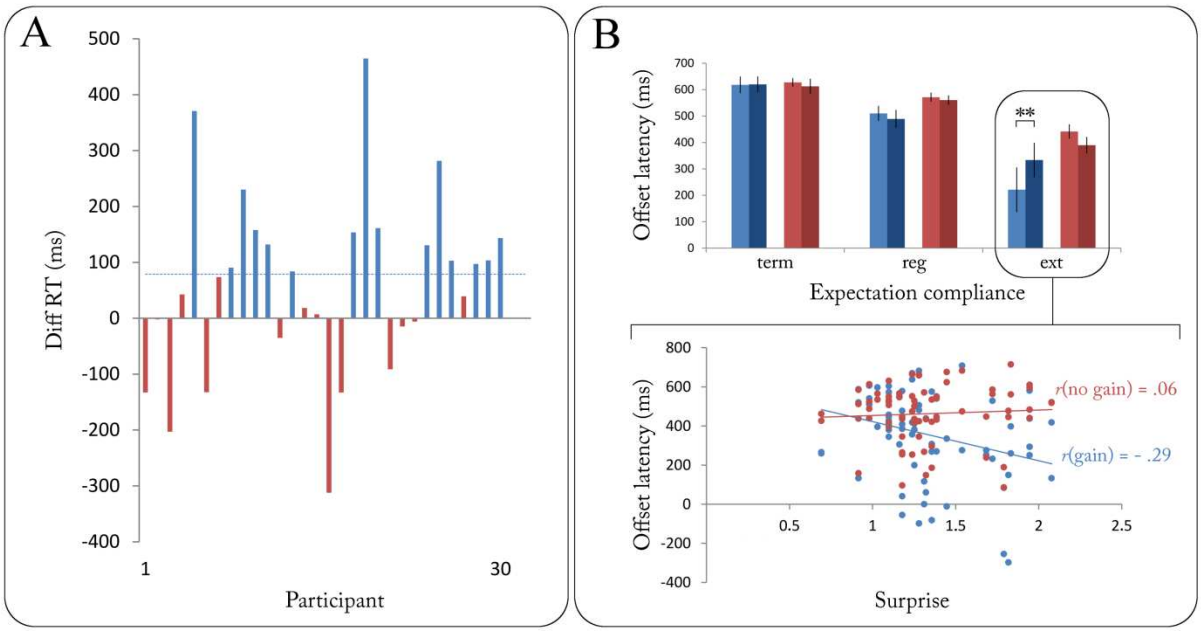

Figure 3-3. (A) Individual gains in reaction time (defined as the difference in reaction time following new minus learned cues) during post-measurement. Positive values indicate quicker button presses following learned cues. Blue dotted line depicts Mdn_{Diff} = 78.70 ms. Participants were consequently median-split into a *gain* group (blue) and a *no gain* group (red). (B) Upper panel: Mean offset latencies as a function of expectation compliance for gain (blue) and no gain group (red). Significant differences only shown for high vs low uncertainty for the sake of clarity (see Fig. 3-2B for differences between levels of expectation compliance). Lower panel: Correlations between offset latency and trial-specific surprise value of sequential extensions for both groups. ** = $p < .01$

By definition, extensions were on average more surprising under low uncertainty due to their low presentation rate in these blocks. Importantly, however, event-specific surprise values of extensions also fluctuated under high uncertainty (albeit to a lesser extent). Thus, the reported uncertainty effect on the gain group's responses after extended sequences might be generalised across uncertainty levels in such a way that more surprising extensions – regardless of global contextual features – evoked shorter offset latencies in the gain group: If excess reliance on cue information had in fact determined the behavioural effect found for the gain group, these participants should have responded equally fast to locally surprising extensions irrespective of global uncertainty. Corroborating this hypothesis, we found a significant negative correlation of stimulus-bound surprise and offset latency after extended sequences for the gain group ($r(72) = -.29$, $p = .013$) but not for the no gain group ($r(72) = .06$, $p = .617$). The difference between the two correlation coefficients was found to be significant ($Z = 2.11$, $p = .017$, one-tailed).

An intuitive explanation for earlier releases following highly surprising extensions would be that the gain group not only released the button more quickly, but also more often prematurely: Conceivably, the more stable the cue information had been learned, the more likely would the response button be released at the expected sequence end rather than at the actual end. We assessed two additional questions with regard to more specific distinctions in behaviour: First, we hypothesised the gain group to more frequently respond prematurely to extended sequences, i.e. at the 'would-be' end of the sequence had it not been extended. Recall that unexpected extensions occurred at the sequential positions where – based on the cue information – a non-sequential digit was expected. Accordingly, as illustrated in Figure 3-4, we compared the two groups' button releases within the interval of -1000 (onset of the unexpected sequential digit) and 500 ms (offset of the first non-sequential digit) around the end of extended sequences. Supporting our hypothesis, the gain group was found to have a

significantly higher number of releases within the [-1000, 500] ms time frame than the no gain group ($t(15) = 22.28$, $p < .001$, one-sided; see Fig. 3-4A). The group difference in incremental releases per 100 ms window was also found to be significant ($t(15) = 2.35$, $p = .017$, one-sided).

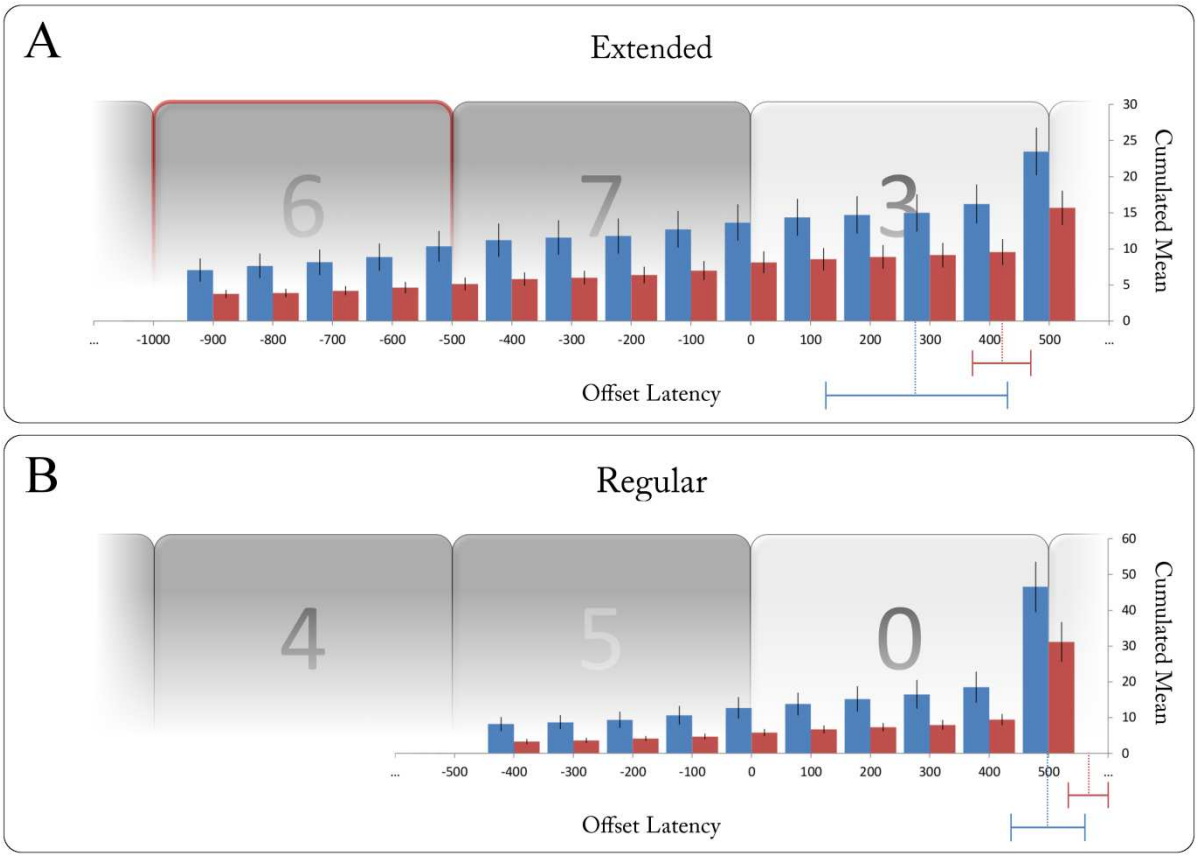

Figure 3-4. (A) Mean count of button releases during the experiment up to selected offset latencies for gain (blue) and no gain group (red). Shown here for an exemplary short extended sequence (length of 7 digits), the gain group was found to release the response button more frequently at offset latencies between -1000 and +500 ms (i.e. between the onset of the unexpected sequential digit [red frame] and the offset of the first non-sequential) following extended sequences. Dotted lines and bars depict mean offset latencies for regular sequences per group ± 2 SEM. (B) Similarly, shown here for a short regular sequence (length of 5 digits), the gain group was found to release the response button more frequently at offset latencies between -500 and +500 ms (i.e. between the onset of the last sequential digit and the offset of the first non-sequential digit) following regular sequences. Dotted lines and bars depict mean offset latencies for extended sequences per group ± 2 SEM.

Second, stronger expectations by means of more accessible cue information within the gain group could conceivably lead to a similar pattern of early responses following regular sequences. The gain group could therefore be expected to more frequently release the response button within a brief interval of ± 500 ms around the end of regular sequences. Responses during the last sequential digit

(i.e. offset latency between -500 and 0 ms) would reflect an anticipatory release of the response button whereas responses during the first non-sequential digit (0 – 500 ms) would reflect a quick detection of the sequence end. Both anticipatory and quick releases after the end of a regular sequence were hypothesised to be positively associated with the degree to which the colour-length association had been learned. Fittingly, the gain group was found to have a significantly higher number of releases within the ± 500 ms time frame than the no gain group ($t(10) = 9.47$, $p < .001$, one-sided; see Fig. 3-4B). The group difference in incremental releases per 100 ms window was found to be marginally significant ($t(10) = 1.81$, $p = .05$, one-sided).

EEG results

Event-related potentials

PE – STD

Based on our hypotheses, we first tested prediction errors and sequential standards for reliable differences in the P300 time frame. We analysed all time points between 300 and 600 ms (1350 comparisons in total) from two subsets of electrodes: one parieto-central subset (CP1, CPz, CP2, P1, Pz, P2) to detect a posterior P3b component and a fronto-central subset (F1, Fz, F2) controlling for anterior P3a effects. Supporting our hypothesis, we found a significant P3b over the parieto-central electrodes peaking around 388 ms (Fig. 3-5A).

Subsequently, all time points between 0 and 600 ms were included in a two-sided whole-brain analysis to assess reliable differences exceeding our hypotheses (18600 comparisons in total). In addition to the reported P3b effect, we found a significant P600 component with a right-lateralised parietal scalp distribution peaking around 500 ms (Fig. 3-5A).

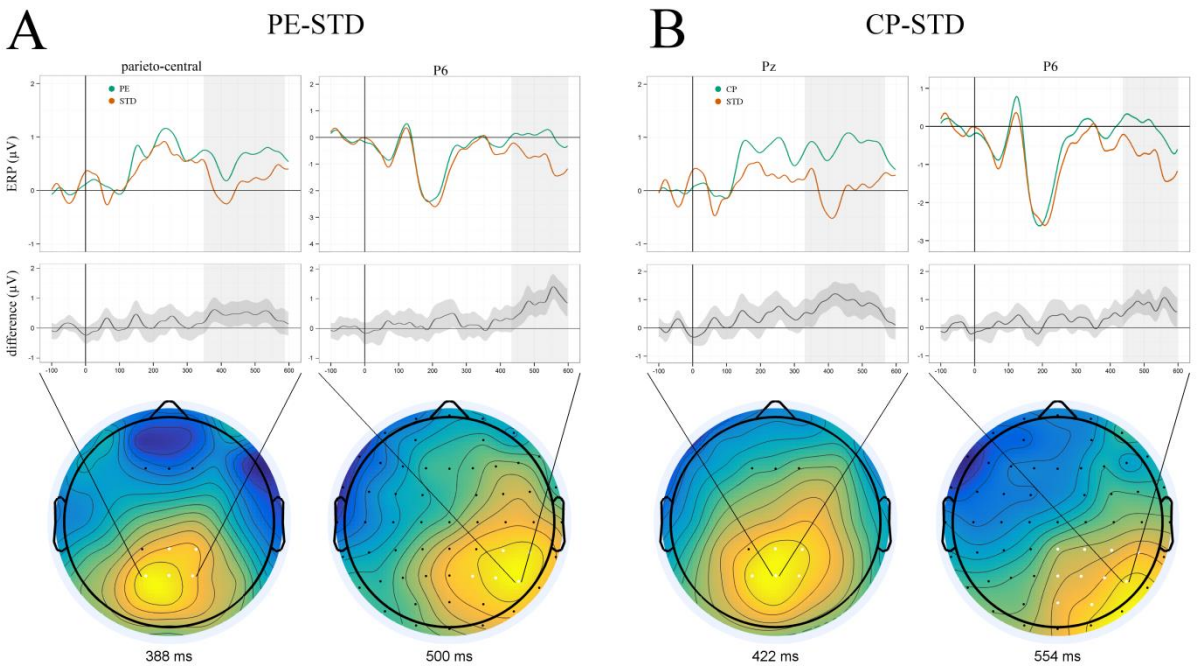

Figure 3-5. (A) Significant ERP differences between prediction errors and sequential standards included a parieto-central P3b (left) as well as a right-lateralised P600 component (right). P3b topography shows the frontal and parietal subsets of electrodes used for the analysis (bottom left). Significant clusters are marked in white. (B) ERP differences between checkpoints and sequential standards were equally reflected in significant P3b (left) and P600 components (right).

CP – STD

Like PEs, checkpoints are probabilistic, highly informative sequential positions with an immediate relevance for behaviour. Therefore, one would expect a certain degree of similarity between PE and checkpoint ERPs when compared with deterministic, behaviourally non-informative standard trials. ERPs from checkpoints and sequential standards were submitted to a one-sided analysis including all time points between 300 and 600 ms (1350 comparisons in total) and the two electrode clusters described above. The analysis revealed a pattern very similar to previous prediction error ERPs, including both a significant posterior P3b (peaking around 422 ms) and a right-lateralised P600 (peaking around 554 ms, Fig. 3-5B). Notably, P3b and P600 peak latencies thus occurred slightly earlier for prediction errors than for checkpoints.

Since strategic adaptation of CP processing as a function of context uncertainty was one of the central objectives of the previous fMRI study, we subsequently split the analysis to separately assess high and

low uncertainty checkpoint ERPs. P3b and P600 were found for checkpoints in both uncertainty conditions (Fig. 3-6). Interestingly, while the P3b component was virtually identical in both latency and scalp distribution, we found subtle differences regarding the P600: At the group level, the activation peak occurred ~50 ms earlier and slightly more frontally for high (500 ms at CP4) than for low uncertainty checkpoints (554 ms at P6, see Fig. 3-6).

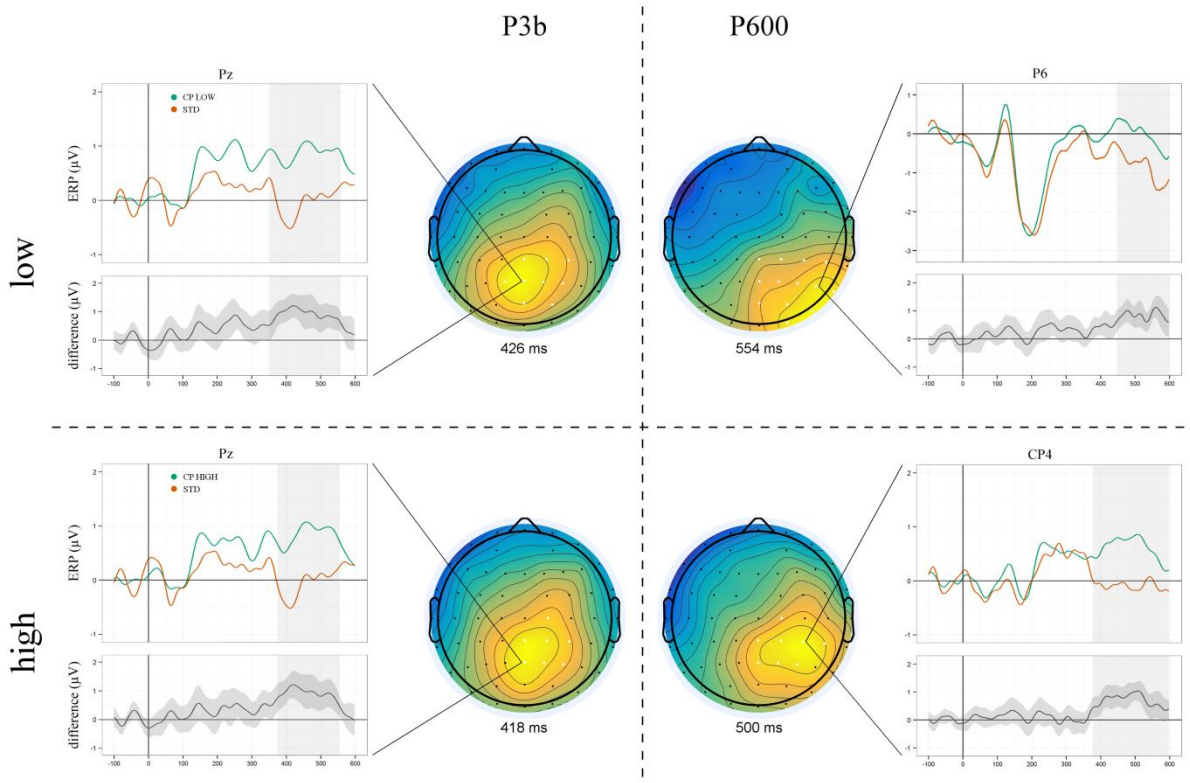

Figure 3-6. Grand averaged ERPs of low (top row) vs high uncertainty checkpoints (bottom row) and sequential standards. Checkpoints elicited significant P3b (left) and P600 components (right) irrespective of the uncertainty level. Note that, while uncertainty did not modulate P3b scalp distribution or peak latency, the P600 elicited by high uncertainty checkpoints showed an earlier peak and a slightly more frontally distributed topography.

Recall that we observed an earlier P3b peak for PEs (388 ms) than for low (426 ms) and high uncertainty CPs (418 ms). In contrast, P600 peak latencies were identical for PEs and high uncertainty CPs (500 ms) and earlier than for low uncertainty CPs (554 ms). This pattern of ERP results suggests a close functional relationship of PEs and (particularly high uncertainty) CPs (see Fig. 3-5A and 3-6).

This relationship and its variation under uncertainty were the main objective of our subsequent single-trial and multivariate segmentation analyses (see below).

PE – CP

Given the reported conceptual and functional similarities between PE and CP, their direct contrast was meant to reveal the correlate of expectation violation definitive of PE. We hypothesised this mismatch to be reflected in an enhanced N400 component. Accordingly, we included all time points between 300 and 500 ms in a one-sided whole-brain analysis (6262 comparisons in total). PEs were found to elicit a significantly enhanced N400 over parieto-central electrodes peaking around 418 ms (Fig. 3-7). No additional significant components were found in the subsequent whole-brain analysis including all time points between 0 and 600 ms (18600 comparisons in total).

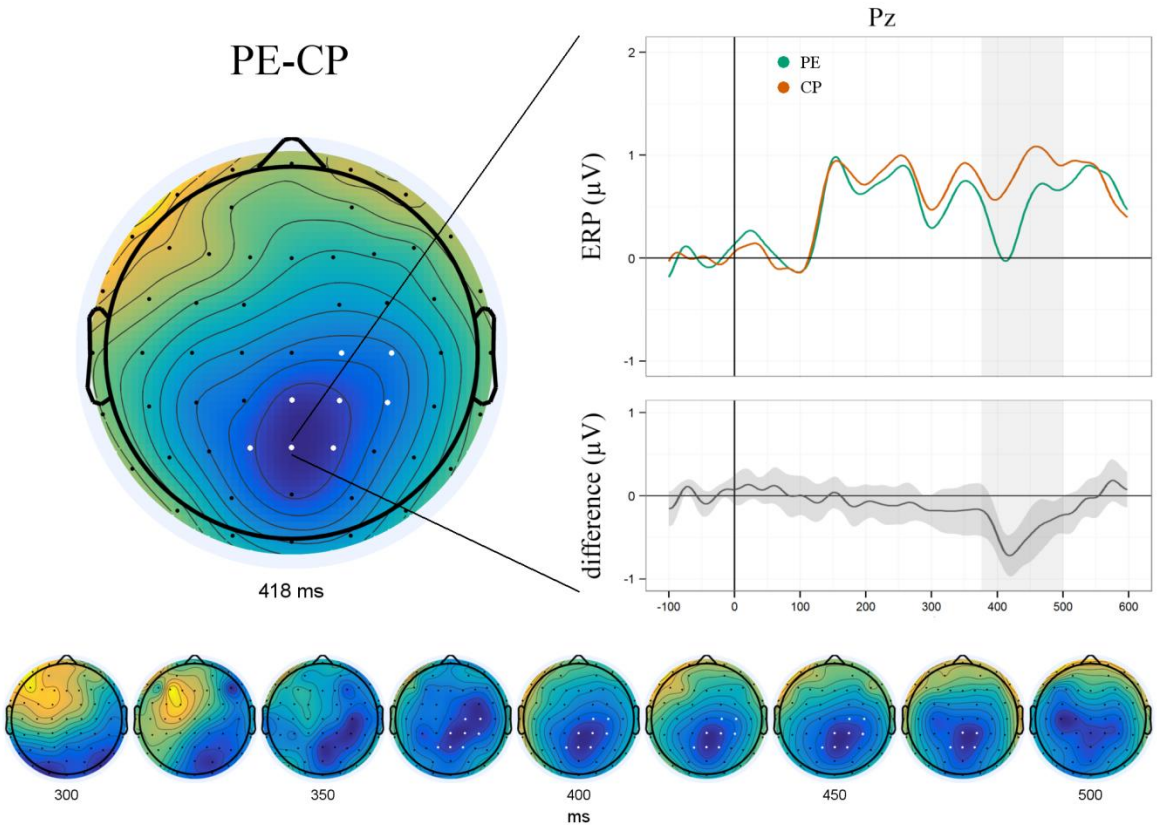

Figure 3-7. The direct comparison of prediction errors and checkpoints revealed a significant N400 component peaking around 418 ms over parieto-central electrodes. Bottom panel shows component evolution over time.

Multivariate segmentation

Cartool's meta-criterion showed that group-averaged ERPs of CPs, PEs, and STD were optimally described by a set of 12 topographic template maps (TM). Figure 3-8 shows the temporal progressions of these topographies for each condition. Visual inspection suggested notable differences between conditions within two main time frames. First, following a virtually simultaneous onset of fronto-centrally distributed TM 11 (around 204 ms), PEs and high uncertainty CPs exhibited a sustained frontal cluster (TM 2, 284 – 326 ms) after transitioning through a more global TM 12 (Fig. 3-8A). Whereas this frontal shift was not found for low uncertainty CPs, it was even more pronounced for STD (i.e., with a higher amplitude and an earlier onset). After fitting the group-level template maps onto individual subject data, one-sided *t*-tests confirmed a significantly greater global field power of TM2 for STD compared to PEs ($t(13.08) = 1.91$, $p = .039$) and CP HIGH ($t(14.83) = 1.83$, $p = .044$). Similarly, onsets of TM2 occurred significantly earlier for STD than for CP HIGH ($t(20.86) = -1.95$, $p = .033$). The comparison of STD and PEs was marginally significant ($t(15.42) = -1.67$, $p = .057$).

Second, ERP time courses showed differential topographic as well as temporal configurations during a later time frame (starting at around 360 ms). PEs and both CP conditions shared a frontal-to-parietal shift (TM 3 – 5) with particular differences in cluster onset and duration (Fig. 3-8B). In contrast, sequential standard trials showed a distinct ongoing frontal topography with a slight dominance of left hemisphere sources (TM 9, 406 – 540 ms). Group-level onset and duration for the reported topographies are listed in Table 3-1.

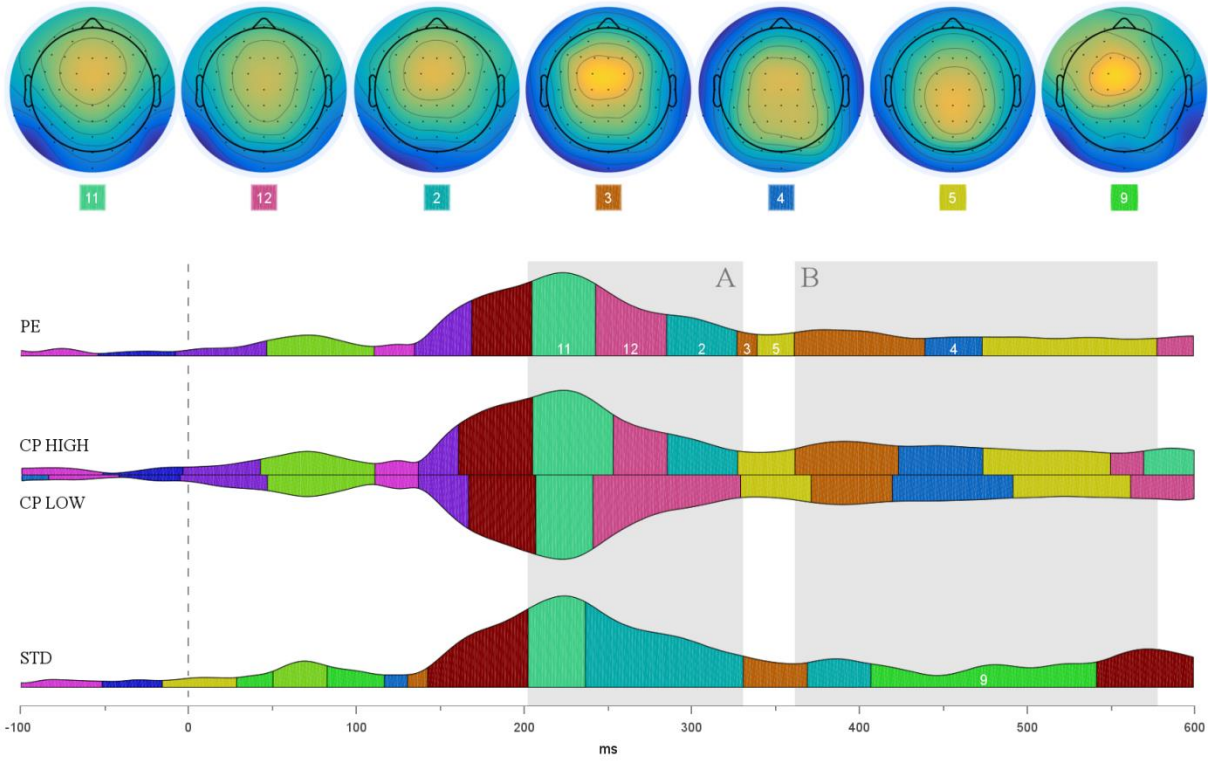

Figure 3-8. Global field power (GFP) of group-averaged ERPs for prediction errors, checkpoints under high/low uncertainty, and sequential standard trials time-locked to stimulus onset. Coloured segments within the area under the curve depict distinct topographic configurations (*template maps*, TM) as revealed by hierarchical clustering. Note that the CP LOW curve was flipped for illustrative purposes only and did not differ in polarity.

Table 3-1. Group-level onset and duration of selected template maps for PE, high/low uncertainty checkpoints, and sequential standard trials. Time frame for grand average ERP analysis [-100, 600] ms.

TM class	Condition			
	PE	*CP HIGH*	*CP LOW*	*STD*
TM 2				
Onset (ms)	284	284	-	236 | 368
Duration (ms)	42	42	-	94 | 38
TM 3				
Onset (ms)	326 | 360	360	370	330
Duration (ms)	12 | 78	62	48	38
TM 4				
Onset (ms)	438	422	418	-
Duration (ms)	34	50	72	-
TM 5				
Onset (ms)	472	472	328 | 490	-
Duration (ms)	104	76	42 | 70	-
TM 9				
Onset (ms)	-	-	-	406
Duration (ms)	-	--	-	134

TM 11				
Onset (ms)	204	204	206	202
Duration (ms)	38	48	30	34
TM 12				
Onset (ms)	242	252	240	-
Duration (ms)	42	32	88	-

3.5 Discussion

Predicting events of everyday life, our internal model of the world is constantly compared to sensory input we perceive. Prediction errors induced by unexpected events are deemed particularly informative in that they instigate learning through model updating. We show here that information is equally sampled from *expected* events at particularly relevant checkpoints, suggesting that under uncertainty, model-affirmative events similarly prompt recourse to the internal model. Both checkpoints and prediction errors showed a significant P3b component when compared to sequential standards, indexing the relative (im)probability of CP and PE occurrence. Conversely, the direct comparison of CPs and PEs revealed a significant N400 component as the mismatch correlate elicited solely by prediction errors. Combined with findings from behavioural and functional microstate analyses, CP characteristics highlight the significance of informative reference points for abstract predictive processing, raising intriguing questions for future research.

Functional Characteristics of Checkpoints

In order to establish a more precise characterisation of checkpoints, they have to be related to and dissociated from two other event types: First, since CPs are regular events, they share the expectedness of sensory input with sequential standards. In contrast to these standards, however, checkpoints are probabilistic and therefore informative with regard to task context and behavioural requirements.

Second, PEs are equally informative but do carry a mismatch signal that requires behavioural adaptation in opposition to the internal model.

As hypothesised, the significance of CPs and PEs as particularly meaningful points in time was reflected in a joint P3b component compared to least informative standards. Often discussed as an index of enhanced information transmission and allocation of resources (Humphrey & Kramer, 1994; Kok, 2001), P3b is well suited to reflect exploitation of information at these sequential positions. More precisely, an incoming stimulus is evaluated in context of previous stimuli by comparing it to information from working memory (Polich, 2003; Polich & Criado, 2006). Such monitoring is immediately beneficial for stimulus classification and – where required – transforming this information into action (Nieuwenhuis et al., 2005; Verleger et al., 2005). These proposals fit well with central findings from the original fMRI study in which we found enhanced activation at CPs under high (vs low) contextual uncertainty. We interpreted these effects as an iterant evaluation of model information retrieved from working memory, pointing towards a strategic adaptation of predictive processing to contextual statistics (Kluger & Schubotz, 2017). Notably, the observed activation pattern included the temporo-parietal junction (TPJ), a hypothesised cortical source of the P3b (Donchin & Coles, 1988). Common ERP components and the similarities in functional microstates thus further illuminate the processing of CPs and PEs as highly informative events, suggesting that positions of potential and actual prediction errors are being exploited in a similar way.

It remains the key difference between CPs and PEs that only the latter violated cue-based predictions. Therefore, despite the similarities reported above, CPs and PEs will eventually be processed differently once consequences of the actual stimulus come into effect. Supporting our initial hypothesis, the mismatch signal for PEs (vs CPs) was reflected in an N400 component. N400 effects have typically been reported when words mismatched semantic expectations shaped by previous context information (e.g., Kutas & Federmeier, 2000). Closely related to the present paradigm,

centro-parietal N400 effects following incorrect (vs correct) solutions in arithmetic tasks (e.g., Galfano et al., 2004) point towards a more general process independent of stimulus modality. Accordingly, Kutas & Federmeier (2011) discuss the N400 as an index of conceptual representations which – when contextually induced predictions are violated – may need to be refined. Such adaptive processes are conceivably reflected by components occurring even later than the PE-related N400, e.g., ERPs related to subsequent digits 'confirming' the initially surprising stimulus. Future research could make use of later time frames to further distinguish prediction errors and checkpoints with regard to the respective consequences they entail.

To summarise, CPs are informative points in time which, despite a lack of unexpected input, show close functional similarities to canonical PEs. Our findings suggest that information from particular sequential positions, irrespective of the actual outcome, is used for evaluation and/or updating of internal models. Importantly, while sensory input at CPs complied with the more likely expectation, their sequential positions were tagged by the statistical structure inherent in the stimulus stream. Previous fMRI results have shown CPs to be exploited particularly in highly uncertain contexts, conceivably in order to solve ambiguity with regard to upcoming sensory information and efficiently adapt behaviour. Overall, the functional profile of checkpoints conceptually relates to *bottleneck states* (Botvinick, 2012; Solway et al., 2014) from the realm of hierarchical reinforcement learning: These states are passages or transition points between larger sets of states, effectively forming natural subgoals in hierarchical representations of behaviour (Moradi et al., 2012; Simşek & Barto, 2009). Similarly, the sequential positions of CPs and PEs mark informative transition points between predictable and non-predictable (random) states. Depending on whether or not the presented stimulus complied with cue-based expectations, checkpoints and prediction errors are supposedly used for model evaluation and updating, respectively.

Implications for Predictive Processing

Combined ERP and microstate findings of the present study revealed considerable similarities between the representations of CPs and PEs. On a broader scale, this suggests overlapping roles of CPs and PEs in predictive processing. Given that PE-based model updating has been established to be fundamental for associative learning (den Ouden et al., 2009), CPs could similarly be used for model *evaluation*. Clearly, expectation-compliant information (as observed at CPs) does not call for corrective model updating. It seems unlikely, however, that potentially critical information extracted from CPs would not be used to evaluate the validity of model statistics on-line. Particularly for the estimation of higher level statistics, the number of regular outcomes at critical time points is no less instructive than the number of PEs. Support for this proposition comes from earlier studies using digit sequences in abstract predictive processing. Kühn and Schubotz (2012) found a distinct frontal correlate of regular, model-compliant events at sequential positions where statistically rare breaches of expectancy had previously been observed. As the actual sensory input neither violated model-based predictions nor called for behavioural adaptation, these frontal responses reflected increased weight of bottom-up signals driving potential model updating solely based on statistical regularities. Another study (Trempler et al., 2017) manipulated the requirement to either ignore or respond to two different expectation violations. Again, violations that could be ignored ('drifts') did confirm the internal model, whereas violations that required a response ('switches') prompted corrective model updating. The pattern of brain activation suggested a two-step neural response to these events, starting with joint processing of stimulus discrimination followed by distinct correlates of behavioural responses prompted by the respective violation type.

In line with these previous findings, we suggest information from CP and PE time points to be evaluated irrespective of the actual outcome (distinguishing both events from STD), especially under

uncertainty. Successive model adaptation is induced only in case of unexpected stimuli (distinguishing PEs from CPs). As the temporal resolution of fMRI did not allow for the inclusion of standard trials in the original study, it remains an intriguing question for future research to determine how context (in)stability influences the expectation and processing of these informative events.

In addition to effects of context uncertainty, behavioural subgroup analyses suggested inter-individual differences in cue learning as a determining factor for CP/PE processing: The more strongly participants had learned the cue-length association, the more often they showed early responses at the end of a sequence. Depending on which sequence was observed, this response pattern had diverging implications on behavioural efficiency: In case of regular sequences, early releases during the last sequential digit showed how strong anticipation of the sequence ending spurred fast and efficient responses. Critically, however, the very same anticipation led some participants to erroneously respond at the 'would-be' end of extended sequences. One explanation could be that (overly) successful cue learning triggered a consistent prediction of sequence length ("Five digits after green") irrespective of context-dependent violations. This way, information from CPs (in regular sequences) or PEs (in extended sequences) would not be exploited, as indicated by the negative correlation between event-specific surprise and offset latency. Overall, these results suggest that participants with increased knowledge of cueing information strongly (and sometimes falsely) relied on these initial cues, virtually disregarding potentially informative transition points during the sequence. In other words, excess reliance on cue information led to less attention being given to these transition points. More formalised accounts of predictive processing have postulated attention to control the involvement of prior expectations at different levels (Friston, 2009). Specifically, attention is conceptualised as a means to increase the weight (or *gain*) of neural responses coding error signals, making them more eligible to drive learning and potential behavioural adjustments. Strict

adherence to cue information conceivably impedes allocating attentional resources to CP/PE time points and, consequently, model adaptation. One promising direction for future studies would thus be to assess the interplay of bottom-up and top-down dynamics underlying information processing at these points in time.

Limitations and Future Directions

The main aim of the present study was to exploit the temporal benefits of EEG for an extension of previous fMRI results. In order to warrant a high degree of comparability between the two studies, we chose a full replication of the experimental paradigm. As a consequence, it remains a limitation of the present study that half of the CPs required a response whereas the other half did not (for discussion, see Kluger & Schubotz, 2017). Therefore, one central direction for current studies is to reduce the number of PE types, effectively ensuring behavioural relevance of all CPs. Furthermore, there are several promising analyses beyond the scope of this paper which would not have been ideal for the current ERP epochs (-100 ms to 600 ms). Going forward, specifically re-epoching the data to include a longer pre-stimulus period would allow ERP and time-frequency analyses of anticipatory CP/PE processing as a function of uncertainty. Relatedly, the microstate analyses presented here motivate a more in-depth multivariate assessment of STD, CP, and PE representations, extending our understanding of similarities and differences between them. For example, STD trials should be reliably discriminable from CPs and PEs already during the pre-stimulus period, reflecting the anticipation of task-relevant information that can be obtained from the latter. Thus, representations of CPs and PEs should be similar during the pre-stimulus period but distinct during later periods reflecting actual outcome processing. Learning about the time course and potential uncertainty modulation of these comparisons will provide a more comprehensive account of the factors driving abstract prediction.

3.6 Conclusion

Checkpoints are probabilistic, cue-compliant events informing predictive processing. Their functional profile closely resembles that of canonical prediction errors, indicating similar roles of CPs and PEs in abstract prediction. Both types of events presumably serve as reference points providing behaviourally relevant information, the central distinction being whether the respective outcome violates the internal model (PE) or not (CP). We suggest that despite the expected input observed at checkpoints, information at these particular positions is exploited on-line in order to adapt behaviour. Intriguing questions remain with regard to underlying network dynamics and their potential modulation as a function of uncertainty.

3.7 Acknowledgements

We would like to thank Monika Mertens, Katharina Thiel, and Alina Eisele for their help during data collection.

4. Study 3

Exploitation of local and global information in predictive processing

Daniel S. Kluger[a,b], Nico Broers[b,c], Marlen A. Roehe[b,c], Moritz F. Wurm[d,e], Niko A. Busch[b,c], and Ricarda I. Schubotz[b,c,f]

[a] Institute for Biomagnetism and Biosignalanalysis, University of Muenster, Muenster, Germany

[b] Otto-Creutzfeldt-Center for Cognitive and Behavioral Neuroscience, University of Muenster, Muenster, Germany

[c] Department of Psychology, University of Muenster, Muenster, Germany

[d] University of Trento, Center for Mind/Brain Sciences, Rovereto, Italy

[e] Harvard University, Department of Psychology, Cambridge, USA

[f] Department of Neurology, University Hospital Cologne, Cologne, Germany

Correspondence should be addressed to

Daniel S. Kluger
Institute for Biomagnetism and Biosignalanalysis
University of Muenster
Malmedyweg 15
D-48149 Muenster
Germany
E-mail: daniel.kluger@uni-muenster.de

Key words: EEG, expectation, local information, global information, model adaptation

4.1 Abstract

While prediction errors have been established to instigate learning through *model adaptation*, recent studies have stressed the role of model-compliant events in predictive processing. Specifically, so-called *checkpoints* have been suggested to be sampled for *model evaluation,* particularly in uncertain contexts.

Using electroencephalography (EEG), the present study aimed to investigate the interplay of such global information and local *adjustment cues* prompting on-line adjustments of expectations. Within a stream of single digits, participants were to detect ordered sequences (i.e., 3-4-5-6-7) that had a regular length of five digits and were occasionally extended to seven digits. Extensions could be unexpected or indicated by incidental colour cues.

Local exploitation of cue information was reflected in significant pattern decoding of cues vs non-cues using pattern classification. Global modulation of checkpoint processing as a function of uncertainty was likewise reflected in significant decoding of high vs low uncertainty checkpoints. In line with previous results, both analyses comprised the P3b time frame as an index of excess information sampled from probabilistic events.

Accounting for cue information, an N400 component was revealed as the correlate of locally unexpected (vs expected) outcomes, reflecting effortful integration of incongruous information. Finally, we employed representational similarity analysis (RSA) to compare the fit of a *global model* (disregarding local adjustments) and a *local model* (additionally considering cue information). Results showed significant correlation of both models to neural data with a better fit of the global model at later latencies (400 – 700 ms).

4.2 Introduction

One of the central faculties of the human brain is to predict upcoming events based on internal models of the world. Incoming sensory information is constantly compared to model-based predictions and resulting mismatches - termed *prediction errors* (PE) - are used to continuously update the model for future reference (Friston, 2005; Mumford, 1992). Consequently, as such unexpected events are particularly informative, the importance of PE for associative learning through model adaptation has long been established (Bastos et al., 2012; Den Ouden et al., 2009). Recent studies have stressed the role of probabilistic events which carry behaviourally relevant information but do not indicate a mismatch (Kühn & Schubotz, 2012; Kluger et al., submitted; Trempler et al., 2017). In contrast to *model adaptation* instigated by PE, these critical points in time have been suggested to serve outcome-independent *model evaluation*. On-line exploitation of model-compliant sensory input at so-called *checkpoints* (CP) has been shown to be modulated by higher-order context statistics such as irreducible uncertainty (Kluger & Schubotz, 2017): In a highly uncertain environment, additional resources are allocated to the processing of CP to resolve ambiguity regarding the validity of the internal model. Notably, we were able to show that this effect was not merely driven by event-related surprise at CPs.

A subsequent EEG study (Kluger et al., submitted) revealed further functional characteristics of CP, demonstrating both electrophysiological commonalities with and distinctions from PE. When compared to fully predictable standard trials, both event types showed a significant P3b component, marking the joint (im)probabilistic - and therefore, informative - nature of CP and PE in accordance with the component's functional significance (Mars et al., 2008; Seer et al., 2016). The direct comparison of the two revealed a significant N400 component to signal a predictive mismatch exclusively induced by PEs. In sum, these findings illustrate two important aspects of predictive

processing: First, CP appear to take on a similar, but not identical role as PE in resolving local ambiguity regarding model validity. Second, the extent to which CP are exploited to evaluate the internal model is modulated by higher-order parameters such as environmental uncertainty. These findings indicate that CPs have the potential to play a central role in predictive processing and are thus worth a more thorough determination, especially in relation to PEs.

Building on the results reported so far, the present EEG study aimed to address the potential influence of learned local contingencies on predictive processing. Normally, CP and PE are informative due to their probabilistic nature, i.e. their occurrence is unknown before it is actually observed. However, these critical events can conceivably be rendered uninformative by local cues reliably indicating a certain outcome - much like a priority road sign telling drivers not to slow down at a particular crossing. We employed a modification of the paradigm from Kluger et al. (submitted) to assess both the electrophysiological signature of these local cues as well as their influence on processing of expected and unexpected outcomes.

Trained participants performed a serial pattern detection task in which they were to press and hold a response button whenever they detected an ordered digit sequence (e.g. 3-4-5-6-7) within an otherwise pseudorandom stream of coloured single digits. Ordered sequences had an expectable length of five digits but were occasionally extended to be seven digits long. This way, the sixth position could either contain a random digit denoting the regular (i.e., expected) end of the sequence (REG) or an unexpected sequential digit indicating an extension (EXT; see Fig. 4-1). Relative presentation rates of REG and EXT were probabilistically modulated by blockwise manipulation of irreducible uncertainty (De Berker et al., 2016; Payzan-LeNestour & Bossaerts, 2011). Crucially, some sequences contained colour-coded *adjustment cues* (AC) at the third position. These cues (AC+) provided excess information regarding the continuation of the sequence, reliably indicating a

sequential extension (p = .75) and thus effectively overriding expectations based on the global task structure.

Using a similar experimental paradigm without adjustment cues, our previous EEG study (Kluger et al., submitted) had revealed a significant N400 component for the direct comparison of PE and CP. In other words, events that violated cue-based expectations (PE) distinctly elicited an N400 when compared to informative events containing model-compliant sensory information (CP). Since the present study comprised adjustment cues (AC) incidentally prompting an on-line shift in model-based expectations, both REG and EXT events could now be locally unexpected as a function of preceding cue information: As AC reliably indicated sequential extensions (EXT+), regular endings following an AC (REG+) were rendered unexpected (see Fig. 4-1B). Conversely, in sequences without AC, regular endings (REG-) were the locally expected outcome and extensions (EXT-) were unexpected. Thus, we aimed to replicate previous findings from the direct contrast of expected and unexpected sequence endings within the N400 time frame (Kluger et al., submitted).

Furthermore, we used multivariate pattern classification of event-related potentials (ERP) to conduct two analyses on local vs global modulations: First, it was one aim of the present study to assess functional characteristics of the newly introduced adjustment cues. Conceivably, when compared to non-informative third position trials (AC-), predictive information provided by adjustment cues (AC+) could manifest in an electrophysiological signature similar to the one reported by Kluger et al. (submitted) for both CP and PE. As our previous study had revealed a joint P3b component to reflect the informativity of both CP and PE, we hypothesised the classifier to reliably decode AC+ and AC- trials within the P3b time frame.

Second, we aimed to validate global effects of CP processing as a function of high vs low uncertainty from our initial fMRI study (Kluger & Schubotz, 2017). Consequently, we sampled regular endings of sequences that did not contain an adjustment cue (REG-), as these events were the precise

equivalent of CP from earlier studies. Consolidating previous findings from fMRI and EEG, we hypothesised significant classifier accuracy of *CP high vs CP low* within the P3b time frame (300-500 ms).

Finally, we applied representational similarity analysis (RSA; Kriegeskorte et al., 2008) to test the explanatory power of competing theoretical models relating the representational structures of critical events (AC, CP, PE) to one another. To that end, we devised two competing binary model dissimilarity matrices (DSMs) to predict the empirical representational similarity of event classes in the EEG data. A *global model* categorically assumed high similarity within and high dissimilarity between the sampled event classes (AC+/- vs REG+/- vs EXT+/-) and thus represented global statistics acquired by means of statistical learning ("Sequences are usually five digits long"). In contrast, a *local model* was set up to reflect these adjustments in expectations due to local information: For sequences containing an adjustment cue (AC+), extensions were no longer unexpected (as in the global model) but rather the expected outcome. In these cases, regular sequence endings (REG+) – otherwise the most probable outcome of a sequence – violated the most recent set of expectations based on the adjustment cue (see above). Consequently, we assumed high similarity of events that violated these potentially adjusted predictions following an adjustment cue (REG+, EXT-) in contrast to those that complied with the adjustment cue (REG-, EXT+).

4.3 Material and methods

Participants

A total of 30 neurologically healthy, right-handed volunteers (25 female) at the age of 24.8 ± 3.46 years (M ± SD) participated in the study for payment or course credit. Participants were recruited from the university's volunteer database and had (corrected-to-) normal vision. Written informed consent was obtained from all participants prior to the start of experimental procedures. Experimental standards complied with the local Ethics Committee of the University of Münster.

Stimulus material

Participants watched pseudorandomly coloured single digits (0 – 9) presented for 500 ms in the centre of a light grey computer screen. Presentation frequencies for all colours and digits were equally distributed both within and across experimental blocks of approximately six minutes. Each block contained *ordered sequences* in which the previous digit was continually increased by one (Fig. 4-1A). Ordered sequences had a *regular* length of five digits and were embedded in *random trials* with no discernible relation between consecutive digits. In order to balance sequential starting points across digits, the ascending regularity necessarily included the 0 character and continued in a circular fashion after the figure 9 (e.g. 8 – 9 – 0 – 1 – 2).

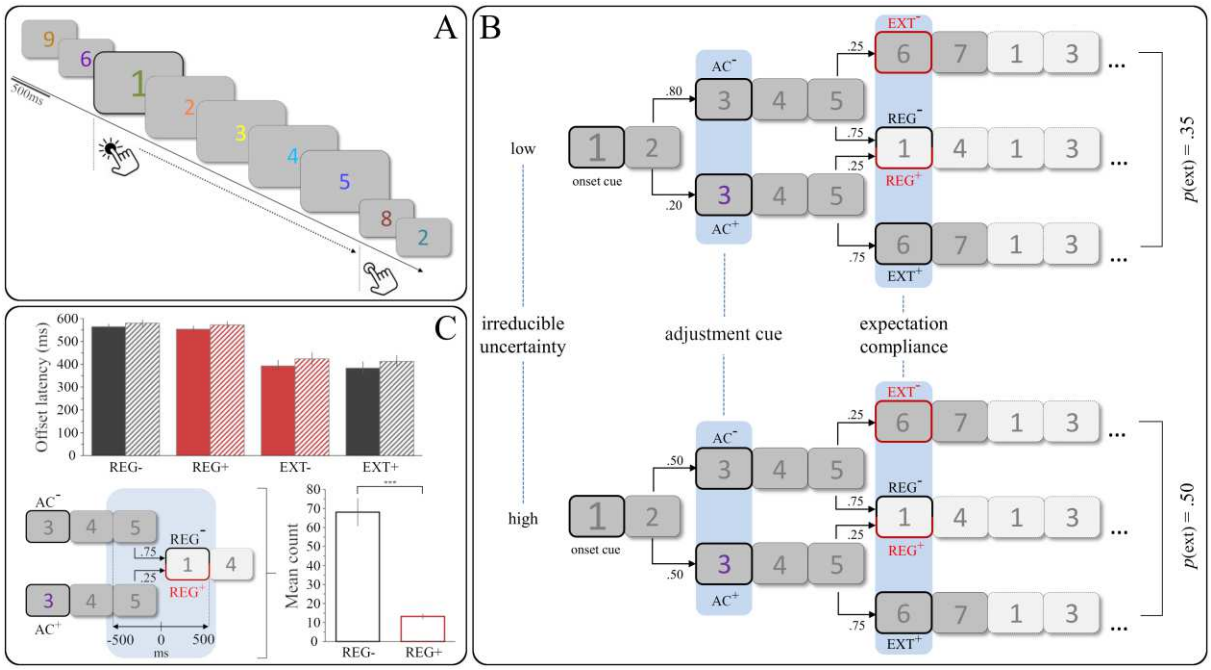

Figure 4-1. (A) Exemplary trial succession and time frame of the corresponding response for ordered sequences. (B) Experimental manipulations and resulting transition probabilities between trials. Bold framing indicates events of interest for EEG analyses; red framing indicates relatively unexpected events. (C) Top: Mean offset latencies for regular and extended sequences as a function of adjustment cueing and irreducible uncertainty (solid bars = low, hatched bars = high). For significance of main effects, please see main text. Bottom left: Schematic of sampling premature and quick button releases for uncued (REG-) and cued regular sequences (REG+). Bottom right: Mean count of releases within [-500, 500] ms for REG- and REG+ sequences. Error bars show standard error of the mean (SEM). AC = adjustment cue, REG = regular, EXT = extended, *** = $p < .001$.

The first digit of every ordered sequence was displayed at 150% of the usual font size, serving as an explicit *onset cue*. Importantly, digit colours at the third position of each sequence were used to vary expectations regarding its continuation. Each participant was unknowingly assigned one fixed colour that served as an implicit *adjustment cue* (AC): Presentation of the third sequential digit in AC colour (*AC+* trials) reliably indicated an *extension* of the respective sequence by two digits ($p_{EXT \mid AC+} = .75$, $p_{REG \mid AC+} = .25$). Conversely, when the third sequential trial was presented in any colour but the AC colour (*AC-* trials), extensions were rare ($p_{EXT \mid AC-} = .25$, $p_{REG \mid AC-} = .75$; see Fig. 4-1B). AC colours did neither occur at any other sequential position nor during random trials.

Finally, the composition of regular and extended sequences within a particular block was varied across blocks. This way, the *irreducible uncertainty* of a block was set to be either *low* or *high* (Fig. 4-1B). Low uncertainty blocks could be seen as statistically stable regarding the expected sequence length whereas highly uncertain blocks provided a less stable statistical context. To ensure constant AC validity, high uncertainty blocks necessarily contained a higher percentage of adjustment cues than low uncertainty blocks.

The experiment was programmed and run using the Presentation 18.1 software (Neurobehavioral Systems, San Francisco, CA, USA).

Task

Participants were instructed to press and hold the left button of a response box with their right index finger whenever an onset cue marked the beginning of an ordered sequence. At the end of an ordered sequence, participants were asked to release the response button as quickly as possible.

Experimental procedures

The study was conducted on two consecutive days. On the first day, participants completed a training session to provide them with implicit knowledge of the adjustment cues and the underlying statistical structure of the experiment. The training consisted of one high and one low uncertainty block with a total duration of approximately 13 minutes (including a one-minute break). On the second day, participants completed the EEG session which consisted of eight blocks (four blocks of each uncertainty level) and a total duration of approximately 56 minutes (including a one-minute break after each block). Participants were comfortably sitting on a chair in a darkened, sound-dampened and electrically shielded EEG booth. Experimental procedure and task during the EEG session were otherwise identical to the training session.

Behavioural data analysis

Statistical analyses of behavioural responses were performed in R (R Foundation for Statistical Computing, Vienna, Austria). First, correct and incorrect responses were aggregated separately for each participant. Incorrect responses were further divided into misses (no response over the course of a sequence) and false alarms (response occurring without presentation of sequential trials). Participants' overall performances were assessed via the discrimination index P_R (probability of recognition; Snodgrass & Corwin, 1988).

Offset latency was calculated as reaction time relative to the onset of the first random trial after a particular sequence. Reaction times occurring either before the go-cue or more than 1500 ms after the end of the sequence were excluded. Repeated-measures analyses of variance (ANOVA) and paired *t*-tests were used to assess potential differences in offset latencies as a function of expectation compliance, adjustment cues, and irreducible uncertainty.

EEG data analysis

EEG data acquisition and data preprocessing

Scalp EEG was recorded from 62 active Ag/AgCl-electrodes mounted in an actiCAP snap electrode cap (Brain Products, Gilching, Germany) using the BrainVision Recorder software (Brain Products, Gilching, Germany). All scalp channels were measured against a ground electrode at position FPz and referenced against FCz during recording. Two additional electrooculogram (EOG) electrodes were applied next to and below the right eye for the detection of horizontal and vertical eye movements. EEG was recorded at a sampling rate of 1 kHz with recording filters set to 0.1 – 1000 Hz bandpass.

EEG preprocessing was conducted in EEGLAB (Delorme & Makeig, 2004). Data segments containing experimental breaks were discarded prior to independent component analysis (ICA). Resulting components distinctly reflecting eye movements were subsequently rejected manually

(mean = 1.80 components) using the SASICA toolbox (Chaumon et al., 2015). Data were then filtered with a 0.1 Hz low cut and 30 Hz high cut filter and recalculated to common average reference. A time frame of [-200, 800] ms was defined for the analysis of event-related potentials (ERP). Epochs containing artefacts were discarded by semiautomatic inspection with an allowed maximum/minimum amplitude of ± 200 μV and voltage steps no higher than 50 μV per sampling. Channels with a high proportion of outliers (kurtosis criterion: $z > 6$) were replaced by a linear interpolation of their neighbour electrodes (mean = 2.13 interpolated channels).

Event-related potentials

Averages of the epochs representing events of interest were calculated separately per participant. The third trial of every ordered sequence was sampled to compare trials with and without the learned adjustment cues (AC+ vs AC-). Since ordered sequences had a regular length of five digits, the subsequent digit was critical with regard to expectation compliance: For regular sequences, this position contained the first random digit and marked the end of the ordered sequence (REG). For extended sequences, this position contained the sixth sequential digit and marked the extension of the ordered sequence (EXT). Both outcomes were further subdivided to reflect whether an adjustment cue had been presented at the third position (REG+, EXT+) or not (REG-, EXT-). Grand average ERPs across participants were calculated for all events of interest.

For stringent hypothesis testing, ERP analyses were restricted to specific time frames. Using the Mass Univariate ERP Toolbox (Groppe et al., 2011), ERPs from the respective conditions were submitted to a repeated measures cluster mass permutation test (Bullmore et al., 1999; family-wise significance level $\alpha = .05$). Repeated-measures *t*-tests were performed for each comparison using the original data and 5000 random within-participant permutations of the data. For each permutation, all *t*-scores corresponding to uncorrected *p*-values of $p = .05$ or less were then bound into clusters. The sum of

the *t*-scores in each cluster defined the "mass" of that cluster and the most extreme cluster mass in each of the 5001 sets of tests was used to estimate the distribution of the null hypothesis. Attempting to replicate previous N400 findings for expectation violations, we directly compared locally unexpected (EXT-, REG+) and expected outcomes (EXT+, REG-) within the N400 time frame (*UNEXP vs EXP*, 300 – 500 ms).

Multivariate classification analyses

Multivariate classification was carried out in Matlab using the CoSMoMVPA toolbox (Oosterhof et al., 2016). To increase signal-to-noise ratio (SNR), EEG time series were first downsampled to 100 Hz (see Carlson et al., 2013). SNR and classifier performance were further increased by continuously averaging n = 4 trials from the same respective categories before decoding (Grootswagers et al., 2017; Isik et al., 2013). Classification was then performed separately for each 10 ms bin by means of a temporal searchlight analysis approach. Specifically, a linear discriminant analysis (LDA) classifier was trained and tested on trials from two categories (e.g., AC+ and AC- trials) using a leave-two-out cross-validation scheme. This way, classification analyses yielded time courses of decoding accuracy with 10 ms temporal resolution within the interval of [-200, 800] ms. Threshold-free cluster enhancement (TFCE; Smith & Nichols, 2009) with default parameters was used to determine periods of significant decoding accuracy (with a fixed chance level of $p = .50$ for pairwise decoding). As implemented in CoSMoMVPA, permutation testing with 10,000 null iterations and subsequent thresholding at $Z > 1.96$ ($p < .05$) was used to correct for multiple comparisons.

With regard to functional event signatures, two comparisons were the focus of pairwise decoding analyses. First, we assessed the classifier's performance decoding AC+ from AC- trials over the ERP time course (see Fig. 4-1B). Third-position trials containing an adjustment cue (AC+) provided excess information with regard to the continuation of the sequence, effectively overriding expectations

based on the global task structure. Exploitation of this particular cueing information should translate to distinct electrophysiological correlates when compared to non-cue trials at the same position (AC-). As we had previously reported P3b effects for the comparison of model-compliant, particularly informative events with non-informative equivalents (Kluger et al., submitted), we hypothesised significant classifier accuracy of *AC+ vs AC-* within the P3b time frame (300-500 ms).

Our second comparison concerned decoding of checkpoints (CP) within the same P3b time frame. The previous finding of a CP-induced P3b component as an index of excess information (see above) raised the intriguing question of whether this potential is modulated by global parameters such as context uncertainty. In the absence of adjustment cues within the original paradigm, our initial fMRI study had established amplitude effects of CP processing as a function of high vs low uncertainty (Kluger & Schubotz, 2017). Consequently, we attempted to decode CP under high vs low uncertainty within the new experimental paradigm. To this end, we sampled regular endings of sequences that did not contain an adjustment cue (REG-), as these events were the precise equivalent of CP from earlier studies. Consolidating previous findings from fMRI and EEG, we thus hypothesised significant classifier accuracy of *CP high vs CP low* within the P3b time frame (300-500 ms).

Representational Similarity Analysis

Whereas the decoding analyses described so far have allowed us to detect category-specific information in the EEG signal, RSA (Kriegeskorte et al., 2008) provides a framework to test hypotheses about the underlying representational structure of activity patterns. Information carried by a given representation (e.g., of event classes) can be quantified in so-called representational dissimilarity matrices (DSMs). Following the assumption that events with distinct representations are fairly easy to decode (and vice versa), these matrices characterise event representations based on their (dis-) similarities among one another (for review, see Haxby et al., 2014). For hypothesis-based

model comparison, multiple hypothetical candidate DSMs can be devised a priori to predict the empirical similarity structure found in the data. By correlating model and empirical DSM, one can then assess the degree to which the model-inherent structure does indeed exist in the activity patterns (in other words, the explanatory power or fit of the respective model DSM).

Two binary model DSMs were set up to predict the empirical representational similarity of event classes in the EEG data (Fig. 4-3B). The first model DSM (termed *global model*) categorically assumed high similarity within and high dissimilarity between the sampled event classes (AC+/- vs REG+/- vs EXT+/-) and did not account for potential adjustments made in response to incidental adjustment cues. In contrast, the second model DSM - termed *local model* - assumed high similarity of events that violated potentially adjusted predictions following an adjustment cue (REG+, EXT-) in contrast to those that complied with the adjustment cue (REG-, EXT+).

4.4 Results

Behavioural results

Participants showed an overall high level of performance with a mean P_R score of $M_{PR} = .97$ ($SD = .02$) during the EEG session, indicating excellent attentiveness throughout the experiment. The repeated-measures ANOVA yielded significant main effects of all three factors on offset latency (*expectation compliance:* $F(1, 29) = 53.63$, $p < .001$; *adjustment cueing*: $F(1, 29) = 19.20$, $p < .001$; *irreducible uncertainty:* $F(1, 29) = 4.74$, $p = .003$). Neither interaction term reached statistical significance (all $p > .254$). First, replicating behavioural findings from earlier studies (Kluger & Schubotz, 2017; Kluger et al., submitted), a post-hoc *t*-test revealed significantly quicker button releases after extended (M = 403 ms, SD = 147 ms) than after regular sequences (M = 567 ms, SD = 77 ms; $t(29) = -7.32$, $p < .001$). Second, participants released the response button significantly quicker

when irreducible uncertainty was low (M = 473 ms, SD = 98 ms) than when it was high (M = 497 ms, SD = 104 ms; $t(29) = -4.38$, $p < .001$). Finally, significantly shorter offset latencies were found for sequences that contained an adjustment cue (M = 480 ms, SD = 104 ms) than for those that did not (M = 490 ms, SD = 97 ms; $t(29) = -2.18$, $p = .038$, see Fig. 4-1).

Building on analyses of an earlier study (Kluger et al., submitted), we assessed anticipatory and particularly quick reactions to regular sequence endings as a behavioural marker of adjustment cue recognition. Following event definitions of previous work, these reactions were defined as button releases during the last sequential and the first non-sequential digit (i.e., offset latency ± 500 ms, see Fig. 4-1C). Since the adjustment cue reliably indicated sequential extensions, participants should respond less frequently to REG+ (i.e. unexpected regular endings) than to REG- sequences (expected regular endings) in this time frame after they successfully learned the adjustment cue. A paired *t*-test supported this hypothesis, revealing significantly fewer responses for REG+ (M = 13.20, SD = 8.40) than for REG- sequences (M = 68.07, SD = 40.25) within the [-500, 500] ms interval ($t(29) = 9.27$, $p < .001$, see Fig. 4-1).

Event-related potentials

UNEXP - EXP

Building on previous EEG results (Kluger et al., submitted), we first aimed to replicate findings of an N400 component for the direct comparison of unexpected and expected outcomes. Accordingly, we included all time points between 300 and 500 ms in a one-sided whole-brain analyses (6262 comparisons in total). As hypothesised, unexpected outcomes (EXT-, REG+) were found to elicit a significant N400 over parieto-central electrodes peaking around 442 ms (Fig. 4-2).

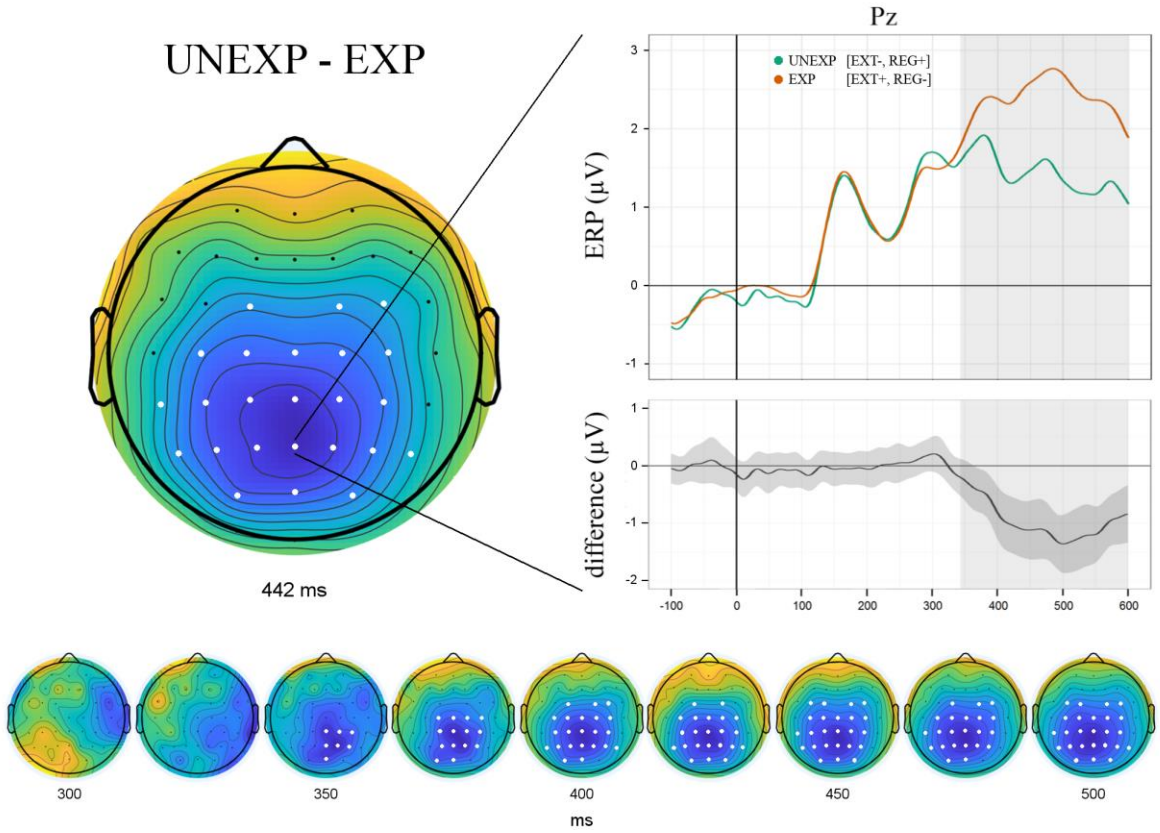

Figure 4-2. The direct comparison of unexpected (EXT-, REG+) and expected outcomes (EXT+, REG-) revealed a significant N400 component peaking around 442 ms over parieto-central electrodes. Difference curve includes M ± STD of individual data, bottom panel shows component evolution over time.

Multivariate classification

AC+ vs AC-

In a first classification analysis, we trained the classifier to decode third-position events that contained an adjustment cue (AC+) from those that did not (AC-, see Fig. 4-1B). Group-level classification analysis revealed temporal clusters with significant decoding accuracy between 140 - 310 ms as well as between 330 – 420 ms (Fig. 4-3A). Intriguingly, the second cluster particularly supported our hypothesis regarding the P3b time frame for this comparison. In addition, two brief significant clusters were found from 570-590 ms and from 770-800 ms. However, since these time frames already fell within the presentation of the next stimulus, these results should be interpreted with caution.

CP high vs CP low

Extending the narrative of previous studies (see Kluger et al., submitted) on modulation by context statistics, the second classification analysis concerned checkpoints (CP) under high and low uncertainty, respectively. As the present experimental design featured locally informative adjustment cues, regular sequence endings with no prior adjustment cue (REG-) immediately corresponded to CP events sampled in previous studies (Kluger & Schubotz, 2017; Kluger et al., submitted). Accordingly, we trained the classifier to decode these CP in high uncertainty contexts from those in low uncertainty contexts. As hypothesised, group-level classification analysis revealed significant temporal clusters within the P3b time frame (360-380 ms). A second small cluster was found for the time frame between 440-450 ms (Fig. 4-3A).

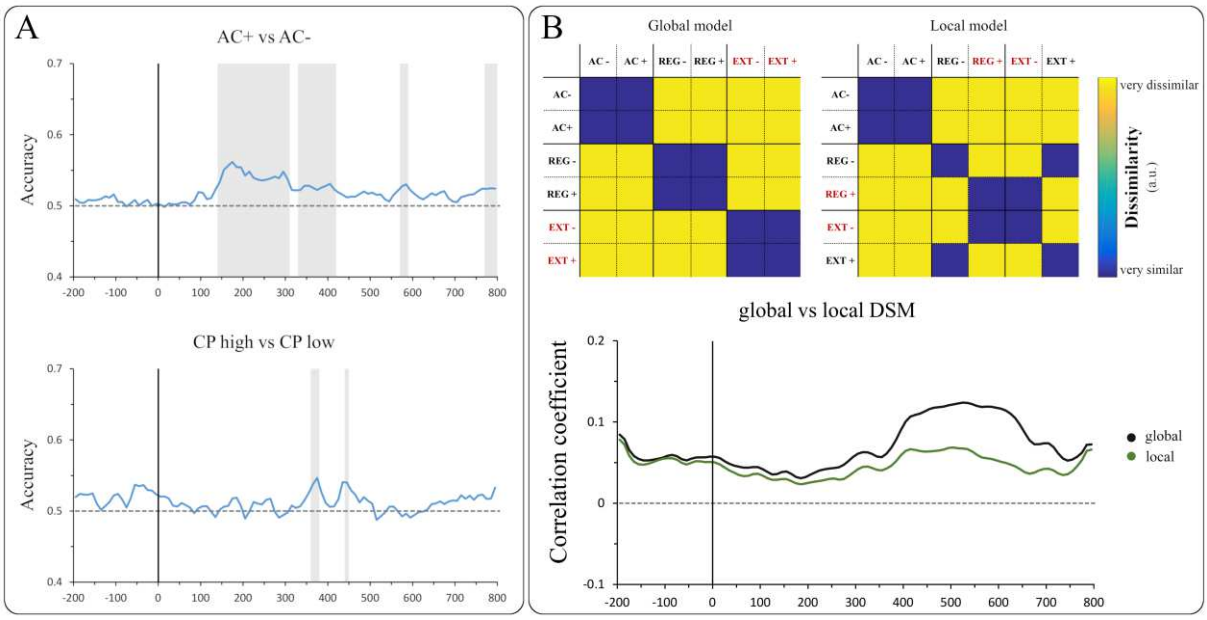

Figure 4-3. (A) Group-level decoding accuracies over time. Shaded areas indicate significant temporal clusters. (B) Top: Model DSMs predicting representational similarity in the EEG data based on distinct event characteristics. Globally (left) and locally (right) unexpected events are marked in red. Bottom: Group-level correlations of both model DSMs with the empirical DSM. AC = adjustment cues, CP = checkpoints, REG = regular endings, EXT = sequential extensions.

Representational Similarity Analysis

We employed searchlight RSA to compare two binary dissimilarity matrices (DSMs) modelling representational similarity of event classes in the EEG data. Both the global and the local model were found to be significantly correlated with the empirical DSM over the entire time course of [-200, 800] ms (Fig. 4-3B). While a statistical comparison of model DSMs in RSA remains difficult (Kriegeskorte & Kievit, 2013; Thirion et al., 2015), the global model clearly had a better fit to the EEG data than the local model later in the time series (around 400-700 ms). The descriptive comparison of the two model DSMs thus suggests that the global event structure acquired through statistical learning took precedence over on-line information conveyed by intermittent adjustment cues. In other words, statistical learning of event categories (REG, EXT) was not cancelled out by temporary shifts in expectations but remained the strongest predictor of overall similarity in EEG activation patterns.

4.5 Discussion

Predictive brain processing vitally relies on top-down hierarchical models continuously matched against incoming sensory information. In case of deficient model assumptions, mismatch signals termed prediction errors (PE) are propagated upstream to instigate model updating for future occasions. Recent fMRI and EEG studies (Kluger & Schubotz, 2017; Kluger et al., submitted) have demonstrated exploitation of information provided by model-compliant events (checkpoints, CP), particularly in contexts of high uncertainty. Building on previous work highlighting such global modulations in predictive processing, the present study was conducted to assess the influence of locally available information. Therefore, we introduced adjustment cues which interfered with the learned statistical task structure, effectively calling for a flexible on-line adjustment of expectations.

Multivariate classification analysis revealed that adjustment cues could be reliably decoded from uninformative events at the same position within the P3b time frame. Moreover, events that were unexpected globally (according to task structure) or locally (as indicated by adjustment cues) elicited a mismatch-related N400 component when compared to expected events. Combined with behavioural data, these findings demonstrate that local cueing information was efficiently exploited to adapt to current task requirements.

Aside from on-line adjustments prompted by local cues, global uncertainty was found to modulate the functional signature of checkpoints: Consolidating previous evidence from fMRI (Kluger & Schubotz, 2017) and EEG (Kluger et al., submitted), checkpoints were distinctly processed under high (vs low) uncertainty within the P3b time frame, as indicated by significant multivariate classification accuracy. Finally, representational similarity analysis (RSA) was used to compare two dissimilarity models representing global and local contingencies, respectively. While both models were found to be significantly correlated with the empirical dissimilarity matrix, the global model was found to fit the neural data better at a later time frame (400 – 700 ms). Conceivably, higher cognitive operations reflected in this time frame are more likely to incorporate global information acquired through long-term statistical learning.

Local information prompts on-line adjustments

It was one of the main aims of this study to assess the influence of two separate sources of information in predictive processing. First, in the present experimental context, global task structure and transition probabilities are acquired through statistical learning, starting with the training session. Higher-order information also comprised different levels of irreducible uncertainty, that is, how

frequently implicit expectations regarding sequence length ("... usually five digits long") were violated by sequential extensions. Secondly, however, local information was introduced in the form of adjustment cues reliably indicating a specific outcome (i.e., an extension of that sequence). This information thus modified the statistically learned contingencies and called for a rapid adjustment of expectation in an ongoing sequence.

Effective exploitation of adjustment cues was reflected in behavioural response patterns at regular sequence ends: Participants responded less frequently to unexpected regular endings (REG+) within a time frame of ± 500 (compared to regular endings, REG-). This time frame was used in a previous study to mark anticipatory and particularly quick reactions, indicating behavioural facilitation effects of cue recognition (Kluger et al., submitted). In the present study, the preceding adjustment cue had reliably indicated an extension, rendering the globally more probable regular ending locally improbable. Recognising adjustment cue information, participants were conceivably less confident to prepare an anticipatory or quick release of the response button. This finding fits well with previous results consistently showing response slowing for sequences that - much like present REG+ trials - ended earlier than indicated by the preceding cue (Kluger & Schubotz, 2017; Kluger et al., submitted).

On the neural level, two key findings provided critical insight into the role of adjustment cue processing: First, supporting our hypothesis, ERP analysis revealed a significant N400 component for locally unexpected (vs expected) sequential events. Critically, unexpected events comprised REG+ and EXT- trials, thus accounting for adjusted expectations due to cue information[1]. This finding replicates results from an earlier study (Kluger et al., submitted) in which we established N400 as the proper mismatch correlate in sequential prediction. The N400 component has been reported to be elicited by a multitude of sensory events and varies in amplitude as a function of event-bound surprise

[1] Note that expected (REG-, EXT+) and unexpected (REG+, EXT-) event classes were defined by means of a weighted average of trials, precluding ERP effects to be caused by mere differences in trial numbers.

in language (Rabovsky et al., 2018; Szewczyk & Schriefers, 2018) and arithmetic problems (Galfano et al., 2004; Niedeggen et al., 1999). More generally, it is discussed as a modality-independent index of conceptual representations which need revision in light of violations (for review, see Kutas & Federmeier, 2011). In sum, our interpretation of N400 as a correlate of locally induced mismatches further corroborates the component's role in integrating incongruous information.

Second, as hypothesised, multivariate pattern classification revealed significant decoding of AC+ and AC- trials within the P3b time frame. In the closely related paradigm of our previous study (Kluger et al., submitted), we were able to show P3b as the joint correlate of probabilistic, informative events irrespective of their outcome, namely checkpoints and prediction errors. We interpreted this finding as an index of excess information use in accordance with previous work showing that P3b is by no means a mere mismatch detector, but rather facilitates adequate responses following behaviourally relevant stimuli (Nieuwenhuis et al., 2011; Verleger et al., 2005). A very similar interpretation was evident for adjustment cues in the present paradigm: Information gained from adjustment cues modified global contingencies and facilitated preparation of the adequate response (i.e., most probably keeping the finger on the response button for an extended sequence).

Adjustment cues (AC+) thus bear a striking conceptual resemblance to checkpoints (CP) defined in previous studies: In both cases, model-compliant sensory input of probabilistic events can be exploited to evaluate the predictive model in the absence of a prediction error. Critically, as reflected by significant decoding within the P3b time frame, this was not the case for uninformative third-trial events (AC-). Sensory input at these positions was model-compliant but did not provide excess local information, making AC- trials comparable to non-informative standard trials.

Given the conceptual similarities of checkpoints and adjustment cues, it was interesting to see global

effects of contextual uncertainty on CP processing also comprising the P3b time frame, spurring promising questions for future research (see Limitations and future directions).

Global effects of contextual uncertainty

To assess the effect of irreducible uncertainty on checkpoint processing, we once again employed multivariate pattern classification for decoding of CPs under high and low uncertainty, respectively. In order to increase comparability between studies, we restricted the present analysis to checkpoints as they were defined in earlier work, namely regular endings of sequences that did not contain an adjustment cue (REG-). Supporting our hypothesis, we found clusters of significant decoding accuracy within the P3b time frame. This finding considerably adds to earlier results from two directions: First, as mentioned above, our previous EEG study revealed P3b as the correlate of excess information gain at checkpoints, effectively separating it from non-informative standard trials (Kluger et al., submitted). Second, a previous fMRI study had established a functional network exhibiting increased neural activity for CP processing under high (vs low) irreducible uncertainty (Kluger & Schubotz, 2017). Notably, this network prominently included the temporo-parietal junction (TPJ), a conceivable cortical source of P3b (Donchin & Coles, 1988). Altogether, the present finding suggests a modulatory influence of contextual uncertainty on information exploitation at checkpoints which translates to distinct correlates within the P3b time frame: As high uncertainty blocks were defined by a higher proportion of extended sequences, validity of the initial model ("... usually five digits long") was reduced in these blocks. Therefore, sites of supposed sequence endings were particularly informative whether a motor response was in fact required (REG-) or not (EXT-). In line with accounts of contextual updating ascribed to both P3b (Kimura et al., 2010; Verleger et al., 2016) and its suggested source in TPJ (Geng & Vossel, 2013), we thus interpret the present uncertainty effect as a correlate of excess information being exploited to estimate global context

features and adapt behaviour accordingly. This closely relates to our proposal of stepwise processing of sequential input, evaluating model validity at CP as informative reference points (Kluger & Schubotz, 2017).

It is worth noting that, since AC+ trials reliably indicated extensions, high uncertainty blocks necessarily contained a higher absolute number of adjustment cues than low uncertainty blocks to ensure constant cue validity (see Fig. 4-1B). In other words, whereas the global portion of extended sequences was higher in these blocks, local probabilities after the fifth sequential position remained unchanged. The finding of significant checkpoint decoding under high vs low uncertainty thus suggests that local information did not fully cancel out global contingencies and that the role of CP as informative reference points should be even more evident in contexts without local adjustment cues (see Limitations and future directions).

Representational similarity in predictive processing

Representational similarity analysis of a local and a global model DSM revealed significant correlation with neural data for both models. The apparent pre-eminence of the global model in a later time frame (400 - 700 ms), however, leaves room for discussion and motivates further research. Critically, the two models differed with regard to which events were jointly classified as unexpected: The local model reflected adjusted probabilities as a function of on-line cues, grouping REG+ and EXT- occurrences as locally unexpected events. In contrast, the global model did not account for any adjustments but rather modelled all extensions (EXT+, EXT-) as similar and as distinct from regular endings (REG+, REG-).

The differential modelling of unexpectedness in the model DSMs as well as the late timing of the difference in model fit strongly suggest these RSA results to be related to the N400 finding reported above: While locally unexpected events could be reliably decoded from expected events in the

preceding classification analysis, the higher fit of the global DSM shows that - across all event classes - late potentials reflected an even higher (dis-)similarity of globally unexpected (vs expected) events. In other words, whereas local information was undoubtedly exploited for on-line adjustments in prediction and behaviour, it did not fully overwrite statistically learned global contingencies. Fittingly, data from previous N400 studies revealed that when information from two different levels (such as word- and sentence level) is available, higher-level context effects tend to take precedence over lower-level ones (Kutas, 1993; van Petten, 1993). Indeed, frequently re-learning by overriding global probabilities seems quite costly in terms of computational efficiency. Instead, we may establish a global situational model (in this case reflecting the expectable sequence length) which can be locally modified - but not overturned - in the light of additional information. In line with interpretations of N400 as an index of information integration (which is more effortful for incongruous input, see above), our findings overall indicate that the effort of integration is invested in a model of global contingencies rather than representing local fluctuations.

Limitations and future directions

The aim of the present study was to investigate the interplay of local and global information in predictive processing. To this end, we employed local adjustment cues which were incidentally presented to interfere with statistically learned, global contingencies. In order to maximise frequency for the events of interest, we were not able to analyse sequential standard trials due to their close temporal proximity to other sampled events. Given the significant decoding of AC+ and AC- within the P3b time frame, a comparison of each event class with non-informative sequential standards would add considerably to the understanding of the functional role of adjustment cues. Conceivably, AC+ and CP as informative, model-compliant events should elicit a P3b component when compared to standard trials. In contrast, AC- trials - due to the lack of excess information - should be more

similar to non-informative standard trials and therefore not be reflected in a P3b.

In a similar vein, manipulating contextual uncertainty irrespective of adjustment cues would allow us to further understand the extent to which checkpoint information is exploited across environments. It would thus be revealing to assess CP processing in an experimental paradigm which includes blocks of high and low uncertainty both with and without adjustment cues.

4.6 Conclusion

Model-compliant events have previously been established as critical reference points informing predictive processing in the absence of prediction errors. A similar functional role on the local level is now suggested for a different class of model-compliant events: Accounting for information from local *adjustment cues*, we were able to replicate a significant N400 component as an index of mismatch-induced model adaptation. Furthermore, using multivariate pattern classification, sequential positions providing excess local information could be reliably decoded from those that did not within the P3b time frame. Global effects of irreducible uncertainty on checkpoint processing remained, as reflected by significant classification of high vs low uncertainty checkpoints in the same time frame. Overall, representational similarity of a global model was found to fit the neural data better than a local model, suggesting that adjustments reflected by late ERPs (e.g., N400) are made to a higher-level model of global contingencies. Intriguing research questions remain with regard to the precise interplay of local vs global information as well as the functional relationship of adjustment cues and checkpoints.

4.7 Acknowledgements

We would like to thank Monika Mertens, Katharina Thiel, Corinna Gietmann, and Marie Kleinbielen for their help during data collection.

5. General discussion

5.1 General summary

The three studies presented in the previous chapters aimed to successively illuminate central aspects of abstract predictive processing. Broadly, we intended to build a functional profile of model-congruent and -incongruent event processing and assess the extent to which exploitation of informative events varies as a function of local and global parameters.

To start with, study 1 set out to pursue two questions: First, we aimed to assess the neural underpinnings of qualitatively distinct non-reward prediction errors. Using fMRI and a sequential pattern detection task, we were able to show differential frontoparietal components for premature terminations (inferior frontal gyrus) and unexpected extensions (superior frontal sulcus, posterior cingulate cortex, and angular gyrus). In line with previous research, these findings were interpreted to distinctly reflect specific rule violations (terminations) or an attentional reorienting process in order to resume the internal model (extensions), respectively. Second, we investigated predictive processing in variably uncertain contexts. Recurring to open questions regarding predictive efficiency raised in the beginning, it was our hope to comprehend how top-down predictions are strategically adapted to highly uncertain contexts. To this end, we proposed so-called *checkpoints* as particularly informative reference points during a sequence: At these sequential positions, prediction errors could potentially occur, but in fact did not. Therefore, checkpoints were conceptualised as probabilistic (and thus informative) events which did not violate the internal model, making them a promising candidate to inform model validity under uncertainty. Supporting this hypothesis, the direct contrast of high vs low uncertainty checkpoints revealed increased activation in a right-lateralised frontoparietal network comprising inferior frontal gyrus and temporoparietal junction (TPJ). Corroborating previous accounts linking TPJ function to contextual updating, we suggested this

pattern of uncertainty-related excess activation to indicate strategic adaptation of predictive processing to unstable environments: High contextual uncertainty presumably resulted in stepwise prediction of upcoming sequential input, iteratively using checkpoint information for model evaluation. Findings from Eigenvector centrality mapping suggested that this adaptation of predictive strategy may be conducted by means of enhanced functional connectivity between frontoparietal circuits and the parahippocampal area.

Exploiting the temporal benefits of EEG, it was the central objective of study 2 to further characterise the functional profile of checkpoints. Specifically, we aimed to elaborate ERP correlates of conceptual commonalities and distinctions between checkpoints and prediction errors. First, when compared to non-informative standard trials, both checkpoints and prediction errors are informative in their probabilistic occurrence and may be used in similar ways to evaluate or adapt the situational model. We found this joint quality to be reflected in a P3b component for both checkpoints and prediction errors (vs STD, respectively). Second, it remained the fundamental distinction between these two event classes that prediction errors violated the internal model whereas sensory input at checkpoints did not. Fittingly, we were able to show a significant N400 as an index of the mismatch signal induced solely by prediction errors (vs checkpoints). Overall, findings from study 2 highlighted checkpoints as informative, model-compliant reference points exploited for model evaluation.

As we had already established the influence of global context features (i.e., uncertainty) on checkpoint processing in study 1, the question remained to what extent local factors potentially modulated predictive processing. Naturally, this implicated intriguing questions with regard to the interplay of local vs global information gain. Therefore, study 3 used a variation of the first two studies' paradigm which now involved incidental adjustment cues. These cues reliably indicated extensions (i.e., otherwise unexpected sequence endings) and, by doing so, locally interfered with the global statistical task structure. Using ERP analysis and multivariate classification, we were able to

consolidate previous findings from two directions: First, taking adjustment cue information into account, we replicated a significant N400 for locally unexpected vs expected sequence endings (cf. study 2). Similarly, extending P3b findings from study 2, exploitation of cue information was reflected in significant decoding accuracy of adjustment cues vs non-informative analogues within the P3b time frame. Second, adding to modulatory effects of context uncertainty from study 1, classification analysis revealed significant decoding of high vs low uncertainty checkpoints within the P3b time frame.

Finally, representational similarity analysis showed that within the time frame of 400 - 700 ms, similarities between event-related potentials were better represented by a *global* model reflecting overall task contingencies (compared to a *local* model reflecting on-line changes therein). Accordingly, while functional and behavioural data conclusively demonstrated recognition and exploitation of cue information for on-line adjustments, global contingencies fundamentally remained the frame of reference. Overall, these findings thus not only bring together individual results from all three studies, but raise compelling questions regarding the mutual influence of local and global information in predictive processing.

In what follows, I will first reconcile key takeaways from studies 1 to 3 in order to argue the bigger picture of non-reward prediction errors and model-compliant information in predictive processing. Relatedly, I will then discuss different scales at which information from a variety of events may be exploited and how this affects predictive efficiency. Finally, I will critically reflect on limitations of studies 1 to 3 and suggest future directions before concluding this thesis.

5.2 Implications on non-reward prediction errors

While a central role of prediction errors for model updating is widely agreed upon, functional characterisation of non-reward prediction errors has merely attracted little attention. Study 1 aimed to provide a functional distinction of two qualitatively different kinds of abstract prediction errors, namely sequential terminations and extensions. To maximise computational efficiency in predictive processing, model adaptations due to prediction errors have to be precise, directed adjustments of model parameters. This is particularly true when current situational contexts demand immediate behavioural responses, as was the case in the paradigm of study 1: Terminations called for an unexpectedly early release of the response button whereas extensions required participants to unexpectedly continue pressing the button.

Our finding of inferior frontal gyrus / BA44 activation for terminations (vs extensions) adds to a rich body of literature of IFG involvement in signalling violations of expected regularities across domains, e.g. language (Friederici & Kotz, 2003), actions (Wurm & Schubotz, 2012), and music (Maess et al., 2001). Conversely, the network we report for extensions (vs terminations) comprised superior frontal sulcus and angular gyrus which are jointly taken to reflect memory-directed reorientation and/or updating of attention (Gottlieb, 2007; Kincade et al., 2005; Schubotz et al., 2012). Overall, we interpret these results to indicate an unexpected resumption of the internal model in the face of surprising sequential digits.

In order to subsequently characterise checkpoints and prediction errors in distinction to one another, study 2 conflated both terminations and extensions into one PE category. The direct comparison of prediction errors and checkpoints revealed a significant N400 component evoked by PE. N400 has been linked to mismatch signals in semantic tasks (reviewed in Kutas & Federmeier, 2000), recognition memory (reviewed in Friedman & Johnson, 2000), and the arithmetic domain (e.g.,

Niedeggen et al., 1999), eventually leading to a more general description of N400 as a reflection of conceptual representations which need updating (Federmeier & Laszlo, 2009; Kutas & Federmeier, 2011). In line with this framing of N400, we interpreted findings from study 2 to reflect mismatch-induced model updating. Intriguingly, results from study 3 further advanced this interpretation. Here, prediction errors were once again aggregated for an ERP analysis of unexpected vs expected events. Importantly, however, this categorisation included local contingency changes due to adjustment cues: As these cues reliably indicated sequential extensions, globally expected events (i.e., regular endings) would become locally unexpected and vice versa. We found a significant N400 component for globally (i.e., uncued extensions) and locally (i.e., cued regular endings) unexpected outcomes when compared to expected events. Comparing these unexpected events to studies 1 and 2, cued regular endings from study 3 paralleled previous terminations (i.e., less sequential input than expected) whereas uncued extensions corresponded to previous extensions (i.e., longer sequential input than expected). While prediction errors in the first two studies violated global contingencies acquired through statistical learning, consistent N400 results for violations of global *and* local contingencies from study 3 highlight the component's greater involvement in denoting representations in need of refinement or updating. The apparent independence of informational sources further supports one of the central notions of predictive processing, namely that prediction errors prompt model adaptations on various time scales ranging from rapid perceptual inference to slower perceptual learning (Clark, 2013).

In sum, we were able to establish a functional distinction of sequential terminations and extensions in study 1, reflecting respective adaptation towards behavioural consequences. When aggregated, prediction errors uniformly evoked a mismatch-related N400 component, irrespective of the time scale on which contingencies were violated (see study 3). In the following sections, I will address the role of model-compliant information and how it benefits predictive efficiency across time scales.

5.3 Model-compliant information in predictive processing

So far, the three studies outlined in the previous chapters have established converging roles of two types of model-compliant information in predictive processing: First, *checkpoints* were initially introduced in study 1 as reference points within a sequence which were particularly informative in highly uncertain contexts. In other words, when statistically unstable environments lowered confidence in model validity, downstream predictive codes would be strategically adapted to use checkpoints for stepwise model evaluation. Critically, since sensory input at checkpoints did not violate model-based predictions, adaptation in response to contextual uncertainty was clearly not induced by mismatch signals.

Second, *adjustment cues* were employed in study 3 to provide reliable local information regarding sequence continuation, thus interfering with global statistical structures. As reported above, we were able to reliably decode adjustment cues from non-cues at the same sequential position, particularly within the P3b time frame. While the interplay of local and global information deserves separate attention (see *5.4 Predictive efficiency across time scales*), joint results from all three studies provide critical insight into functional characteristics of model-compliant events in predictive processing.

A key role in understanding these characteristics falls to the P3b as the common denominator of exploiting informative events. P3b has consistently been reported for informative, task-relevant stimulus evaluation (Huang et al., 2015; Katayama & Polich, 1998; Nieuwenhuis et al., 2011). Critically, and in opposition to several predictive processing accounts, our findings emphasise that stimuli do not necessarily have to violate model assumptions to be informative. Previous studies have shown P3b to be evoked by an informative sequence of standard trials acting as a 'secondary target' (Fogelson et al., 2009) and even in the absence of stimuli if the omission was informative (Nieuwenhuis et al., 2011). Relatedly, when we aimed to distinguish informative checkpoints from

perceptually identical but non-informative standard trials in study 2, a hypothesised P3b component emerged as the index of excess information provided by checkpoints. The parallel comparison of prediction errors vs standard trials revealed a strikingly similar P3b component for PE, emphasising related roles for checkpoints and prediction errors in predictive processing. Based on accumulated findings of studies 1 and 2, we suggested *model evaluation* as the checkpoint-induced counterpart to *model adaptation* widely ascribed to prediction errors (Clark, 2013; Rao & Ballard, 1999): Uncertainty-related activation within TPJ, a suggested cortical source of P3b (Mengotti et al., 2017; Volpe et al., 2007), had established checkpoints as context-sensitive points of reference in study 1, presumably reflecting their exploitation for stepwise processing (see above). In addition to increased TPJ activation for high (vs low) uncertainty checkpoints, study 1 had also shown heightened functional connectivity of parahippocampal cortex during epochs of high uncertainty. Consequently, we suggested stepwise processing in unstable environments to rely on model information retrieved from working memory. This interpretation gains substantial support from previous work declaring P3b a marker of working memory operations initiated in the hippocampal formation whose updated output is then transmitted to parietal cortex (Knight, 1996; Squire & Kandel, 1999).

Finally, Study 3 further refined our understanding of P3b as an index of evaluative information use in light of global uncertainty modulation. Significant multivariate pattern decoding of high vs low uncertainty checkpoints within the P3b time frame provided the missing link between fMRI and EEG findings from studies 1 and 2. Furthermore, replicating the logic of the CP vs STD comparison from study 2, we were able to reliably decode newly introduced adjustment cues from non-informative equivalents. Like checkpoints, adjustment cues provided critical, task-relevant information with regard to the continuation of the respective sequence - a long-standing, central factor of influence on P3b amplitude (Goldstein et al., 2002; Patel & Azzam, 2005). Overall, P3b appears to be evoked by processing of excess information from model-congruent and -incongruent

outcomes, irrespective of the time scale on which this information is used: Checkpoints were shown to be related to higher-level contextual features (i.e. global uncertainty), whereas adjustment cues indicated short-term, local changes of contingencies. Prediction errors, which also elicited a P3b when compared to standard trials, can prompt model adaptation on a variety of time scales (Clark, 2013). As the temporal scope of P3b has been shown to range from stimulus-bound surprise (Mars et al., 2008; Bennett et al., 2015) up to maintenance of broader context information (Kimura et al., 2010; Polich, 2007), we suggest model-compliant events to inform predictive processing on an equally variable time scale.

To summarise, while potential benefits of model-compliant information are oftentimes overlooked in predictive processing accounts, the three studies presented here draw a coherent picture of how non-error events are used for model evaluation. Consistent findings of P3b and its involvement in transmitting information between (para-)hippocampal areas and TPJ demonstrate the component's role in model evaluation across time scales: While information from adjustment cues was exploited to learn short-time contingency changes on-line, checkpoints were informative with regard to broader contextual features. This way, both local and global sources of information appear to be exploited to inform predictive processing for the sake of predictive efficiency.

5.4 Predictive efficiency across time scales

It is one of the core merits of the predictive processing framework to provide a comprehensive account of computational efficiency in brain function. As highlighted in the introduction, its hierarchical structure allows parsimonious comparisons of model predictions and sensory information at every stage of the processing hierarchy, effectively reducing bottom-up signal to unexplained prediction errors (see *1.2 The predictive brain*). These mismatch signals prompt model updating on different time scales: Automatic adjustments of probabilistic representations at the next highest level allow for rapid perceptual inference whereas the very structure of the internal model may be updated for a better fit on future occasions (Clark, 2013).

This error-bound conceptualisation is indisputably well-suited to explain model adaptation following both locally unexpected events and global contingency changes. Extending the narrative of predictive efficiency, this dissertation aimed to examine how top-down predictive codes may be strategically adapted to varying contexts and how different sources of information would be used in the process.

In addition to prediction errors, two classes of model-compliant events emerged as significant reference points informing predictive processing. First, checkpoints were established as points in time where model-compliant information was exploited for stepwise evaluation of model validity under high uncertainty (see study 1). Based on the pattern of uncertainty-related fMRI results, we suggested such stepwise predictive processing to be rooted in iterative evaluation of model information retrieved from working memory within the (para-)hippocampal cortex. Supposedly, this information is then transmitted to parietal cortex (angular gyrus / TPJ) to inform appropriate updating of contextual information. Consistent P3b results for checkpoints from studies 2 and 3 substantially corroborated this interpretation, as P3b has been linked to both contextual updating (Kimura et al.,

2010; Polich, 2007) and information transmission from hippocampal areas to TPJ (Knight et al., 1989; Verleger et al., 1994; Yamaguchi & Knight, 1992).

The second class of non-error events used to inform predictive processing were adjustment cues presented during an ongoing sequence. As described in more detail above, these cues interfered with the global task structure by providing reliable on-line information with regard to the outcome of respective sequences. Just as prediction errors and checkpoints, adjustment cues were found to evoke distinct potentials within the P3b time frame (compared to non-informative equivalents), as revealed by multivariate pattern classification in study 3.

Although intriguing research questions remain to be addressed in future studies (see below), it is worth considering the implications of joint P3b findings for both model-congruent (checkpoints, adjustment cues) and -incongruent events (prediction errors) on a variety of time scales. P3b has long been established as an index of effortful, evaluative processing of particularly informative events (Dien et al., 2004; Goldstein et al., 2002; Huang et al., 2015; Katayama & Polich, 1998). Providing exploitable information was indeed the common denominator of CP, AC+, and PE, as all corresponding analyses throughout studies 2 and 3 compared informative events to non-informative counterparts.

Critically, P3b did neither depend on model-compliance (see above) nor the temporal scope on which the information was used: Modulations due to global contextual features (i.e., checkpoints vs standard trials, study 2) were equally reflected within the P3b time frame as short-term, local changes in contingencies (i.e., informative adjustment cues vs non-informative equivalents, study 3). In light of our previous proposal that model evaluation makes use of model-compliant information in uncertain environments (see study 2), it seems intuitive that iterative validation of the internal model is not limited to either local or global information. As is the case for error-induced model adaptation, the overall hierarchical organisation of predictive processing implies that model

evaluation too can occur at different levels of the processing hierarchy. Empirical studies have highlighted this multi-level organisation by demonstrating distinct error signals for the violation of local probabilities and global rules (Wacongne et al, 2011; cf. Bekinschtein et al., 2009). In other words, similar to the paradigm of our third study, the authors were able to distinguish processing of unexpected from expected mismatch signals. Therefore, representational similarity analysis from study 3 was particularly designed to provide a comprehensive account of how local and global information is weighted in the experimental context (see *5.5 Critical reflection* for considerations of potential influence factors).

As described earlier, RSA model comparison revealed a better fit of a model reflecting global statistical contingencies than the one modelling local changes. This finding corroborates a fundamental notion of hierarchical predictions, namely that global models always rank higher than local expectations in predictive processing. Once a global set of rules has been acquired as a frame of reference - in our studies, by means of statistical learning - local deviations from this global model are quite literally 'exceptions to the rule'. With respect to computational efficiency, it appears considerably less effortful to propagate occasional error signals upstream than to develop highly detailed model assumptions on principle. This marks a critical distinction regarding the time scale of model adaptation: While rapid, low-level adaptation occurs automatically when even a single unexpected stimulus is perceived, broader changes to model assumptions require integration of multiple mismatch occurrences over time.

Balancing computational costs and benefits of maintaining highly specified internal models vs committing prediction errors is particularly relevant in contexts where model validity is uncertain. The stepwise processing account we proposed in study 1 provides a highly efficient solution regarding model specificity in uncertain environments: As participants had completed a training session prior to the experiment, they were (albeit implicitly) aware of variably uncertain contexts. When

uncertainty was high, checkpoints emerged as reference points whose model-compliant information was exploited for model evaluation. In other words, rather than preparing a separate model for the changed contingencies in high uncertainty contexts, original full-length predictions were segmented into smaller portions and incrementally evaluated at checkpoint positions.

Underscoring the precedence of global over local information, study 3 showed successful multivariate decoding of checkpoint potentials under high vs low uncertainty. This finding is quite instructive in that local information from adjustment cues could have indeed cancelled out the functional significance of checkpoints altogether: Extensions were in fact more frequent under high (vs low) uncertainty, but were still reliably preceded by adjustment cues. Therefore, when no adjustment cue was presented[1], local probabilities of extensions were factually identical in high and low uncertainty contexts (see Fig. 4-1, panel B). Distinct behavioural and functional indices of adjustment cue information notwithstanding (see study 3), differential processing of high vs low uncertainty checkpoints thus demonstrates how global models are - to a certain extent - shielded against energetically costly refinement. Conceivably, the circumstances of eventual global changes to the internal model are subject to modulation by a variety of parameters, some of which will be addressed below.

[1] It is critical to recall that study 3 only sampled checkpoints from sequences which did not contain an adjustment cue at the third sequential position. This was the only way to a) ensure that original, global task contingencies were still in place and b) maximise applicability of transferring the concept of checkpoints established in the first two studies into the new paradigm of study 3.

5.5 Critical reflection

The three studies presented in this thesis were conceptualised to highlight select aspects of predictive processing with a focus the way in which model-compliant information may be exploited similarly to prediction errors. Despite careful planning, neither of the studies comes without its limitations – some only apparent in hindsight – and a few points deserve critical reflection.

One issue to consider is the interplay of surprise and uncertainty. As the two measures are naturally correlated, we set up a parametric surprise regressor in the GLM of study 1 to assess net effects of higher-order uncertainty on checkpoint processing. While study 2 employed an identical replication of the paradigm in every respect[2], the mass univariate approach we consistently applied throughout studies 2 and 3 did not allow for an equivalent analysis. One way to address the influence of stimulus-bound surprise on event-level P3b amplitudes is by means of trial-by-trial analyses (e.g., Bennett et al., 2015; Mars et al., 2008). As such additional analyses were beyond the scope of studies 2 and 3, it remains an important objective for future efforts to illuminate differential contributions of single-trial surprise and higher-level context uncertainty.

Relatedly, we would have liked to examine in more detail the influence of irreducible uncertainty on reported potentials evoked by adjustment cues and checkpoints in study 3. Due to constraints on the duration of the EEG measurement, we were unfortunately not able to include a sufficient amount of trials to allow these analyses. For example, it would be highly instructive to see whether distinct potentials of adjustment cues vs non-informative equivalents vary as a function of context statistics. Potential interactions of adjustment cue exploitation and contextual uncertainty would further underscore the functional similarities of adjustment cues and checkpoints: Conceivably, adjustment cue positions, like checkpoints, are of particular interest under high uncertainty to assess whether any

[2] Note that study 2 contained more experimental blocks (i.e., a higher overall trial count) than study 1 due to the longer duration of the EEG measurement (vs fMRI).

further information can be gained locally. In a similar vein, it would have been instructive to be able to include uncertainty as a factor in the global and local model DSM of the RSA analysis. This would allow us to model even more detailed relations between the events of interest, for example with regard to how functionally similar checkpoints or adjustment cues are under high vs low uncertainty.

Finally, RSA results from study 3 give reason to fruitful speculation regarding the distinction of local and global information. As outlined above, better model fit of the global model corroborated how hierarchical principles in predictive processing rank higher order assumptions over local ones. In other words, local information, in this case provided by on-line adjustment cues, does not overwrite learned global contingencies. However, it is worth considering the experimental circumstances under which these results were obtained and how they may vary in everyday life. As discussed earlier, local information prompting short-term updating is initially the exception to an established set of rules. This global frame of reference is somewhat robust to structural changes, as it would be highly effortful to maintain an unnecessarily complex model at any time. Only if expectation violations are encountered repeatedly over time does slower learning occur in the form of higher-order model adaptation.

Conceivably, however, such changes may be spurred by a number of motivational as well as cognitive parameters. For example, participants learned both local probabilities and global uncertainty changes implicitly. While consistent response patterns throughout all three studies demonstrated behavioural indices of successful learning, *explicit* instructions regarding adjustment cues would certainly increase the prominence of such informative events. Just as a priority road sign explicitly indicates short-term changes of global rules ('Yield to the right'), allocation of attentional resources towards explicitly instructed cues would facilitate their integration into model assumptions.

From a motivational point of view, it is worth noting that participants were exclusively driven by intrinsic motivation. In the absence of reward, one could argue that incentives to feel involved in task

performance were rather low. While the focus on non-reward prediction errors was a critical novel feature of all three studies, the question remains to what extent available information is used as a function of personal commitment – either because correct predictions prompt rewarded responses or because, as exemplified in study 1, avoidance of prediction errors prevents harmful consequences like dropping a fridge in the hall. Much like explicit instructions, personal investment (as is often the case in everyday life) may very well lead to increased attention towards informative events regardless of model-compliance and, therefore, to differential neural signatures.

5.6 Future directions

Some critical objectives for subsequent efforts have already been outlined in the discussion sections of all three studies. Especially since study 3 has provided first insight into the functional similarities of both model-congruent and –incongruent events, future studies should aim to extend our understanding of the parameters modulating the exploitation of model-compliant information.

Further research is needed for a more detailed characterisation of adjustment cues. While study 3 has demonstrated differential processing of informative vs non-informative cue positions, it would be revealing to assess the extent to which these positions are processed differently from non-informative sequential standards: Even if no cue is presented, the *potential* information gain at respective sequence positions may be sufficient to single out these events compared to deterministic standard trials. This kind of functional prominence, in turn, conceivably depends on how vital potential information from cue positions would be in the current task context. A similar logic lead to the discovery of checkpoints in study 1: Cast as sites of potential prediction errors, checkpoints were highly informative in uncertain environments which compromised confidence in model validity.

Consequently, the more instructive adjustment cues are, the more resources could will presumably be allocated to sequential positions from which such information could potentially be gained, irrespective of the actual outcome. To this end, future studies should assess neural correlates not only in response to, but also leading up to, these events. Upcoming efforts should assess expectations prior to potentially informative sequential positions by means of pre-stimulus ERPs or time frequency analyses. One promising candidate marker of information expectation is the stimulus-preceding negativity (SPN; see Mnatsakanian & Tarkka, 2002), a slow potential shown to increase in amplitude prior to the presentation of particularly informative events (Morís et al., 2013). It remains to be examined whether potential cue positions are preceded by correlates of increased expectations when compared to non-informative standard trials. In a similar vein, as distinct components of predictive processing have been suggested to be transmitted by separate oscillatory rhythms (e.g., Arnal & Giraud, 2012), assessment of pre-stimulus differences in the power of β- and γ-frequency bands would considerably add to the understanding of adjustment cue and checkpoint processing.

Finally, much like study 1 demonstrated the influence of context uncertainty on the exploitation of checkpoints, intriguing questions remain with regard to which factors supposedly modulate pre-stimulus expectation of potentially informative cue positions. As suggested above, one promising variation would be to decouple context uncertainty and adjustment cue validity in an effort to make cue information particularly vital under high uncertainty.

5.7 Conclusion

In this dissertation, I presented three studies designed to illuminate how both model-congruent and –incongruent information is used to maintain computational efficiency in predictive processing. A culmination of fMRI and EEG findings highlighted the benefits of checkpoints (CP) for stepwise processing in contexts where confidence in full-length predictions was compromised due to high irreducible uncertainty. Checkpoints were subsequently shown to share central functional characteristics with prediction errors (PE), in that both events provided vital, highly task-relevant information (compared to sequential standard trials). Distinct event-related potentials were shown for joint information provided by checkpoints and prediction errors (P3b) and for the mismatch signal evoked only by PE (N400). We concluded that checkpoint information is used for context-induced model evaluation whereas prediction errors instigate model adaptation.

Finally, on-line task-relevant information provided by adjustment cues (AC) was found to be exploited similarly to checkpoints, but on a different time scale: While checkpoints emerged as a function of global changes in uncertainty, adjustment cues provided local model-compliant information. Representational similarity analysis corroborated the primacy of global over local information in predictive processing, raising intriguing questions for future research regarding the variation of model-compliant event processing across time scales.

6. References

Alexander WH, Brown JW (2011). Medial prefrontal cortex as an action-outcome predictor. Nat Neurosci 10:141338.

Allen TA, Salz DM, McKenzie S, Fortin NJ (2014). Nonspatial sequence coding in CA1 neurons. J Neurosci 36:1547-1563.

Arnal LH, Giraud AL (2012). Cortical oscillations and sensory predictions. Trends Cogn Sci 16:390-398.

Bar M (2009). The proactive brain: memory for predictions. Philos Trans R Soc B Biol Sci 364:1235-1243.

Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Dale AM, Halgren E (2006). Top-down facilitation of visual recognition. Proc Natl Acad Sci 103:449-454.

Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ (2012). Canonical microcircuits for predictive coding. Neuron 76:695-711.

Bayer HM, Glimcher PW (2005). Midbrain Dopamine Neurons Encode a Quantitative Reward Prediction Error Signal. Neuron 47:129–141.

Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007). Learning the value of information in an uncertain world. Nat Neurosci 10:1214–1221.

Bekinschtein TA, Dehaene S, Rohaut B, Tadel F, Cohen L, Naccache L (2009). Neural signature of the conscious processing of auditory regularities. Proc Natl Acad Sci USA 106:1672-1677.

Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B 57:289-300.

Bennett D, Murawski C, Bode S (2015). Single-trial event-related potential correlates of belief updating. ENeuro 2.

Bestmann S, Harrison LM, Blankenburg F, Mars RB, Haggard P, Friston KJ, Rothwell JC (2008). Influence of uncertainty and surprise on human corticospinal excitability during preparation for action. Curr Biol 18:775-780.

Bland AR, Schaefer A (2012). Different varieties of uncertainty in human decision making. Front Neurosci 6:85.

Bonacich P (2007). Some unique properties of eigenvector centrality. Soc Netw 29:555-564.

Bornstein AM, Daw ND (2012). Dissociating hippocampal and striatal contributions to sequential prediction learning. Eur J Neurosci 35:1011–1023.

Botvinick MM (2012). Hierarchical reinforcement learning and decision making. Curr Opin Neurobiol 22:956-962.

Brunet D, Murray MM, Michel CM (2011). Spatiotemporal analysis of multichannel EEG: CARTOOL. Comput Intel Neurosci 2.

Bullmore ET, Suckling J, Overmeyer S, Rabe-Hesketh S, Taylor E, Brammer MJ (1999). Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. IEEE Trans Med Imag 18:32-42.

Cabeza R, Ciaramelli E, Olson IR, Moscovitch M (2008). The parietal cortex and episodic memory: an attentional account. Nat Rev Neurosci 9:613–625.

Carlson T, Tovar DA, Alink A, Kriegeskorte N (2013). Representational dynamics of object vision: the first 1000 ms. J Vis 13:1-19.

Chaumon M, Bishop DV, Busch NA (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. J Neurosci Methods 250:47-63.

Clark A (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci 36:181–204.

Clark A (2015). Embodied prediction. Open MIND. Frankfurt am Main: MIND Group.

Clower DM, West RA, Lynch JC, Strick PL (2001). The inferior parietal lobule is the target of output from the superior colliculus, hippocampus, and cerebellum. J Neurosci 21:6283-6291.

Corbetta M, Shulman GL (2002). Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci 3:201–215.

Davachi L, DuBrow S (2015). How the hippocampus preserves order: the role of prediction and context. Trends Cogn Sci 19:92-99.

De Berker A, Rutledge RB, Mathys C, Marshall L, Cross G, Dolan R, Bestmann S (2016). Computations of uncertainty mediate acute stress responses in humans. Nat Com 7:10996.

Delgado MR, Locke HM, Stenger VA, Fiez JA (2003). Dorsal striatum responses to reward and punishment: effects of valence and magnitude manipulations. Cogn Affect Behav Neurosci 3:27–38.

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

Den Ouden HEM, Friston KJ, Daw ND, McIntosh AR, Stephan KE (2009). A dual role for prediction error in associative learning. Cereb Cort 19:1175-1185.

Den Ouden HEM, Daunizeau J, Roiser J, Friston KJ, Stephan KE (2010). Striatal Prediction Error Modulates Cortical Coupling. J Neurosci 30:3210–3219.

Dien J, Spencer KM, Donchin E (2004). Parsing the late positive complex: mental chronometry and the ERP components that inhabit the neighborhood of the P300. Psychophys 41:665-678.

Donchin E, Coles MG (1988). Is the P300 component a manifestation of context updating?. Behav Brain Sci 11:357-374.

Egner T, Monti JM, Summerfield C (2010). Expectation and Surprise Determine Neural Population Responses in the Ventral Visual Stream. J Neurosci 30:16601–16608.

Federmeier KD, Laszlo S (2009). Time for meaning: Electrophysiology provides insights into the dynamics of representation and processing in semantic memory. Psych Learn Mot 51:1-44.

Feldman H, Friston K (2010). Attention, uncertainty, and free-energy. Front Hum Neurosci 4:215.

Fiebach CJ, Schubotz RI (2006). Dynamic anticipatory processing of hierarchical sequential events: a common role for Broca's area and ventral premotor cortex across domains? Cortex 42:499-502.

Fisher RA (1921). On the probable error of a coefficient of correlation deduced from a small sample. Metron 1:3-32.

Fogelson N, Wang X, Lewis JB, Kishiyama MM, Ding M, Knight RT (2009). Multimodal effects of local context on target detection: evidence from P3b. J Cog Neurosci, 21:1680-1692.

Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC (1995). Improved assessment of significant activation in function magnetic resonance imaging (fMRI). Use of a cluster-size threshold. Magn Res Med 33:636-647.

Fortin, NJ, Agster KL, Eichenbaum HB (2002). Critical role of the hippocampus in memory for sequences of events. Nat Neurosci 5:458-462.

Friederici AD, Kotz SA (2003). The brain basis of syntactic processes: functional imaging and lesion studies. Neuroimage 20:8-17.

Friedman D, Johnson Jr R (2000). Event-related potential (ERP) studies of memory encoding and retrieval: A selective review. Micr Res Tech 51:6-28.

Friston KJ (2003). Learning and inference in the brain. Neur Netw 16:1325-1352.

Friston KJ (2005). A theory of cortical responses. Philos Trans R Soc B Biol Sci 360:815-836.

Friston KJ (2009). The free-energy principle: a rough guide to the brain? Trends Cogn Sci 13:293-301.

Friston KJ (2010). The free-energy principle: a unified brain theory?. Nat Rev Neurosci 11:127.

Friston KJ (2011). What is optimal about motor control?. Neuron 72:488-498.

Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ (1993). Functional connectivity: the principal-component analysis of large (PET) data sets. J Cereb Blood Flow Metab 13:5-14.

Friston KJ, Daunizeau J, Kilner J, Kiebel SJ (2010). Action and behavior: a free-energy formulation. Biol Cybern 102:227-260.

Galfano G, Mazza V, Angrilli A, Umiltà C (2004). Electrophysiological correlates of stimulus-driven multiplication facts retrieval. Neuropsychologia 42:1370-1382.

Geng JJ, Vossel S (2013). Re-evaluating the role of TPJ in attentional control: contextual updating?. Neurosci Biobehav Rev 37:2608-2620.

Goldstein A, Spencer KM, Donchin E (2002). The influence of stimulus deviance and novelty on the P300 and novelty P3. Psychophys 39:781-790.

Gottlieb J (2007). From thought to action: the parietal cortex as a bridge between perception, action, and cognition. Neuron 53:9-16.

Grootswagers T, Wardle SG, Carlson TA (2017). Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time series neuroimaging data. J Cog Neurosci 29:677-697.

Groppe DM, Urbach TP, Kutas M (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. Psychophys, 48:1711-1725.

Harrison LM, Duggins A, Friston KJ (2006). Encoding uncertainty in the hippocampus. Neur Netw 19:535-546.

Haxby JV, Connolly AC, Guntupalli JS (2014). Decoding neural representational spaces using multivariate pattern analysis. Ann Rev Neurosci 37:435-456.

Hohwy J (2012). Attention and conscious perception in the hypothesis testing brain. Front Psych 3:96.

Hohwy J (2013). The predictive mind. Oxford University Press.

Hohwy J, Roepstorff A, Friston KJ (2008). Predictive coding explains binocular rivalry: An epistemological review. Cognition 108:687-701.

Holroyd CB, Coles MGH (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psych Rev 4:679-709.

Huang Y, Rao RP (2011). Predictive coding. Wiley Int Rev Cogn Sci 2:580-593.

Huang WJ, Chen WW, Zhang X (2015). The neurophysiology of P 300-an integrated review. Eur Rev Med Pharmacol Sci, 19: 1480-1488.

Huettel SA, Mack PB, McCarthy G (2002). Perceiving patterns in random series: dynamic processing of sequence in prefrontal cortex. Nat Neurosci 5:485–490.

Humphrey DG, Kramer AF (1994). Toward a psychophysiological assessment of dynamic changes in mental workload. Hum Factors, 36: 3-26.

Ishihara S (1917). Test for Colour-Blindness. Tokyo: Hongo Harukicho.

Isik L, Meyers EM, Leibo JZ, Poggio T (2013). The dynamics of invariant object recognition in the human visual system. J Neurophys 111:91-102.

Jones DS (1979). Elementary information theory. Oxford: Oxford University Press.

Katayama JI, Polich J (1998). Stimulus context determines P3a and P3b. Psychophysiology 5:23-33.

Kimura M, Widmann A, Schröger E (2010). Top-down attention affects sequential regularity representation in the human visual system. Int J Psychophys 77:126-134.

Kincade JM, Abrams RA, Astafiev SV, Shulman GL, Corbetta M (2005). An event-related functional magnetic resonance imaging study of voluntary and stimulus-driven orienting of attention. J Neurosci 25:4593-4604.

Kluger DS, Schubotz RI (2017). Strategic adaptation to non-reward prediction error qualities and irreducible uncertainty in fMRI. Cortex 97:32-48.

Kluger DS, Quante L, Kohler A, Schubotz RI (submitted). Being right matters: Model-compliant events in predictive processing. PLoS One.

Knight RT (1996). Contribution of human hippocampal region to novelty detection. Nature 383:256.

Knight RT, Scabini D, Woods DL, Clayworth CC (1989). Contributions of temporal-parietal junction to the human auditory P3. Brain Res, 502:109-116.

Kok A (2001). On the utility of P3 amplitude as a measure of processing capacity. Psychophysiology 38:557-577.

Kriegeskorte N, Mur M, Bandettini PA (2008). Representational similarity analysis-connecting the branches of systems neuroscience. Front Sys Neurosci 2:4.

Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009). Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neur 12:535.

Kriegeskorte N, Kievit RA (2013). Representational geometry: integrating cognition, computation, and the brain. Trends Cogn Sci 17:401-412.

Kühn AB, Schubotz RI (2012). Temporally remote destabilization of prediction after rare breaches of expectancy. Hum Brain Map 33:1812-1820.

Kumaran D, Maguire EA (2006). An unexpected sequence of events: mismatch detection in the human hippocampus. PLoS Biol 4:e424.

Kutas M (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. Lang Cogn Proc 8:533-572.

Kutas M, Federmeier KD (2000). Electrophysiology reveals semantic memory use in language comprehension. Trends Cogn Sci 4:463-470.

Kutas M, Federmeier KD (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Ann Rev Psych 62:621-647.

Kutas M, Hillyard SA (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. Science 207:203-205.

Lee TS, Mumford D (2003). Hierarchical Bayesian inference in the visual cortex. JOSA A 20:1434-1448.

Lehmann D, Ozaki H, Pal I (1987). EEG alpha map series: brain micro-states by space-oriented adaptive segmentation. Electroencephalogr Clin Neurophysiol 67:271-288.

Lisman JE (1999). Relating hippocampal circuitry to function: recall of memory sequences by reciprocal dentate-CA3 interactions. Neuron 22:233–242.

Lisman JE, Redish AD (2009). Prediction, sequences and the hippocampus. Philos Trans R Soc London B 364:1193-1201.

Little DYJ, Sommer FT (2013). Maximal mutual information, not minimal entropy, for escaping the "Dark Room". Behav Brain Sci 36:220-221.

Lohmann G, Müller K, Bosch V, Mentzel H, Hessler S, Chen L, Zysset S, von Cramon DY (2001). Lipsia—a new software system for the evaluation of functional magnetic resonance images of the human brain. Comput Med Imaging Graph 25:449-457.

Lohmann G, Margulies DS, Horstmann A, Pleger B, Lepsien J, Goldhahn D, Schloegl H, Stumvoll M, Villringer A, Turner R (2010). Eigenvector Centrality Mapping for Analyzing Connectivity Patterns in fMRI Data of the Human Brain. PLoS One 5:e10232.

Maess B, Koelsch S, Gunter TC, Friederici AD (2001). Musical syntax is processed in Broca's area: an MEG study. Nat Neurosci 4:540-545.

Makris N, Kennedy DN, McInerney S, Sorensen AG, Wang R, Caviness JVS, Pandya DN (2005). Segmentation of subcomponents within the superior longitudinal fascicle in humans: a quantitative, in vivo, DR-MRI study. Cereb Cortex 15:854–869.

Makris N, Papadimitriou GM, Sorg S, Kennedy DN, Caviness JVS, Pandya DN (2007). The occipitofrontal fascicle in humans: a quantitative, in vivo, DT-MRI study. Neuroimage 37:1100-1111.

Marr D (1982). Vision: A computational approach. Freeman.

Mars RB, Debener S, Gladwin TE, Harrison LM, Haggard P, Rothwell JC, Bestmann S (2008). Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. J Neurosci 28:12539-12545.

Mengotti P, Dombert PL, Fink GR, Vossel S (2017). Disruption of the right temporoparietal junction impairs probabilistic belief updating. J Neurosci 37:5419-5428.

Mnatsakanian EV, Tarkka IM (2002). Task-specific expectation is revealed in scalp-recorded slow potentials. Brain Topogr 15:87-94.

Moradi P, Shiri ME, Rad AA, Khadivi A, Hasler M (2012). Automatic skill acquisition in reinforcement learning using graph centrality measures. Intell Data Analys 16:113-135.

Morís J, Luque D, Rodríguez-Fornells A (2013). Learning-induced modulations of the stimulus-preceding negativity. Psychophys 50:931-939.

Mumford D (1992). On the computational architecture of the neocortex. Biol Cybern 66:241-251.

Murray MM, Brunet D, Michel CM (2008). Topographic ERP analyses: a step-by-step tutorial review. Brain Topogr 20:249-264.

Neal RM, Hinton GE (1998). A view of the EM algorithm that justifies incremental, sparse, and other variants. In: Learning in graphical models (pp. 355-368). Springer, Dordrecht.

Nelson SM, Cohen AL, Power JD, Wig GS, Miezin FM, Wheeler ME, Velanova K, Donaldson DI, Phillips JS, Schlaggar BL, Petersen SE (2010). A parcellation scheme for human left lateral parietal cortex. Neuron 67:156–170.

Niedeggen M, Rösler F, Jost K (1999). Processing of incongruous mental calculation problems: Evidence for an arithmetic N400 effect. Psychophys, 36:307-324.

Nieuwenhuis S, Aston-Jones G, Cohen JD (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. Psychol Bull 131:510.

Nieuwenhuis S, De Geus EJ, Aston-Jones G (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. Psychophys, 48:162-175.

Norman DA (1981). Categorization of action slips. Psych Rev 88:1–15.

O'Connor AR, Han S, Dobbins IG (2010). The inferior parietal lobule and recognition memory: expectancy violation or successful retrieval? J Neurosci 30:2924–2934.

Oosterhof NN, Connolly AC, Haxby JV (2016). CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. Front Neuroinf 10:27.

O'Reilly JX, Schüffelgen U, Cuell SF, Behrens TEJ, Mars RB, Rushworth MFS (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. P Natl Acad Sci USA 38:3660–3669.

Patel SH, Azzam PN (2005). Characterization of N200 and P300: selected studies of the event-related potential. Int J Med Sci 2:147.

Payzan-LeNestour E, Bossaerts P (2011). Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. PLoS Comp Biol 7:e1001048.

Petrides M, Pandya DN (1999). Dorsolateral prefrontal cortex: comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. Eur J Neurosci 11:1011-1036.

Petrides M (2000). Mapping prefrontal cortical systems for the control of cognition. In: Toga AW, Mazziotta JC, editors. Brain mapping: The systems (pp. 159-176). San Diego: Academic Press.

Polich J (2003). Theoretical overview of P3a and P3b. In: Detection of change. Boston: Springer.

Polich J (2007). Updating P300: an integrative theory of P3a and P3b. Clin Neurophysiol 118: 2128-2148.

Polich J, Criado JR (2006). Neuropsychology and neuropharmacology of P3a and P3b. Int J Psychophysiol 60:172-185.

Rabovsky M, Hansen SS, McClelland JL (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. Nat Hum Behav 2:693.

Rao RPN, Ballard DH (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79–87.

Raichle ME, Snyder AZ (2007). A default mode of brain function: a brief history of an evolving idea. Neuroimage 37:1083-1090.

Reason JT (1992). Cognitive underspecification: Its varieties and consequences. In: Baars BJ, editor. Experimental slips and human error: Exploring the architecture of volition (pp. 71-91). New York: Plenum.

Rescorla RA, Wagner AR (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical conditioning II: Current research and theory 2:64-99.

Rissman, J, Gazzaley A, D'Esposito M (2004). Measuring functional connectivity during distinct stages of a cognitive task. Neuroimage 23:752-763.

Rushworth MFS, Ellison A, Walsh V (2001). Complementary localization and lateralization of orienting and motor attention. Nat Neurosci 4:656-661.

Rolls ET (2013). The mechanisms for pattern completion and pattern separation in the hippocampus. Front Syst Neurosci 7:74.

Ross RS, Brown TI, Stern CE (2009). The retrieval of learned sequences engages the hippocampus: Evidence from fMRI. Hippocampus 19:790-799.

Schapiro AC, Gregory E, Landau B, McCloskey M, Turk-Browne NB (2014). The Necessity of the Medial Temporal Lobe for Statistical Learning. J Cogn Neurosci 26:1736–1747.

Schendan HE, Searl MM, Melrose RJ, Stern CE (2003). An fMRI Study of the Role of the Medial Temporal Lobe in Implicit and Explicit Sequence Learning. Neuron 37:1013–1025.

Schiffer AM, Ahlheim C, Wurm MF, Schubotz RI (2012). Surprised at all the entropy: hippocampal, caudate and midbrain contributions to learning from prediction errors. PloS One 7:e36445.

Scholl J, Kolling N, Nelissen N, Stagg CJ, Harmer CJ, Rushworth MFS (2017). Excitation and inhibition in anterior cingulate predict use of past experiences. eLIFE 6:e20365.

Schubotz RI, Korb FM, Schiffer AM, Stadler W, von Cramon DY (2012). The fraction of an action is more than a movement: Neural signatures of event segmentation in fMRI. Neuroimage 61:1195–1205.

Schuck NW, Gaschler R, Wenke D, Heinzle J, Frensch PA, Haynes JD, Reverberi C (2015). Medial prefrontal cortex predicts internally driven strategy shifts. Neuron 86:331-340.

Schultz W (1998). Predictive reward signal of dopamine neurons. J Neurophysiol 80:1–27.

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

Seghier M (2013). The Angular Gyrus: Multiple Functions and Multiple Subdivisions. Neuroscientist 19:43–61.

Silvetti M, Seurinck R, Verguts T (2013). Value and prediction error estimation account for volatility effects in ACC: a model-based fMRI study. Cortex 49:1627–1635.

Şimşek Ö, Barto AG (2009). Skill characterization based on betweenness. In: Advances in neural information processing systems (pp. 1497-1504). Boston: MIT Press.

Smith SM, Nichols TE (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44:83-98.

Snodgrass JG, Corwin J (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. J Exp Psych Gen 117:34.

Solway A, Diuk C, Córdova N, Yee D, Barto AG, Niv Y, Botvinick MM (2014). Optimal behavioral hierarchy. PLOS Comput Biol 10.

Strange BA, Duggins A, Penny W, Dolan RJ, Friston KJ (2005). Information theory, novelty and hippocampal responses: unpredicted or unpredictable? Neural Netw 18:225–230.

Summerfield C, Egner T (2009). Expectation (and attention) in visual cognition. Trends Cogn Sci 13:403-409.

Squire L, Kandel E (1999). Memory: From molecules to mind. Scientific American Library, New York.

Szewczyk JM, Schriefers H (2018). The N400 as an index of lexical preactivation and its implications for prediction in language comprehension. Lang Cog Neurosci 33:665-686.

Szűcs D, Csépe V (2005). The effect of numerical distance and stimulus probability on ERP components elicited by numerical incongruencies in mental addition. Cogn Brain Res 22:289-300.

Taylor PCJ, Muggleton NG, Kalla R, Walsh V, Eimer M (2011). TMS of the right angular gyrus modulates priming of pop-out in visual search: combined TMS-ERP evidence. J Neurophysiol 106:3001–3009.

Thirion B, Pedregosa F, Eickenberg M, Varoquaux G (2015). Correlations of correlations are not reliable statistics: implications for multivariate pattern analysis. In: ICML Workshop on Statistics, Machine Learning and Neuroscience (Stamlins 2015).

Tobia MJ, Iacovella V, Davis B, Hasson U (2012). Neural systems mediating recognition of changes in statistical regularities. Neuroimage 63:1730–1742.

Todorovic A, van Ede F, Maris E, de Lange FP (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. J Neurosci 31:9118-9123.

Trempler I, Schiffer AM, El-Sourani N, Ahlheim C, Fink GR, Schubotz RI (2017). Frontostriatal contribution to the interplay of flexibility and stability in serial prediction. J Cogn Neurosci 29:298-309.

Tribus M (1961). Information theory as the basis for thermostatics and thermodynamics. J Appl Mech 28:1-8.

Turk-Browne NB, Scholl BJ, Johnson MK, Chun MM (2010). Implicit perceptual anticipation triggered by statistical learning. J Neurosci 30:11177–11187.

Van Petten C (1993). A comparison of lexical and sentence-level context effects in event-related potentials. Lang Cogn Proc 8:485-531.

Verleger R, Heide W, Butt C, Kömpf D (1994). Reduction of P3b in patients with temporo-parietal lesions. Cogn Brain Res 2:103-116.

Verleger R, Jaśkowski P, Wascher E (2005). Evidence for an integrative role of P3b in linking reaction to perception. J Psychophys 19:165-181.

Verleger R, Grauhan N, Śmigasiewicz 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.

Vincent JL, Snyder AZ, Fox MD, Shannon BJ, Andrews JR, Raichle ME, Buckner RL (2006). Coherent spontaneous activity identifies a hippocampal-parietal memory network. J Neurophysiol 96:3517–3531.

Volpe U, Mucci A, Bucci P, Merlotti E, Galderisi S, Maj M (2007). The cortical generators of P3a and P3b: a LORETA study. Brain Res Bull 73:220-230.

Von Helmholtz H (1860). Handbuch der physiologischen Optik (Vol. 9). Voss.

Wacongne C, Labyt E, van Wassenhove V, Bekinschtein T, Naccache L, Dehaene S (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. Proc Natl Acad Sci USA 108:20754-20759.

Waskom ML, Kumaran D, Gordon AM, Rissman J, Wagner AD (2014). Frontoparietal Representations of Task Context Support the Flexible Control of Goal-Directed Cognition. J Neurosci 34:10743–10755.

Wurm MF, Schubotz RI (2012). Squeezing lemons in the bathroom: Contextual information modulates action recognition. Neuroimage 59:1551-1559.

Yamaguchi S, Knight RT (1992). Effects of temporal-parietal lesions on the somatosensory P3 to lower limb stimulation. Clin Neurophys 84:139-148.

Yu AJ, Dayan P (2005). Uncertainty, neuromodulation, and attention. Neuron 46:681-692.

7. Curriculum vitae

Name

Daniel Simon Kluger

Born

07.08.1989 in Oberhausen

Education

2019 | Ph.D., Cognitive Neuroscience, University of Münster
Title: *Informational segmentation in event prediction: Temporal dynamics and predictive efficiency*
Dissertation Advisors: Prof. Ricarda I. Schubotz, Prof. Pienie Zwitserlood, Moritz F. Wurm, Ph.D.

2014 | M.Sc., Psychology, University of Münster
Thesis: *Mapping multilevel object dimensions in the human premotor cortex*
Thesis Advisor: Prof. Ricarda I. Schubotz

2012 | B.Sc., Psychology, University of Münster
Thesis: *Preference and judgement in music evaluation*
Thesis Advisor: Prof. Günther Kebeck

Employment

2014 – 2019 | Researcher, Predictive Cognition and Action Lab (Prof. Ricarda I. Schubotz), Department of Psychology, University of Münster
2013 – 2014 | Student research assistant, Predictive Cognition and Action Lab (Prof. Ricarda I. Schubotz), Department of Psychology, University of Münster
Teaching assistant 'Introductory Research Class' (B.Sc., Psychology)
2012 – 2013 | Teaching assistant 'Diagnostic Interviewing' (B.Sc., Psychology)
2011 – 2013 | Student research assistant, Institute for Biomagnetism and Biosignalanalysis, Münster
2010 – 2011 | Student research assistant, LVR Children's Psychiatric Clinic, Essen

Teaching experience

2017 | Introductory Research Class (B.Sc., Psychology)

2016 | Methodology Course: fMRI and TMS (M.Sc., Psychology)

2016 | Introductory Research Class (B.Sc., Psychology)

2015 | Methodology Course: fMRI and TMS (M.Sc., Psychology)

2015 | Biological Psychology Seminar: 'The mental cinema' (B.Sc., Psychology)

2014 | Introductory Research Class (B.Sc., Psychology)

Publications

Kluger DS, Quante L, Kohler A, Schubotz RI (in press). Being right matters: Model-compliant events in predictive processing. PLoS One.

Quante L, Kluger DS, Ekman M, Bürkner PC, Schubotz RI (2018). Graph measures in task-based fMRI: Functional integration during read-out of visual and auditory information. PLoS One 13:e0207119.

Kluger DS, Schubotz RI (2017). Strategic adaptation to non-reward prediction error qualities and irreducible uncertainty in fMRI. Cortex 97:32-48.

Presentations

Kluger DS, Schubotz RI (2017). Strategic adaptation to non-reward predicition error qualities and contextual volatility in fMRI. Poster presentation, Annual Meeting of the Cognitive Neuroscience Society (CNS). San Francisco, USA.

Quante L, Kluger DS, Schubotz RI (2017). Integration and segregation of task-specific areas during task preparation - A graph theoretical analysis. Poster presentation, Annual Meeting of the Cognitive Neuroscience Society (CNS). San Francisco, USA.

Language skills

German (native), English (fully fluent)