Uncovering contextual biases in human decision-making

A multivariate analysis approach for patterns of functional magnetic resonance imaging data and event-related potentials

Stefan Bode

aus Celle

Habilitationsschrift zur Erlangung der *venia legendi* für das Fach Psychologie an

der Humanwissenschaftlichen Fakultät der

Universität zu Köln

Melbourne, Australia, 26. April 2016

Foreword and Acknowledgements

This work is the result of a long scientific journey, and a rather quick decision to write it all up as my 'Habilitation' thesis. Perhaps, however, it was the other way round, and whilst my scientific journey has not been very long, the decision to write this thesis was the product of a longer process that zigzagged through my unconscious brain, and only surfaced in my consciousness during a conversation with Prof Jutta Stahl on New Years Eve 2013. It was another two years until I actually took some time off, locked myself away in a small hotel room with amazing ocean views of Cape Schanck on the beautiful Mornington Peninsula, and finally finished this project. Through writing this thesis I realised how I have developed a far more nuanced understanding of the scientific framework in which my recent work is embedded than I had two years ago, so I guess the long journey was worth it.

Neuroscience Lab in Melbourne for their various contributions to my research over the past years. My sincere thanks also extent to all members of the former Attention & Awareness research group at the MPI CBS, as well as to all other researchers, students and personnel at the MPI, who have contributed to my work during this time.

Notably, three colleagues have been particularly important for my academic development since I started working at the UoM in 2010, and I want to express my sincere gratitude to those people here. Firstly, Prof Philip L. Smith's support was absolutely indispensable for establishing a research career in Melbourne, and I will always be grateful for his trust in me. Secondly, I was very lucky to meet Dr Carsten Murawski in Melbourne with whom I founded the Decision Neuroscience Lab in 2011, which made it possible to pursue our joint research agenda over the past years to investigate the neural mechanisms underlying contextual biases in decision-making. Finally, I like to sincerely thank my 'Habil-Mami' Prof Jutta Stahl, who has supported me throughout my academic career, offered me her resources at times when I didn't have access to any, and convinced me to persevere with research during times of doubt.

Last but not least, a heart-felt thanks to everyone else who contributed in any way to my work, wellbeing, health and sanity over the past years; in particular to Kathleen Charles-Walsh, who additionally went through the process of proofreading this work.

This work is dedicated, once more, to the greatest supporter I have ever had:

Käthe Bilgenroth (1923 – 2013)

Table of Content

1. Introduction

Decision-making is a fundamental aspect of human cognition and behaviour. Every day, we make a multitude of decisions, ranging from rather simple perceptual choices (e.g., about which objects we encounter in the world) to complex financial decisions (e.g., whether to spend money on a car today, or to save up for the future). The underlying cognitive and neural mechanisms appear to directly deploy external information, gathered by our senses, as well as internal information, such as preferences and beliefs. Ideally, this results in well-informed decisions and successful goal-directed behaviour. In reality, however, we are often faced with decision situations in which we do not have clear preferences, or access to all information. For example, we might be unsure about road signs while driving in heavy rain; we might not have clear preferences for all potential choice options when buying a new car; or we might be in doubt about whether or not to buy a car in the first place. Moreover, there seem to be a variety of rather arbitrary choice situations in which decisions appear random and inconsequential, for example when we have no idea behind which wardrobe door the rain jacket that we are searching for might be hidden, and we have to decide whether to try the left door or the right door first. It is in these situations, where external information is absent or not helpful and decision preferences are unstable, that contextual factors appear to have a stronger influence on decision-makers. For example, greater saliency might bias object choice (Itti & Koch, 2001), encountering other objects with similar attribute profiles might shift preferences (e.g., Trueblood, Brown, Heathcote, & Busemeyer, 2013), certain contextual cues might shift individuals to be more reward-seeking (Woelbert & Goebel, 2013), and previous decisions might come to inform ensuing choices (Akaishi, Umeda, Nagase, & Sakai, 2014). The presented work will highlight research supporting the hypothesis that not only can the representation of goals be unconsciously biased (Custers & Aarts, 2010), but contextual

information can also exert significant biases on a variety of decisions outside decision-makers' awareness.

The decision scenarios outlined above roughly fall into three categories: a) perceptual decision-making, b) financial decision making / preference formation, and c) voluntary decision-making. Of course, this is not an exhaustive characterisation of all types of human decision-making, but these categories constitute three key research areas in psychology and cognitive neuroscience over the past few decades. They will also constitute the three strands of research described in the present work. From each of these fields, studies will be presented that exemplify a *content-based cognitive neuroscience approach* to human decision-making research, building on multivariate analysis techniques for brain imaging data that utilise machine learning algorithms to directly predict the content of decision-outcomes and other decision-related variables from brain activity. These techniques found their way into mainstream cognitive neuroscience, predominately for the analysis of spatially distributed patterns of functional magnetic resonance imaging (fMRI) data, in the first decade of the 21st century. They provided novel tools for making use of the full patterns of brain activation, as measured using blood-oxygen-level-depended (BOLD) signals, for the prediction of the content of cognitive processes (Haynes & Rees, 2006; Norman, Polyn, Detre, & Haxby, 2006). Recently, similar approaches have also been applied to patterns of brain activity as measured using electroencephalography (EEG), and in particular to spatio-temporal patterns of the event-related potential (ERP; Bode, Sewell, Lilburn, Forte, Smith, & Stahl, 2012b; King & Dehaene, 2014). While the application of machine learning techniques to both neuroimaging methods is not entirely new, the discovery that cognitive processes might be 'encoded' in – and moreover, be predictable from – patterns of noninvasively measurable brain signals has opened up new avenues for an information-based (or, content-based) cognitive neuroscience approach (Heinzle et al., 2012; Kriegeskorte, Goebel, & Bandettini, 2006). These methods have since been used to study the representation of a variety of

6

cognitive functions, ranging from perceptual processes (e.g., Cox & Savoy, 2003; Haxby, Gobbini, Furey, Ishai, Schouten, & Pietrini, 2001; Haynes & Rees, 2005; Kay, Naselaris, Prenger, & Gallant, 2008; Kamitani & Tong, 2005; Williams, Dang, & Kanwisher, 2007), to more abstract cognitive processes, including the ones of interest for the present work: perceptual decisions (Bode, Bogler, Soon, & Haynes, 2012a; Bode, Bogler, & Haynes, 2013; Hebart, Schriever, Donner, & Haynes, 2016; Serences & Boynton, 2007), economic and financial decisions (e.g., Clithero, Carter, & Huettel, 2009; Murawski, Harris, Bode, Domínguez D, & Egan, 2012), preferences (Bode, Bennett, Stahl, & Murawski, 2014a; Tusche, Bode, & Haynes, 2010), and voluntary decisions (e.g., Bode, He, Soon, Trampel, Turner, & Haynes, 2011; Gilbert, 2011; Haynes, Sakai, Rees, Gibert, Frith, & Passingham, 2007; Soon, Brass, Heinze, & Haynes, 2008; Soon, He, Bode, & Haynes, 2013).

The present work will first provide a short methodological introduction into multivariate pattern analysis (MVPA) for both fMRI and ERPs in Chapter 2. This chapter will briefly explain the core concepts and principles by referring to two exemplary studies: **Bode and Haynes (2009)** for fMRI (Appendix I), and **Bode and Stahl (2014)** for ERPs (Appendix II). Next, Chapter 3 will discuss the application of MVPA for the investigation of perceptual decision-making and decision biases in perceptually ambiguous situations. In **Bode, Bogler, Soon, and Haynes (2012a)**, we showed that MVPA could be used for fMRI to predict decision outcomes for perceptual decisions as well as for perceptual guessing when participants had insufficient access to perceptual information (Appendix III). In **Bode, Bogler, and Haynes (2013)**, again using MVPA for fMRI, a direct demonstration of strong pattern similarity between perceptual guesses and voluntary decision-making was provided, suggesting a shift from utilising external information to internal information in perceptually ambiguous situations (Appendix IV). In **Bode, Sewell, Lilburn, Forte, Smith, and Stahl (2012b)**, we provided evidence from MVPA for ERPs that the internal information utilised in perceptual guessing was related to pre-existing information, which biased guessing decisions

on the following trial. We further showed, using mathematical modelling, that these biases originated from carry-over effects of participants' previous choices, and therefore that the choice history itself constituted a significant, unconscious contextual driver for decision-making (Appendix V). Chapter 4 discusses a study by **Murawski, Harris, Bode, Domínguez D, and Egan (2012)** showing that MVPA for fMRI could be used to predict choice outcomes for specific financial decisions, namely intertemporal decisions, in which participants made consecutive decisions between immediately available smaller rewards and larger rewards, which were only available in the future (Appendix VI). Furthermore, the study showed that strongly masked images of rewarding brand logos, used as prime images, could shift the neural representation of subjective values, highlighting the potential of contextual cues to exert unconscious biases on financial decisions. This chapter also discusses the study by **Bode, Bennett, Stahl, and Murawski (2014a)** that used MVPA for ERPs to investigate potential pathways for the unconscious influence of contextual cues on decision-making (Appendix VII). In particular, the study showed that some abstract stimulus features, such as a stimulus' reference to time, were processed fast and unconsciously, even without directing attention to these particular features. If such automatically processed information maps onto important aspects of a subsequent decision problem, it could bias decision-making outside of awareness. Chapter 5 will then introduce several studies on voluntary decision-making. The study by **Soon, He, Bode, and Haynes (2013)** built on earlier work, which showed that arbitrary voluntary decisions between left and right button presses could be predicted from brain activity several seconds before decision-makers' became aware of having made a decision. Soon and colleagues (2013) extended these findings using MVPA for fMRI to the prediction of more abstract voluntary decisions between two arithmetic operations (Appendix VIII). Using MVPA for ultra-high field 7 Tesla fMRI, **Bode, He, Soon, Trampel, Turner, and Haynes (2011)** consolidated evidence that arbitrary voluntary decisions between left and right button presses could be predicted several seconds before the reported conscious

decision, and additionally reported evidence for a slow build-up of decision-related patterns over time (Appendix IX). This chapter concludes with a discussion of how these findings can be explained by assuming the influence of hidden contextual effects – here participants' choice history as well as their implicit interpretation of the decision task instructions – on the outcome of voluntary decisions. This discussion will draw on ideas presented in **Bode, Murawski, Soon, Bode, Stahl, and Smith (2014b)**, which suggested that an evidence accumulation framework could account for the variety of decision situations as well as for contextual bias effects (Appendix X). Finally, Chapter 6 will provide a short conclusion of the work discussed here and remarks for future research.

2. Multivariate pattern analysis (MVPA) for brain activity

2.1 MVPA for functional magnetic resonance imaging (fMRI)

One major challenge for a cognitive neuroscience of decision-making is the gap between the neural level of description (e.g., differences in BOLD signal for fMRI) and the cognitive level (Poldrack, 2006; Poldrack & Farah, 2015). In order to successfully inform cognitive models it is often not sufficient to show the involvement of a brain region during decision-making, for example by contrasting the average activation in a given region during a decision task with activation in a control task. This is because any observed difference in activation could simply be due to additional effort, modulation by attention, or some other confounded cognitive process, and must not necessarily reflect the decision process itself. Importantly, this problem is not unique to fMRI, and no neuroimaging method to date can directly monitor cognitive processes without strongly relying on reverse inference (Poldrack, 2006). However, it has been argued that demonstrating a *representation of the content* of cognitive processes (for example, the *outcome* of a decision) would indeed constitute stronger (albeit arguably still not conclusive) evidence for a more direct involvement of the respective brain regions in the decision process (e.g., Haynes, 2015; Heinzle et al., 2012), and thus, linking its function closer to decision-making *per se*.

Multivariate pattern classification analyses (e.g., Haxby, et al., 2001; Haynes & Rees, 2006; Kriegeskorte, et al., 2006; Norman, et al., 2006; Tong & Pratte, 2012), often referred to as neural 'decoding' methods, provide a novel set of tools that have been used to predict the content of cognitive processes from patterns of spatially distributed brain activity. These methods have been increasingly applied to fMRI, which is based on the blood-oxygen-level-dependent (BOLD) signal (Logothetis & Wandell, 2004). In short, oxyhaemoglobin (Hb), which is needed for glucose metabolism during local neural activity, is diamagnetic, and

henceforth enhances the local BOLD signal. Because task-induced neural activity is accompanied by a local increase in – and oversupply of – oxygenated blood, the BOLD signal can be utilised as an (indirect) indicator of the 'involvement' (activation) of a particular brain region during the cognitive process, which is assumed to be critical for the experimental task. Classically, fMRI data is analysed using a mass-univariate general linear model (GLM) approach (Friston, Holmes, Worsley, Poline, Frith, & Frackowiak, 1995). Signals in each three-dimensional, cubic measurement unit, termed 'voxel' (usually 1-4 mm^3 in size), are separately fitted by a GLM, which convolves the heamodynamic response function (HRF) with all event onsets, which are expected to elicit a neural response in the experiment. It is then statistically analysed how well the model fits the data, and the model fit is interpreted as task-induced activity at this particular position in the brain, usually compared to a baseline or an adequate control task (Friston et al., 1995). In order to account for anatomical variability, and to satisfy the assumption of Gaussian random field theory, activity from adjacent voxels is pooled by spatial smoothing, preceding statistical group-level analysis. In publications, usually only the activation in larger, connected clusters of voxels is reported, while sub-threshold clusters and single, isolated supra-threshold voxels are treated as noise. This classical approach, however, neglects the possibility that the aggregation and smoothing of activation might result in a significant loss of information, which may have been contained in the full fine-grained local spatial activation patterns.

MVPA, on the other hand, can provide a more sensitive measure than conventional mass-univariate analyses, because even if the signal in single voxels might not be strongly associated with a cognitive process, larger *patterns of activity* might indeed show stable and reliable associations (Kamitani & Tong, 2005; Haynes & Rees, 2006). This could be the result of weak but nevertheless stable associations (or biases) of single voxels, related to the encoding of stimulus properties. For example, orientation preference has been suggested to be organised in cortical columns in early visual cortex, and thus neurons coding for single spatial

orientations would cluster together. If orientation columns can be assumed to be spatially distributed across parts of the visual cortex, and the voxel-grid is randomly applied to the cortical surface during scanning (as it usually is the case), this would result in weak, idiosyncratic biases for some orientations in single voxels (Kamitani & Tong, 2005; Haynes & Rees, 2006; Swisher, et al., 2010; for further discussions of coding principles see: Haynes, 2015; Op de Beeck, 2010; Kamitani, & Sawahata, 2010; Swisher, Gatenby, Gore, Wolfe, Moon, Kim, & Tong, 2010). A cluster of adjacent voxels, each with a weak but stable coding bias, would then result in meaningful patterns, containing information that can be picked up with MVPA. The assumption of weak but stable biases can be extended to other visual properties for higher-level representations of object categories in the ventral (and potentially the dorsal) visual pathway (e.g., Carlson, Schrater, & He, 2003; Cox & Savoy, 2003; Williams et al., 2007), and also to other modalities (e.g., Howard, Plailly, Grueschow, Haynes, & Gottfried, 2009). Whether or not similar clustering could explain decodable representations in other brain areas, such as prefrontal and parietal cortex, is still an open question; however, a variety of cognitive processes have been predicted from regions beyond sensory cortex, including stimulus saliency (Bogler, Bode, & Haynes, 2011), deception (Davatzikos et al., 2005), emotion regulation strategies (Morawetz, Bode, Baudewig, Jacobs, & Heekeren, 2016), decisions (Soon, Brass, Heinze, & Haynes, 2008; Soon, et al, 2013; Bode et al., 2011), intentions (Haynes et al., 2007; Gilbert, 2011), and decision rules (Bode & Haynes, 2009; Woolgar, Thompson, Bor, & Duncan, 2011). The increased sensitivity of MVPA has led some authors to suspect that sometimes, one might decode only confounds rather than the cognitive processes of interest (Todd, Nystrom, & Cohen, 2013). Firstly, however, simple confound controls can be incorporated into experimental designs and GLMs that protect from such spurious findings; and secondly, an increased sensitivity *per se* is always advantageous as it increases a method's applicability and should not be regarded as a limitation (for a detailed discussion see: Woolgar, Golland, & Bode, 2014).

The general approach taken in MVPA is to train a pattern classifier, usually adopted from machine learning (e.g., Chang & Lin, 2011) on several exemplars of neural activation patterns from a cluster of voxels (as opposed to each voxel separately), acquired when participants are engaged in an experimental task. In one of the first studies to use this approach for the investigation of higher-level cognitive processes, Bode and Haynes (2009) (Appendix I) asked participants to apply one of two task rules while being scanned using 3 Tesla MRI: the display of cue A instructed them to press a joystick to the right when presented with one specific colourful, moving stimulus, and to the left when presented with the alternative stimulus. Cue B instructed the reverse mapping. The presentation of task components (cue: 1400ms, delay: 2800ms, stimulus: 4200ms, response screen: 2800ms) was spaced in time, which allowed the temporally staggered prediction of task rules, presented stimuli, and responses during the entire process of task preparation (see *Materials and Methods* in Bode & Haynes, 2009). The experimental data was then divided into four separate sets, corresponding to the four experimental runs (comprising 64 trials each). After data pre-processing, which only involved the spatial realignment to the first recorded image (volume) of the first run to correct for head motion, and slice-timing correction of each volume to correct for differences in recording time of each slice, a GLM (Finite Impulse Response model, FIR; for details see *Materials and Methods* in Bode & Haynes, 2009) was fitted to the data, modelling the two rules, two target stimuli, and both motor responses separately for each task phase. Then, a linear support vector machine (SVM) classifier was trained on data from three of the four runs (i.e., the beta-images, representing the model-derived regressors for each event-of-interest in each run) for a series of three-dimensional 'searchlight' clusters (Kriegeskorte et al., 2006; Haynes et al., 2007) with a radius of r=4 voxels. Note that classifiers can be applied to brain activation patterns in different ways, including analysing the entire brain, or analysing specific regions of interests (ROIs). The 'searchlight' approach, however, as used in all studies described in this work, first requires the definition of a cluster,

usually a sphere with a radius of 3-5 voxels, created around an arbitrary voxel in the brain, whose activation pattern is then analysed (Figure 1A). In Bode and Haynes' (2009) study, this cluster was subsequently rebuilt around every voxel in the brain, and all possible positions were analysed separately and independently. One important advantage of this approach is that it is spatially unbiased and does not rely on any assumptions or potentially circular analyses (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009) to define or identify candidate brain regions for MVPA. For each 'searchlight' cluster, the classifier then estimated a decision boundary (or, hyperplane) to distinguish between patterns associated with each rule, stimulus and motor response (also see: Mur, Bandettini, & Kriegeskorte, 2009). In order to determine whether these activation patterns contained information about the experimental conditions, the trained classifier was then used to predict rules, stimuli and responses from patterns of independent data (i.e., the left-out run). For statistical testing, a leave-one-out cross-validation procedure was applied, and data from each run served as test data once while training was performed on the data from the other runs (Figure 1A). Finally, the resulting average classification (or 'decoding') accuracy from each 'searchlight' cluster was attributed to the central voxel of the respective cluster, and a three-dimensional map of classification accuracies was generated to represent information about each of the three task components in the brain. This procedure was repeated for all consecutive time steps (2800ms, corresponding to the TR) of the experiment, making it possible to map the temporal evolution of information regarding rules, stimuli and responses as they became available to the decision-maker during the task. The main finding of this study was that the location of the representation of abstract rules, which guided decisions for motor responses, shifted over time, depending on the task phase. Rules were first encoded in visual cortex during the presentation of the cue, but during the following memory delay period, their representation could be decoded from the intraparietal sulcus, a region in posterior parietal cortex that is known to be involved in task switching and cognitive control (Singh-Curry & Hussain, 2009). Subsequently, during active

decision-making, and during preparation and execution of the motor response, the task rules, together with a representation of the stimulus, were represented in ventrolateral prefrontal cortex (VLPFC). These VLPFC regions have also been suggested to be key regions for cognitive control and decision-making (Bunge, 2004). Methodologically, this study paved the way for the application of MVPA in the following fMRI studies discussed in this work.
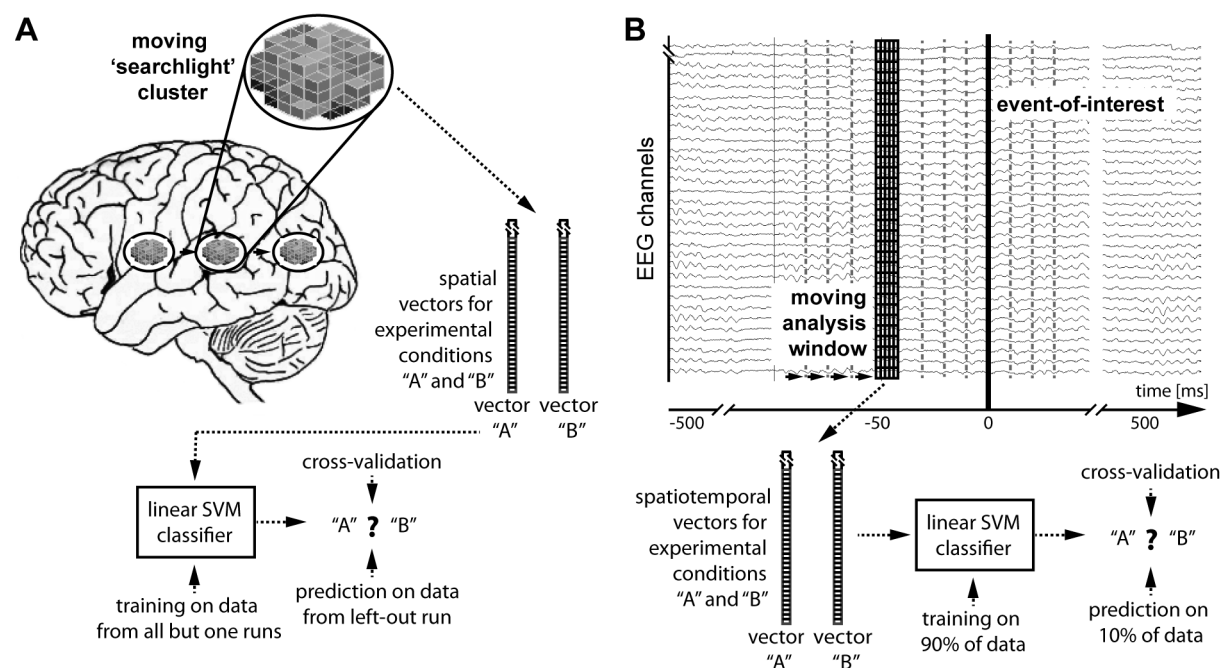
Figure 1: *Illustration of multivariate pattern classification analysis (MVPA) A) MVPA for fMRI. After fitting a GLM, data is extracted from moving 'searchlight' clusters with a pre-defined radius for the experimental conditions and transformed into vectors. These represent the local spatial activation patterns for each condition. A classifier, usually a linear support vector machine (SVM) classifier, is trained on data from all but one run and consecutively tested on data from the left-out run. A cross-validation procedure is applied, which means that classification is repeated until each run served as the test data set once. The average classification accuracy is assigned to the central voxel of the 'searchlight' cluster. This procedure is repeated for all possible 'searchlight' clusters in all participants (modified from Bode & Haynes, 2009). B) MVPA for ERPs. A moving analysis time window is defined for event-locked ERP data and applied to all possible time steps throughout the trial. For each analysis step, spatiotemporal patterns of activity are extracted and transformed into vectors. A classifier,*

usually a linear SVM classifier, is trained on 90% of all trials and consecutively tested on the left-out 10% of

trials. For cross-validation this is repeated until all sets of 10% of trials have served as the test data set once.

This procedure is usually repeated ten times to avoid drawing biases, and results are tested against a shuffled-

label test distribution for each analysis time window (modified from Bode & Stahl, 2014).

2.2 MVPA for event-related potentials (ERPs)

One major shortcoming of fMRI for the investigation of cognitive processes is its very limited

temporal resolution. Even though latest methodological improvements (for example the use of

multi-band sequences and ultra-high field MRI at 7 Tesla and above) can push the time it

takes to record a brain volume (*time to repeat*, TR) to one second or below, most standard

fMRI studies using 3 Tesla MRI scanners typically operate with a TR of 1.5 to 3s. The

attempts to track the evolution of predictive brain activation patterns therefore remain at a

rather coarse scale (Bode & Haynes, 2009; Bode et al., 2011; Soon et al., 2008; 2013). EEG,

on the other hand, offers a continuous measurement up to 4 kHz, meaning that one or more

data points are available per millisecond (500-1000 Hz are standard). Given that most

decision processes are fast, and their temporal scale is often below 100 ms (e.g., Smith &

Ratcliff, 2004), the EEG signal allows for more precise correlations with the processes of

interest. In particular, the event-related potential (ERP) method has been widely applied to

study cognition. For this, electrical activity recorded at the scalp is time-locked to

experimental events-of-interest (Luck, 2005). The general approach is to repeatedly measure

the cortical response to identical events, defined as voltage differences recorded at the scalp,

and to derive an estimate of the ERP by averaging across measurements. This has the

advantage that specific components of the ERP, which are observed at known scalp locations,

can be linked to manipulations in the experimental paradigm, which are thought to involve, or

modulate, different cognitive processes. As in fMRI, averaging serves the improvement of the

signal-to-noise ratio. The most commonly used parameters, which are believed to be linked to

relevant aspects of cognitive processes, are the polarity of the difference signal, the mean or peak amplitude, and the latency of specific components (for an introduction see Luck, 2005).

While classical ERP analyses have been (and still are) highly successful, they also face some serious limitations. First, at times it is difficult to clearly identify specific ERP components at the level of single participants. Second, there is a great deal of variance in the expression of these components. Third, for a particular research question, no specific ERP components might have been described in order to formulate meaningful *a priori* hypotheses. Finally, similar to mass-univariate analysis of fMRI data, there might be more information contained in the distributed patterns of ERPs then is extracted by analysing specific ERP components at circumscribed locations. In particular, the question of whether the *content* of a specific cognitive process is represented in brain activity at a given point in time is often difficult to answer using classical approaches. However, patterns of ERPs are also multivariate in nature as they consist of distinct time series of signal at all recorded sites (channels; usually between 32 and 128). This means, instead of dismissing signals from most channels as noise, it is possible to apply MVPA to the full pattern of activity with the aim to predict the content of the cognitive process at hand. This approach has been taken recently to investigate decision-making and error processing (Bode & Stahl, 2014; Bode et al., 2012; Charles, King, & Dehaene, 2014; Das, Giesbrecht, & Eckstein, 2010; related multivariate approaches can be found in: Philiastides & Sajda, 2006; van Vugt, Simen, Nystrom, Holmes, & Cohen, 2012). It has been suggested that MVPA also allows for investigating the temporal stability of mental representations (King & Dehaene, 2014).

The general MVPA approach for ERPs, as applied in all studies discussed in the present work, will be illustrated in the reminder of this chapter, using the example of the study by Bode and Stahl (2014) (Appendix II). In this study, participants were presented with a display containing a central digit between 1 and 9, which could be odd or even, flanked by

two different, identical digits, which could also be odd or even. Their task was to make simple parity judgements about the central digit while the EEG was recorded at 61 channels using active Ag/AgCl electrodes (actiCAP, Brain Products) at 500 Hz. Standard artefact removal was applied that involved screening for technical artefacts (±100 microvolt [µV]), the elimination of eye movements by applying an ocular correction algorithm (Gratton, Coles, & Donchin, 1983), and re-referencing using a current-source density (CSD) analysis (Perrin, Bertrand, & Pernier, 1987). The CSD analysis rendered the signal independent of the location of the reference electrode, acted as a spatial filter that decreased the blur distortion caused by skull resistance (Katznelson, 1981), and reduced the effect of adjacent currents on the local recordings by emphasizing shallow neural generators. Then, a spatio-temporal MVPA analysis was conducted using the Decision Decoding ToolBOX (DDTBOX; Bode, Bennett, Feuerriegel, & Alday, 2016). The aim of this analysis was to predict whether each parity decision was correct or incorrect *before* the motor response was made. This was based on the observation that all known ERP components, which reflect error processing, reach their peak only *after* the response is made. For example, the error-related negativity (Ne/ERN) (Falkenstein, Hohnsbein, Hoormann, & Blanke, 1991; Gehring, Goss, Coles, Meyer, & Donchin, 1993) peaks ~80–100 ms after an overt erroneous response, and the error positivity (Pe) (Falkenstein et al., 1991) peaks ~300 ms after the response error. However, the Ne/ERN component displays an early rise, starting before response execution (typically without differentiation between errors and correct responses), which suggests the existence of early generators of error-related signals. Bode and Stahl (2014) first identified the onset of the motor response at a cortical level ~100 ms preceding the response by means of analysing the response-locked lateralised readiness potential (R-LRP) (for details see *Material and Methods* in Bode & Stahl, 2014). They then used a sliding window approach to dissect each trial into small analysis time windows of 10 ms leading up to the overt response (containing 305 data points per analysis time window with a sampling rate of 500 Hz; 5 data points x 61 EEG

channels). Spatio-temporal patterns of ERPs were extracted from these time windows for when participants made errors and when they made correct responses. Data from all available trials (strictly controlled for the number of trials per condition) were then randomly sorted into ten sets; nine of which were used for training a linear SVM classifier, and the left-out set was used for testing. Using a ten-fold cross-validation procedure, which was additionally repeated ten times (resulting in 100 analyses), the average classification accuracy was determined for each analysis time window (see Figure 1B). Results were then statistically compared to classification results from an identical analysis, which used the same data but shuffled labels in each iteration of the cross-validation procedure, providing a strict empirical chance distribution (i.e. permutation test). Note that several variations of this approach have been applied in the studies described in this work: for example, the sets of trials can be de-noised first by averaging across trials within each set (similar to fMRI beta-estimates), and classification is subsequently performed on these average sets (Bode et al., 2012). It is also possible to classify based only on purely spatial patterns (by averaging across all temporal information within the analysis time windows), or only on temporal patterns (by conducting separate, independent analyses for each channel and treating the temporal information as a pattern) (Bode et al., 2012). Furthermore, instead of using SVM classification for two conditions, multivariate support vector regression (SVR) can be used to regress, for example, continuous data such as ratings (e.g., Bode et al., 2014). Other parameters that can be varied are the analysis window width and the step size (Figure 1B). To date, there is no conclusive guideline for which parameters are favourable, and their effectiveness might strongly depend on the nature of the cognitive processes of interest and their associated ERP profile.

The result of each classification step described above is an average classification accuracy, which can then be assigned to the respective analysis time window, resulting in an information time course for the entire the trial. Using this approach, Bode and Stahl (2014) showed that information about upcoming errors indeed started to be predictable at ~90 ms

before response execution. This information most likely originated from strong contributions of visual and motor channels, supporting the hypothesis that it was based on an ongoing analysis of the perceptual input, which increasingly accumulated evidence that the already initiated response was incorrect. Furthermore, decision errors were committed using less response force compared to correct responses. It could be speculated that this reduction in force reflected unsuccessful attempts to use the accumulated sensory information to prevent the erroneous response (Bode & Stahl, 2014). More importantly for the present work, these findings confirmed the high sensitivity of MVPA for ERPs to detect even subtle and early decision-related information in patterns of brain activity that would be overlooked if only conventional analyses were used.

3. Contextual biases in perceptual decision-making

Perceptual decision-making is a fast process, which usually happens without effort and deliberation. It involves comparing noisy sensory representations of objects to memory representations of choice options. Cognitive models of perceptual decisions usually share the general idea that noisy evidence is accumulated over time until a decision boundary (threshold, or criterion) is reached (Ratcliff, 1978; Smith & Ratcliff, 2004; Ratcliff, Smith, Brown, & McKoon, 2016). Once sufficient evidence has been accumulated for one choice alternative, a decision is triggered, and a response can be prepared. The dominant Diffusion Decision Model (DDM) conceptualises this evidence accumulation as a stochastic, noisy drift, starting at a neutral mid-position (in the standard model), and continuing until a decision threshold is reached. The evidence accumulation process is always noisy, due to internal noise (e.g., noise due to fluctuations in firing rates of neurons) and external noise (e.g., the number of photons reaching the retina). Perceptual decisions are relatively well understood psychologically as well as neurobiologically, and psychological models that were designed to explain behaviour have been successfully applied to models of activity of single neurons and populations of neurons (Heekeren, Marrett, & Ungerleider, 2008; Gold & Shadlen, 2007; Romo & Salinas, 2003; Smith & Ratcliff, 2004). These models, however, often fall short in explaining decision-making under *ambiguous perceptual situations,* and do not explicitly address decisions, which cannot be guided by sensory evidence. Fluctuations in neural noise have been discussed as a key factor in determining choice outcomes in these cases (Deco & Romo, 2008). For example, in an ambiguous face vs. vase detection experiment, incidental activation levels in a brain region in the ventral visual cortex, specialised for face processing, have been suggested to be associated with an advantage for face-decisions (Hesselmann, Kell, Eger, & Kleinschmidt, 2008). The two fMRI MVPA studies described in the following paragraphs (Bode, et al., 2012a, Appendix III; and Bode, et al., 2013, Appendix IV) were

designed to investigate whether perceptual decisions for objects in ambiguous perceptual situations could nevertheless be predicted from brain activation patterns in sensory, object-processing brain regions.

In the fMRI study by Bode and colleagues (2012a), participants were presented in each trial with one of two object stimuli (an image of a piano or a chair) or a pure-noise stimulus, which contained no meaningful object and consisted of scrambled phase textures with the same power spectra as the images from both object categories. Stimuli were presented in a randomised order in two possible discriminability conditions: either relatively easy to detect (~79-90% correct decisions) or near impossible to detect with performance around chance level (~34-45% correct decisions) due to differences in the masking procedure. Images were presented for 66.67 ms, followed by a post-mask for 433.33 ms for high discriminability, or for 16.67 ms followed by a post-mask for 483.33 ms for low discriminability (for details see *Materials and Methods* in Bode et al., 2012a). Importantly, participants attempted to make perceptual decisions in both conditions but needed to revert to guessing in the difficult condition. Participants indicated their choices using three response buttons, operated with the index-, middle and ring finger of the right hand. The response mapping was randomised on a trial-by-trial basis to de-correlate category choices and motor responses. The 'searchlight' variant of MVPA, as described above, was used to search for brain regions that allowed the prediction of choices under high discriminability and extremely low discriminability. It was found that in the high discriminability condition choices could be predicted from regions in ventral visual cortex, namely the lateral-occipital complex (LOC) and medial temporal cortex (Figure 2A). These regions have been associated with the encoding and processing of object information (e.g., Schwarzlose, Swisher, Dang, & Kanwisher, 2008). Hence, these findings support the assumption that under high discriminability, visual information was successfully sampled to guide subsequent decision-making. However, no decision-related information was found in the ventral object-related

brain regions for choices made under low discriminability. Instead, choices in the guessing situation could only be predicted from activity patterns in the medial posterior parietal cortex, a region located in the precuneus, and lateral parts of the posterior parietal lobe (Figure 2A). There are several potential interpretations for finding these representations. First, the precuneus is not known to be part of the visual system, but it has been discussed to be involved in mental imagery and self-related cognitive functions (Cavanna & Trimble, 2006), and this area also overlapped with posterior parietal brain regions that were found to be predictive of voluntary decisions (Soon et al., 2008). Furthermore, electrophysiology studies with monkeys have repeatedly shown that posterior parietal cortex, in particular the lateral intraparietal area (LIP), was related to processing the computation of a decision variable, indexing the integration of all available information into a decision signal (e.g., Shadlen & Newsome, 2001). This means, the findings by Bode and colleagues (2012a) could reflect either an attenuation of internal signals (as opposed to external sensory information) during decision-making, or some other properties of the decision-making process that might be enhanced, or only decodable, under perceptual ambiguity.

To further investigate the association between perceptual guessing and voluntary decision-making, Bode and colleagues (2013) replicated the previous findings using a close-to-identical fMRI paradigm with some minor modifications in the same 3 Tesla MRI scanner. Since the overlap in predictive patterns between the previous study and voluntary decision-making studies (Soon et al., 2008) could be incidental, for example because the medial parietal cortex is involved in multiple cognitive functions (Margulies et al., 2009), Bode and colleagues' (2013) study directly investigated *how similar* the neural activation patterns for perceptual guesses and voluntary decision were. For this, a different sample of participants made perceptual decisions between the same two object categories (pianos and chairs), again for highly discriminable objects as well as fully invisible objects, using the same masking procedure as described above. In order to rule out any effects of residual visual information

on decision-making, this time no object stimuli were actually shown in the low discriminability condition, but only a phase-randomised noise image was presented below perceptual threshold. This manipulation made it a true zero-discriminability (or, guessing) condition (for details see *Materials and Methods* in Bode et al., 2013). As a third condition, this experiment included a voluntary choice condition (indicated by the colour of the fixation cross shown at the beginning of each trial) using the same decision between pianos and chairs. In this condition, participants were instructed to freely and spontaneously choose one of the two categories, regardless of the visual presentation (which was identical to the guessing condition). The response mapping was again randomised on a trial-by-trial basis, thus de-correlating category choices and motor responses. 'Searchlight' MVPA was used to find brain regions, which allowed for the prediction of choices in each condition. The crucial analysis involved a *cross-condition classification* approach: As before, the classifier was trained on activation patterns associated with piano and chair choices from four out of five runs from the guessing condition. The trained classifier, however, was then used to predict category decisions from activation patterns from the respective left-out run from the voluntary decision condition. This analysis was then repeated the other way round, this time training on data from the voluntary decision condition and predicting test data from the guessing condition. The average cross-condition classification accuracy then denoted the *degree of inter-changeability* between activation patterns, which was used as an indicator for the similarity in the neural patterns between the two different types of decisions. It was shown that the posterior parietal regions found in the previous study (Bode et al., 2012a) indeed cross-predicted both types of decisions, while medial prefrontal cortex only predicted choices in the voluntary decision condition. However, patterns in medial prefrontal cortex did not generalise to guesses.
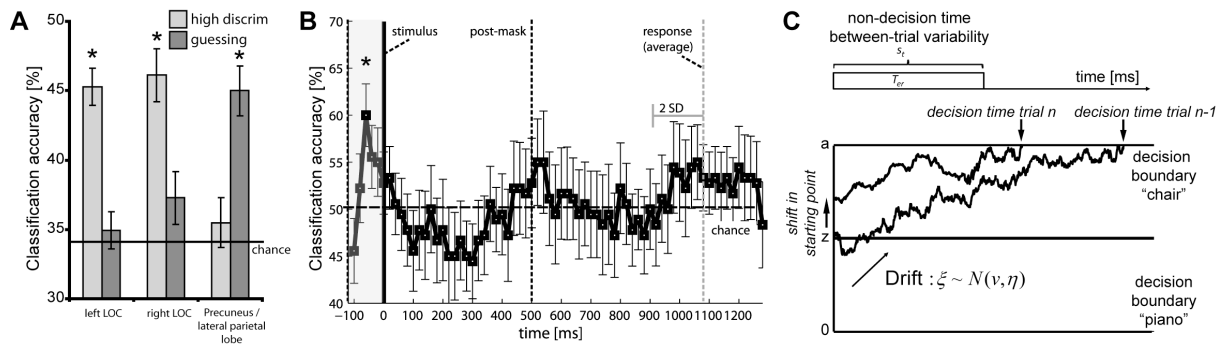
Figure 2: Prediction of perceptual guesses. A) Choice classification for highly discriminable objects and perceptual guessing. 'Searchlight' MVPA for fMRI revealed significant classification accuracies for high discriminability in bilateral LOC, and for guessing in medial and lateral posterior parietal cortex (modified from Bode et al., 2012a), B) Prediction of choice outcomes for guesses from patterns of pre-stimulus ERPs. Using a moving analysis window MVPA approach, significant classification accuracies were found in the 100ms preceding the stimulus presentation when the following object stimulus did not contain any object-related visual information (modified from Bode et al., 2012b). C) Illustration of the shift in starting point (as conceptualised in the Diffusion Decision Model) for evidence accumulation in a series of consecutive guessing trials in which stimuli did not contain any object-related visual information (modified from Bode et al., 2012b).

The findings of both fMRI studies (Bode et al., 2012a; Bode et al., 2013) strongly suggested an attenuation of internal (as opposed to external sensory) decision-related signals, which might drive choice under conditions of strong perceptual ambiguity. However, the questions regarding the origins of these choice-predictive signals, and how they might relate to the decision-making process, remained unanswered. One plausible explanation, which does not require the postulation of a fundamentally different symmetry-breaking decision mechanism, is that this activation reflected the state of the decision system *before* decision-making. However, both fMRI studies used a TR or 2.8 s and were therefore unable to provide the required temporal resolution to test this assumption. In order to overcome this limitation, Bode and colleagues (2012b) conducted a study that used a similar perceptual decision-making paradigm in combination with MVPA for ERPs (Appendix V) that allowed for the

prediction of decision outcomes on a millisecond time-scale before and after the decision was triggered. The study incorporated four discriminability conditions, again by varying the ratio of the presentation of an object stimulus (pianos and chairs at 16.67 ms, 33.33 ms, 50.00 ms, and 66.67 ms) and the post-mask (500 ms minus target object duration). Additionally, a true zero-discriminability condition was again included in which only a strongly masked noise stimulus was shown for 16.67 ms, as in Bode and colleagues' (2013) study. A randomised response-mapping screen was used as in the studies before, and participants indicated their category decisions using the left and right index finger. 63-channel EEG was recorded at a sampling rate of 500 Hz. After standard off-line pre-processing, a current source density (CSD) analysis was performed (for details see *Material and Methods* in Bode et al., 2012). For the MVPA analysis, signals time-locked to the pre-mask were averaged from piano trials and chair trials individually for each participant within each of the five runs and each experimental condition. A sliding analysis time window of 80 ms was moved in steps of 20 ms through the trial. Similar to the fMRI studies described above, a linear SVM classifier was trained on patterns of ERPs associated with piano choices and chair choices from all but one run, and tested on data from the left-out run. Notably, averaging was performed across all data points at each channel within each analysis time window (*spatial decoding*). A five-fold cross-validation procedure was applied, and statistical testing was performed against chance level as well as using a permutation test with randomised labels (see Chapter 2). In addition to this purely spatial decoding approach, a *temporal decoding* analysis was conducted separately for each of the 63 channels. For this, all 40 data points within each analysis time window were used as the relevant patterns on which classification was conducted. This analysis served as an independent method to localise channel-specific information for significant time windows from the spatial decoding analysis.

The results of this study confirmed that spatial activity patterns from 200 ms after the presentation of fairly discriminable object stimuli predicted the outcome of perceptual

decisions. Classification accuracies were strongest for the highest discriminability condition during the entire post-mask period and into the response period, but accuracies dropped with decreasing discriminability. The temporal decoding analyses revealed that during early stages of the trial information was as mainly represented in occipital and occipito-parietal channels, while during later stages, around response initiation, information was primarily present in prefrontal channels (see *Fig. 5* in Bode et al., 2012b). This study further replicated the prediction of decision outcomes in the zero-discriminability / guessing condition. Importantly, however, it was found that only the 100 ms time period *preceding* the stimulus presentation contained information about the upcoming choices. The temporal decoding analysis showed that for this pre-stimulus time period, prefrontal, posterior parietal and occipito-parietal channels were predictive of choice outcomes (Figure 2B). These findings confirmed the hypothesis that under perceptual ambiguity, the decision-related information predicted from patterns of brain activity might indeed reflect the state of the decision system immediately before the decision situation. This is particularly plausible given that the following stimulus did not contain any additional perceptual information, which could have over-written this pre-existing state to then drive the decision process. To further investigate the origin of this pre-stimulus information, Bode and colleagues (2012b) fitted the DDM (Ratcliff, 1978; Ratcliff et al., 2016; Smith & Ratcliff, 2004) to their behavioural data and obtained an excellent fit. They then allowed various model parameters to vary. It was found that the best fitting model showed substantial systematic starting point variability, meaning that the starting point for the evidence accumulation process in each trial was not neutral but systematically shifted towards the decision boundary of the previous choice (see Figure 2C). An analysis of choice ratios for zero-discriminability trials further confirmed these results: on consecutive pure-noise trials, participants were more likely to repeat their previous choices, and they were faster when repeating choices, most likely due to the shift in starting point towards the previous decision boundary, as predicted by the model.

In conclusion, the internal decision-related signals that were predictive for choice outcomes in ambiguous perceptual situations in all three studies (Bode et al., 2012a; 2012b; 2013) were most likely related to carry-over effects of decision biases, reflected in a systematic shift of starting point for evidence accumulation. Notably, these biases were dependent on participants' own choice history. This finding has several important implications. First, it shows that even in the absence of usable perceptual information, the decision system will never be in a true stalemate situation, because each decision will influence the following decisions to some degree. Second, contextual biases on perceptual decision-making exist (for explicit manipulations of biases see, e.g.: Chen, Jimura, White, Maddox, & Poldrack, 2015; Mulder, Wagenmakers, Ratcliff, Boekel, & Forstmann, 2012; van Ravenzwaaij, Mulder, Tuerlinckx, & Wagenmakers, 2012), and can even result from the decision-makers' own previous choices (see also Fründ, Wichmann, & Macke, 2014; for other effects of previous choice see also: Bronfman, Brezis, Moran, Tsetsos, Donner, & Usher, 2015; Hawkins, Brown, Steyvers, & Wagenmakers, 2012). Third, MVPA appears to have sufficient sensitivity to detect these biases, and in combination with an appropriate analysis of behavioural data (here computational modelling), it has the potential to shed light on otherwise hidden cognitive decision processes.

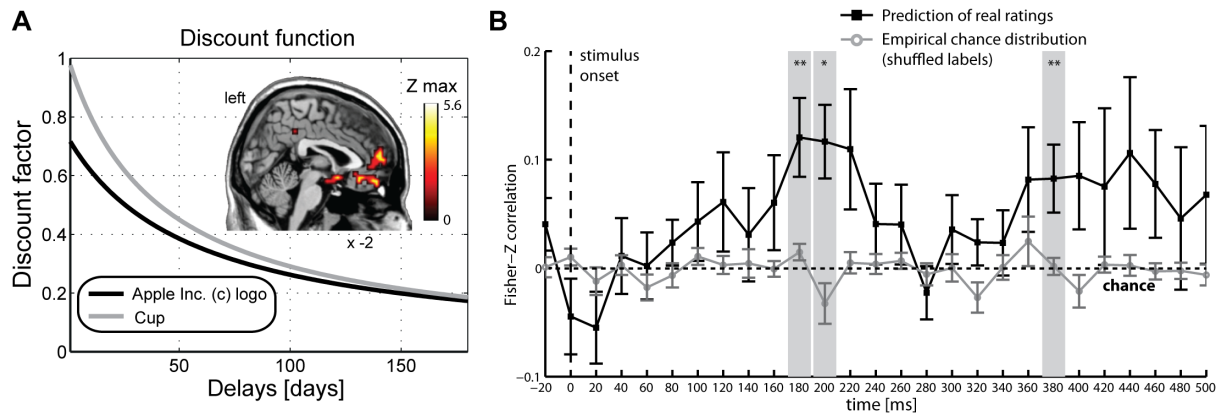## 4. Contextual biases in financial decision-making

Many everyday decisions differ from perceptual decisions in that they are more complex, based on existing preferences and subjective values, and might involve explicit monetary returns or other benefits. The decision to purchase a car, for example, requires the integration of several sources of information, often including knowledge about potential products or product features (e.g., quality, prices, brands), as well as information inherent to the decision-maker, including financial situation, motivational state, past experiences, attitudes and preferences (Glimcher, 2004; Rangel, Camerer, & Montague, 2008). The integration of all these aspects will result in assigning subjective values to potential choice options, which then might drive choice behaviour. However, this is not necessarily a conscious process, and contextual factors might bias decision goals (Custers & Aarts, 2010), even though, intuitively, we might believe that such decisions should depend strongly on exiting, stable preferences. The interesting scenarios, however, are those in which no strong preferences for any of the choice options exist, which will be termed *preference ambiguity* here, similar to perceptual ambiguity as in the previous chapter. One experimental task, which has frequently been used to create such ambiguous preference situations, is temporal discounting (TD) in which participants are asked to choose between a smaller, sooner reward and a larger, later reward (e.g., Berns, Laibson, & Loewenstein, 2007). In this task participants usually discount the delayed reward, meaning that they need to be compensated for waiting, which typically leads to hyperbolic discounting (e.g., Kable & Glimcher, 2007). Based on such a discounting model, it is then possible to vary reward sizes and delay periods to elicit participants' indifference point, or *indecision point*, which is the point where the smaller, sooner reward and the larger, later reward are statistically equally likely to be chosen, constituting a measure of participants' *intertemporal preferences*. These preferences are rather stable and often show systematic differences between personality types, such as trait impulsivity, and associations

with specific clinical populations (Peters & Büchel, 2011). However, it has also been suggested that situation-specific intertemporal preferences can be biased by contextual cues, such as rewarding images and positive mental imagery (Cheng, Shein, & Chiou, 2012; Peters & Büchel, 2010; Simmank, Murawski, Bode, & Horstmann, 2015; Wilson & Daly, 2004; Woelbert & Goebel, 2013; Zhong & DeVoe, 2010). In the following section, the study by Murawski and colleagues (2012) will be described that combined a classical TD task with a subliminal priming procedure (e.g., Dehaene et al., 2003; Krüger, Klapötke, Bode, & Mattler, 2013; Vorberg, Mattler, Heinecke, Schmidt, & Schwarzbach, 2003) conducted in a Siemens MRI scanner at 3 Tesla, to investigate the influence of rewarding but otherwise incidental contextual cues on participants' intertemporal preferences.

Murawski and colleagues (2012) (Appendix VI) asked their participants to choose between a smaller, immediately available reward (always $20 today) and a larger reward available in the near future (e.g. $50 in 10 days) in every trial. The study incorporated the combination of six different delays (1, 10, 21, 55, 90, and 180 days) and six different amounts per delay (ranging from $20.10 to $385.16). Indifference points (assuming hyperbolic discounting; for details see *Materials and Methods* in Murawski et al., 2012) were calculated from participants' choice behaviour. In order to encourage decisions that accorded with participants' true preferences, they could win one of their choices after the experiment (paid out immediately, or after the chosen delay). The crucial manipulation in the experiment was the implementation of two priming conditions: preceding each temporal discounting decision, participants were presented with either an image of the Apple Inc. ® logo, acting as the incidental rewarding prime (its effectiveness was determined using an independent behavioural experiment; see Murawski et al., 2012), or a perceptually closely matched 'neutral' image of a coffee cup, which had no strong rewarding properties. Each prime image was shown for only 16.67 ms and was strongly masked using a sandwich pattern mask such that participants reported not having been able to identify the images during the experiment.

The results showed that in trials in which participants were exposed to the Apple Inc. ® logo, they significantly shifted their intertemporal preferences towards immediate rewards, compared to neutral coffee cup priming trials in which no shift in preferences was observed. In support, a computational model, which incorporated the priming effect, fitted the behavioural data better than a standard discounting model, which did not incorporate the priming effect (confirmed by likelihood ratio tests for individual participants by means of Bayesian Information Criterion, BIC, and Akaike Information Criterion, AIC; significant for nine out of 13 participants). On average, participants needed to receive $1.17 more for the delayed option at a delay of 180 days in order to be indifferent between both options. The fMRI analysis first used 'searchlight' MVPA (see Chapter 2) to identify brain regions in which spatial activation patterns were predictive for the decision outcomes ('sooner option' vs. 'later option'). These were found in the medial orbito-frontal cortex (mOFC), medial prefrontal cortex (mPFC), anterior parts of the anterior cingulate cortex (ACC), and medial frontopolar cortex (FPC). Second, in a mass-univariate GLM analysis, parametric regressors were first used to search for brain region in which the BOLD signal was systematically altered by changes in assigned subjective values while controlling for the priming effect; and the second, crucial analysis searched for brain regions in which activation reflected *changes in subjective value due to priming* while controlling for subjective value encoding that was not influenced by the primes (for details see *Materials and Methods* in Murawski et al., 2012). It was found that most regions, which encoded subjective value, were also sensitive for the priming-induced changes in subjective value, as predicted by the computational model. These regions included mPFC, mOFC, ACC, and extended to medial posterior parietal cortex (mPPC), inferior temporal sulcus, and the striatum (Figure 3A). Importantly, these regions in which activation reflected the priming-related modulation of subjective value were not restricted to the regions that encoded choice outcomes, as revealed by MVPA. In fact, most of

the regions comprising the brain's reward-network appeared to change their activation profile when participants were primed with the incidental rewarding images.

Figure 3: Contextual effects on financial decision-making. A) Priming intertemporal preferences using subliminal rewarding cues. Displayed is the shift in discount function when primed with the logo of Apple Inc. ©, as compared to an image of a coffee cup as a control stimulus for an exemplary participant. The brain image shows some medial prefrontal, striatal and posterior parietal brain regions in which the priming effect systematically changed the representation of subjective value at group level. See text for a full list of regions (modified from Murawski et al., 2012). B) Information time-course for the prediction of post-experiment 'time-reference' ratings for rewarding images, which were passively presented in the background during the experiment. Results from a SVR analysis on distributed spatio-temporal patterns of ERPs showed that significant prediction was possible during early visual processing, providing evidence for an automatic processing of abstract, decision-relevant stimulus features (modified from Bode et al., 2014a).

The study by Murawski and colleagues (2012) provided behavioural and neural evidence for the influence of incidental, rewarding contextual cues on financial choices. In another study it has since been shown that rewarding images might not only bias decision-makers towards more impatient choices (i.e., amplifying the tendency to choose immediate rewards), but some images might have the opposite effect and bias decision-makers towards more future-oriented decisions (Simmank et al., 2015). This means that contextual biases

might operate at a higher level than simply activating the brain's reward network, and that some decision-relevant *semantic* properties of contextual cues might be automatically processed. In support of this hypothesis, it has been shown that neural activation patterns measured using fMRI during exposure to task-irrelevant stimuli, such as images of cars in the background, allowed for the prediction of hypothetical future purchase decisions for the same cars later on (Tusche, Bode, & Haynes, 2010). These results strongly suggest that high-level abstract features of these stimuli might be automatically processed during passive exposure, even though participants did not pay attention to the images because they were distracted using an attention-demanding foreground task (Tusche et al., 2010). Further evidence for this assumption comes from several studies showing that exposure to unattended stimuli can indeed elicit neural activation, which reflects more abstract category information, such as attractiveness and preferences (O'Doherty, Winston, Critchley, Perrett, Burt, & Dolan, 2003; Tusche, Kahnt, Wisniewski, & Haynes, 2013). In conclusion, the observed effects of incidental rewarding images on intertemporal preferences (Murawski et al., 2012; Simmank et al., 2015) could potentially have resulted from fast and automatic (and not necessarily conscious) processing of the temporal dimension (here termed *time reference*) of the stimuli, i.e. whether stimuli were perceived by the decision-makers as being related to the present or to the future. 'Present-related' rewarding images might then bias decision-makers towards impatient choices for immediate rewards (e.g., the Apple Inc. ® logo might be strongly associated with the wish to immediately acquire highly desirable products), while 'future-related' images might exert biases towards more patient choices and delayed gratification. In the following section, the study by Bode and colleagues (2014a) is described that used MVPA for ERPs to investigate whether information about the abstract *time reference* dimension of rewarding images could be processed automatically and unconsciously during passive stimulus exposure (Appendix VII).

In each trial, Bode and colleagues (2014) presented their participants with one of 24 positive, rewarding images (according to available norms, and confirmed by participants' post-experimental ratings) for a duration of 3.2 s. Images were taken from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008) and preselected in an independent behavioural experiment such that they were all positive, but varied substantially on the dimensions *arousal* and *time reference* (for details see *Material and Methods* in Bode et al., 2014a). In the EEG experiment, each image was shown three times in a randomised order in each of the six experimental runs. Importantly, the images were completely task-irrelevant, and at this stage, participants were unaware of the purpose of the experiment. Participants were only asked to closely monitor a small fixation box in the centre of the screen and press the left or right response button with their respective index finger whenever the box opened to the left or to the right. Each trial always began with the presentation of an image and a superimposed fixation box for 400 ms, which subsequently opened to a randomly selected side for a period of 400 ms, and continued to alternate between being closed and open in steps of 400 ms until the trial concluded. This task was adapted from similar attention-demanding foreground tasks used in previous studies to withdraw attention from background images (Tusche et al., 2010; Bogler et al., 2011). During the experiment, 64-channel EEG was recorded at a sampling rate of 512 Hz. Standard off-line pre-processing was conducted using the EEGLab toolbox for MATLAB (Delorme & Makeig, 2004), followed by a current source density (CSD) analysis (Perrin et al., 1987) (for details see *Materials and Methods* in Bode et al., 2014a). After the experiment, participants were shown the same images again and were asked to rate them on several dimensions, including *arousal* ("How arousing/exciting do you think the image is?") and *time reference* ("How strongly do you think the image is associated with the present or the future?"). They gave their answers on a nine-point Likert-scale (from "very arousing / exciting" to "not arousing / exciting at all"; and "strongly related to the present" to "strongly related to the future"), which was subsequently

converted into increments of 'low', 'medium' and 'high' for both dimensions to facilitate the MVPA ERP analysis. For this analysis, individual trials were sorted according to their ratings on both dimensions independently, and a linear SVR algorithm was used to regress individual ratings from the distributed patterns of ERPs in independent analysis time windows of 40 ms, which were moved through the trial with a steps size of 20 ms. This analysis was similar to SVM classification, but it allowed the regression of the continuous measures directly from patterns of ERPs (see Chapter 2). Each analysis time window contained spatial information from all 64 channels as well as temporal information from 20 data points (= 1280 features). The results showed that ratings for both dimensions could be regressed significantly above chance level (as determined using a strict shuffled-label control analysis; see Chapter 2) from distributed patterns of ERPs within the first 100 to 200 ms of visual processing (see Figure 3B for *time reference* results). The underlying feature weight maps (an approximate scalp topography index for the sources of this information) as well as the results of an additional temporal SVR analysis for each channel suggested that the origin of the predictive information differed between both dimensions.

These results strongly suggest that the abstract feature dimension *time reference* was automatically processed during relatively early stages of visual processing, even though participants did neither pay attention to the stimuli nor were they aware of the importance of their temporal aspect. These findings therefore lent plausibility to the assumption that *time reference* could be an important stimulus feature, which decision-makers automatically assess when being exposed to stimuli in their environment. This is particularly plausible when considering that, when processing most (if not all) rewarding stimuli in the environment, their potential for consumption and their hedonistic value might be analysed fast and unconsciously, as one must constantly decide whether the stimulus constitutes an immediately available option, or whether it should be considered in the future. This is surely true for any food displayed in windows of bakeries and restaurants, or for most non-consumable products

in shopping windows, but also for more abstract stimuli such as advertisements for holiday trips. The same further applies to all situations, which lend themselves to rewarding social interactions, as these could also take place immediately but might sometimes be wisely postponed to the future. Thus, the results from Bode and colleagues' (2014a) study provide an important insight into potential mechanisms driving the influence of contextual cues on decision-making, and suggest that decision-relevant properties of rewarding cues are automatically extracted and could have carry-over effects on incidental decision problems under preference ambiguity. This means, similar to decisions in perceptually ambiguous situations, more complex financial and intertemporal decisions are also susceptible to the decision context, potentially modulated by similar decision-making mechanisms (Hare, Schultz, Camerer, O'Doherty, & Rangel, 2011; Krajbich, Lu, Camerer, & Rangel, 2012).

5. Contextual biases in voluntary decision-making

In recent years, another area that has received substantial attention and has been a vibrant field of research is voluntary decision-making (Brass, Lynn, Demanet, & Rigoni, 2013; Haggard, 2008; Haynes, 2011). This topic draws its attractiveness from its apparent relevance to philosophical problems such as the 'free will' debate, and some earlier attention-grabbing findings in cognitive neuroscience demonstrating decision-related neural signals in the human ERP several hundred milliseconds before the decision-makers were aware of 'wanting' to make a decision (Libet, Gleason, Wright, & Pearl, 1983; Libet, 1985). These findings have been replicated and expanded on many times (e.g., Fried, Mukamel, & Kreiman, 2011; Haggard & Eimer, 1999; Schneider, Houdayer, Bai, & Hallett, 2013), but their validity as well as the origin and meaning of these signals have been intensely debated (e.g., Batthyany, 2009; Bode et al., 2014; Brembs, 2012; Haynes, 2011; Joordens, van Duijn, & Spalek, 2002; Roskies, 2010; Trevena & Miller, 2002). One attractive feature of voluntary decisions for the current work is that they are conceptualised as *ambiguous decision scenarios* in which neither decision outcome is associated with a stronger preference, with more sensory evidence, or with any other benefit, which should drive choice. The idea behind the experimental approach is to create decision problems that are "maximally non-stimulus driven" (Haggard, 2008) such that participants are as free as possible to choose one out of several (usually two) available choice options (e.g., Mueller, Brass, Waszak, & Prinz, 2007), or to choose the time point at which a behaviour (usually a motor action) occurs (e.g., Brass & Haggard, 2007). This obviously constitutes an extremely artificial choice situation, and it is indeed questionable to what extent such a free-choice scenario can truly be achieved. We have argued that such choice scenarios are more likely to conceal the true drivers of individual decisions than being truly non-stimulus driven, as the latter would require the absence of all external stimulation and all potential internal variables (e.g., memory, preference, physiological states), which is

highly unlikely to be possible (Bode et al., 2014; also see Schüür & Haggard, 2011). Nevertheless, as an extreme case of ambiguous choice scenarios, Libet-style voluntary decision experiments are highly suitable for the investigation of contextual effects on decision-making. The following section will first briefly introduce a seminal experiment by Soon and colleagues (2008), which used a novel adaptation of Libet's original experiment for fMRI and MVPA. Then our experiment will be discussed, which incorporated one important extension to the original study and provided evidence for early choice predictive brain activity for *abstract* voluntary decisions (Soon et al., 2013; Appendix VIII). This will be followed by a discussion of a direct replication of Soon and colleagues' (2008) original study, conducted by Bode and colleagues (2011) that additionally allowed for the investigation of the potential formation of choice biases (Appendix IX).

In their fMRI study, Soon and colleagues (2008) asked participants to make simple voluntary motor decisions by pressing the left or the right button with the respective index finger at a time point of their own choosing. Participants were additionally asked to report the time point at which they became aware of their decision immediately before performing the action, by recounting the letter that was displayed on the screen at that moment. For this purpose, letters on the screen were refreshed every 500 ms, providing participants with a 'clock' to measure their 'time of will'. The authors then used MVPA in a time-resolved fashion to predict the outcome of the decisions before the reported time point of conscious decision-making. It was found that decisions could be predicted significantly above chance from spatial patterns of brain activation in the medial posterior parietal cortex (mPPC), mainly in the precuneus and the posterior cingulate cortex (PCC), as well as in the medial frontopolar cortex (FPC), which has been associated with high-level planning and information integration (e.g., Burgess, Dumontheil, & Gilbert, 2007; Koechlin & Hyafil, 2007). Crucially, the onset of predictive activation was found ~7 s (possibly 10 s when taking into account the lag in the haemodynamic response) before the reported time point of conscious decision-

making. As it has been argued that 'real' decisions require more complex choice options than motor actions, which may be driven by simpler unconscious 'urges' (Batthyany, 2009), the study by Soon and colleagues (2013) aimed at replicating the original findings using abstract decisions in a similarly ambiguous choice situation. Their task required participants to make decisions between two possible arithmetic tasks: adding or subtracting two numbers. Participants were again placed in a Siemens 3 Tesla MRI scanner and were instructed to relax while a continuous stream of letters and single digit numbers was presented in the centre of the screen, refreshed every 500 ms. The letters again served as the clock, and participants were asked to remember the letter on the screen when they first became aware of their decision to add or to subtract. Once they had made their decision, they were instructed to immediately perform the mental calculation using the two digits presented on the next two consecutive screens. This procedure prevented the use of favourite numbers to trigger decision-making, as the numbers to perform the arithmetic operation on were revealed only after the decision was made. In order to keep both mental operations equally simple, the smaller number always had to be subtracted from the higher number. In addition to the central numbers and letters, each screen also contained four single digit numbers displayed in the four corners of the screen. These always contained the correct answers for adding and subtracting the previous two numbers, as well as two incorrect solutions (as a control for task engagement), which were always one digit smaller or higher than the correct solutions. The positions of the correct solutions were pseudo-randomised across screens. Participants used four response buttons, mapped onto the four corners of the screen, to indicate their responses. After having revealed their choice by selecting one of the solutions, a second response screen asked them to state which letter was on the screen when they made their conscious decision to add or subtract. The MVPA analysis then followed the logic of Soon and colleagues' (2008) study and used a time-resolved 'searchlight' decoding approach based on an FIR model that allowed for the analysis of brain activation patterns in time-steps of 2 s (see Chapter 2, and

Experimental Procedures in Soon et al., 2013). The results showed that the same brain regions as reported in the original study, comprising of mPPC, including the precuneus and PCC, and anterior FPC, encoded the upcoming abstract decisions up to ~5s before (possibly even earlier) the reported conscious time point of decision-making. These findings confirmed that predictive information for decisions of different levels of abstraction was represented in activation patterns seconds before the decision-makers' reported consciously deciding on an option. While the time point of decision prediction was astonishingly early in these studies, the question how these signals were generated was left unanswered.

In order to further investigate the nature of the choice-predictive activity patterns, in particular in medial FPC, Bode and colleagues (2011) replicated Soon and colleagues (2008) paradigm for voluntary motor decision in combination with 7 Tesla ultra-high field fMRI. Their task was identical to the original study: participants were asked to relax until they felt the urge to make a decision for either the left or the right response button, and they again used the letters on the screen (refreshed every 500 ms) to report the time point of their conscious decision. Using an improved TR of 1.5 s, and an improved spatial resolution of 1 mm^3 isotropic voxels (i.e., 27-times smaller MVPA 'searchlight' clusters; Figure 4A), this study could investigate the evolution of pattern similarity and stability over time (for details see *Materials and Methods* in Bode et al., 2011). The study replicated the original findings and showed that decision outcomes could be predicted with significant accuracy from clusters in anterior FPC ~7 s before the reported the conscious decision. Furthermore, with increasing temporal proximity to the time point of the decision, the prediction accuracy increased (Figure 4B); and importantly, the similarity of activation patterns with the activation pattern at the time point just before becoming aware of the decision (i.e. the most prototypical decision-related activation pattern) increased as well (Figure 4C). One plausible explanation for this finding is that during the unconscious phase of decision-formation, the activation patterns slowly evolved towards a pattern that was then strongly associated with the final conscious

decision. This process at the neural level had remarkable resemblance to a diffusion process towards a decision boundary (albeit much slower), as conceptualised in the DDM (Smith & Ratcliff, 2004). It can therefore be speculated that something similar was reflected in these slowly evolving neural patterns, potentially the accumulation of internal decision-related signals. In support, other authors have offered similar explanations for evolving voluntary decisions that are based on evidence accumulation models (Mattler & Palmer, 2012; Schurger, Sitt, & Dehaene, 2012).
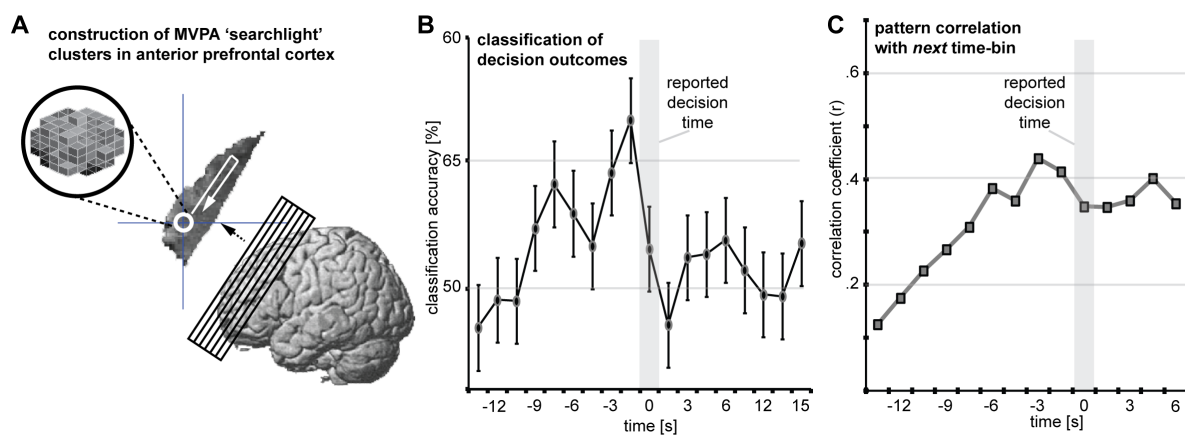
Figure 4: *Prediction of voluntary decisions for left and right button presses from brain activity before reported conscious awareness. A) Illustration of the construction of 'searchlight' clusters from anterior FPC using ultra high-field fMRI at 7 Tesla. The voxel size was 1 x 1 x 1 mm, and the cluster radius was 3 voxels (modified from Bode et al., 2011). B) Prediction of decision outcomes from most informative cluster in anterior FPC over time. These findings show a slow build-up of information towards the reported time point of conscious decision-making, starting several seconds before (modified from Bode et al., 2011). C) Correlation of activation patterns from the most informative cluster in anterior FPC from each time window with the activation patterns from the following analysis time windows (averaged across both decisions). An increase in pattern similarity was observed until the time point of the reported conscious decision was reached (modified from Bode et al., 2011).*

Finally, the question that arises is why predictive information in brain activity patterns is found so early, and what might drive decision-making in these seemingly arbitrary choice

scenarios. The following discussion is partly based on ideas expressed in Bode and colleagues (2014b), which provided an in depth theoretical analysis of the nature of voluntary decisions, and the role of contextual information (Appendix X). In the absence of external factors, several scenarios that could explain the early predictive activation patterns are conceivable. First, decisions in each trial might be truly random. This is unlikely, however, considering that Bode and colleagues (2012b) demonstrated significant choice sequence effects in a related scenario where external information was absent. Indeed, while on average participants in Bode and colleagues (2011) chose each choice option equally often, and indicated having been 'spontaneous most of the time' (as assessed by post-experimental questionnaires), the analysis of sequential effects at a participant level remained inconclusive due to the low number of trials in each run (on average only 11 trials per 5min run were available). If anything, it should be surprising that participants were rather balanced in their choices, given that truly random sequences could be expected to be far more imbalanced for such short sequence lengths. This suggests a more likely second possibility – that participants produced, to some extent, predictable choice patterns. The higher frequency of short sequence lengths in these studies further point towards this possibility (Lages, Boyle, & Jaworska, 2013). In support, it has been shown recently that in the absence of external information that could otherwise guide decision-making, the previous choice can become the only 'feedback' available (Akaishi et al., 2014). These authors suggested that having chosen an option before might add value to the respective choice option, thereby increasing the likelihood of choice repetitions. Complex choice history effects have also been shown in monkeys, and the expression of these effects strongly depended on the exact nature of the task (Mochizuki & Funahashi, 2014). The most direct demonstration of choice history effects in voluntary decision-making comes from a behavioural study by Lages and Jaworska (2012), who directly replicated the experimental paradigms by Soon and colleagues (2008) and Bode and colleagues (2011). These authors found that in ten out of twelve of their participants it was

possible to use the same multivariate classification approach as in the fMRI studies to predict the upcoming choices from the previous choice history alone, without any reference to brain activity. Their classification accuracy was significant at group level and of similar magnitude to the accuracies reported in the original studies (Lages & Jaworska, 2012). Even though this demonstration of dependencies in sequences of voluntary decisions does not invalidate that brain activation patterns preceding choices strongly predict the upcoming choice (Soon et al., 2014), their existence again supports the hypothesis that contextual effects have a significant influence in shaping decisions in ambiguous decision situations (Bode et al., 2014).

Moreover, it is highly likely that in voluntary decision tasks another important factor might influence choice behaviour, namely participants' implicit understanding of task instructions. These tasks are, as argued above, extremely artificial in that there appear to be neither preferences, external evidence, nor explicit instructions for why one should choose one alternative over the other. But participants are always instructed to choose something – they are offered a limited set of alternatives (usually two), from which they are expected to choose repeatedly, and both alternatives are always available in each trial. It is difficult to conceive how such a situation would not trigger an implicit idea of what kind of choice behaviour might be expected from them. Indeed, in most voluntary decision-making studies mentioned above (Bode et al., 2011; Soon et al., 2008; 2013), while participants were not explicitly told how to choose, they were nevertheless preselected with respect to their natural tendency to be "inherently" balanced between the available choice options (e.g., Bode et al., 2011). This means, their initial *implicit* understanding of the task instructions might have been to be rather balanced between the two options – and this then constituted an important – but hidden – contextual driver of their decision-making. As argued above, typical voluntary decision tasks conceal these influences, and participants who might become aware of their existence and report them afterwards might even be excluded from the analyses. Importantly, however, counter-measures such as not explicitly instructing balanced choices and post-hoc

exclusions do not protect against the existence of such biases – these measures would simply render their origins invisible, make them inaccessible for analyses, and create the spurious assumption that participants' decision behaviour was random and spontaneous. A decision situation, which is truly independent from any external and internal context, is simply unconceivable (Bode et al., 2014; Schüür & Haggard, 2011). As already Thorndike (1911) noted at the beginning of the 20th century, *"[...] voluntary behaviour might be organised in loops rather than in single chains of events, as they do not need an initial cause; a free decision might be best described as "intelligent interaction with the animal's current and historical context*". In that sense, voluntary decisions would be, even more than any other type of decision-making discussed in this work, susceptible to subtle contextual influences, because they create their own context.

6. Conclusion and remarks

Using a MVPA approach for patterns of fMRI data and ERPs (Bode & Haynes, 2009; Bode & Stahl, 2014), the present work has demonstrated significant biases exerted by contextual information on decision-making across a variety of decision types. These biases can originate from external factors, such as contextual cues in the environment. This was shown using the example of shifts in intertemporal preferences after exposure to incidental rewarding cues in situations of preference ambiguity (Murawski et al., 2012). An explanation for this effect could be that participants automatically processed higher-level semantic attributes, such as stimuli's *time reference*, even from contextual stimuli that were not attended and task-irrelevant (Bode et al., 2014a). Contextual biases can further originate from the history of previous choices, as shown for perceptual decisions in situations of perceptual ambiguity in which guessing was required (Bode et al., 2012b). Results from this study provided support for the hypothesis that these biases were expressed as a shift in starting point for the following evidence accumulation process, as conceptualised in the DDM. This was reflected in the neural state of the decision system prior to perceptual processing, and became assessable using MVPA (Bode et al., 2012a; 2012b; 2013). Similarly, participants' choice history has been argued to be an important contextual factor explaining early decision-predictive brain signals in fMRI studies on voluntary decision-making (Bode et al., 2011; Soon et al., 2013). In addition, the internalised implicit understanding of the task instructions in these ambiguous choice situations is highly likely to constitute another, often overlooked contextual factor biasing sequences in voluntary decision-making (Bode et al., 2014b).

The present work demonstrated that multivariate analysis methods constitute a powerful toolset for the investigation of human decision-making and contextual biases. First, in several of the studies discussed here, the decision-related information could not have been revealed using conventional methods. This is because activation patterns can represent

information without a significant enhancement of the local average BOLD signal (e.g., Bode et al., 2011; Bode & Haynes, 2009), and without strong differences between conditions of interest in single ERP components (e.g., Bode et al., 2014a; Bode & Stahl, 2014). Often, however, it is not sufficient to demonstrate the presence of information in brain activity that may give rise to the misleading impression of having discovered a novel neural mechanism (see the initial interpretations in Bode et al., 2012a; 2013), and that alone do not necessarily add much to the understanding of the underlying cognitive processes (as in Bode et al., 2011; Soon et al., 2013). Instead, the careful combination of MVPA with analysis of behaviour, for example by means of response times, error rates, choice probabilities, response force, or cognitive modelling (e.g., Bode et al., 2012b; Bode & Stahl, 2014a; Murawski et al., 2012), is highly advisable to gain a better understanding of the origins of the uncovered information in the neural signals.

The studies discussed here also reflect a recent trend in the cognitive sciences to investigate contextual influences on choice behaviour. However, despite this topic's current attractiveness, it is also obvious that much more work is needed to understand the full extent – as well as the limits – of contextual biases on various decision types. In particular, the field of voluntary decision-making still mostly ignores the idea of sequential effects, and several authors maintain their (unrealistic) assumption that decisions can be truly spontaneous and unbiased. It would be most desirable if this assumption could finally be abandoned; and instead of concealing contextual influences, studies would focus more strongly on addressing them. In other scenarios, it will be of great importance to further explore under which circumstances contextual information can bias decisions, in particular for applications such as 'nudging' decision-makers towards desirable choice options (e.g., Felsen & Reiner, 2015), for example when changing behaviour towards more healthy choices (e.g., Schonberg, Bakkour, Hover, Mumford, & Poldrack, 2014). All examples discussed here explored the influence of contextual cues in situations of strong preference ambiguity, perceptual ambiguity, or choice

ambiguity. But how indifferent between available options must a decision-maker be in order to be successfully biased by contextual cues? Or, in other words, how strong must a contextual cue be in order to still be successful if the decision-maker already has some (weak) preference for a specific option?

Finally, despite the great importance of contextual biases, this work is not meant to depict humans as purely context-driven agents. It should be noted that situational contextual factors are only one class of variables that can bias behaviour, while in real life the much more important driver for decision-making is the interplay of people's preferences, personality, motivations, memory, and learning. These factors constitute the (often unconscious) 'internalised' context against which decisions are made every day. Gaining a better understanding of how such preferences and motivations are shaped by constant updating of beliefs, learning, and making of new experiences in our interactions with the world, and how these updates are represented in the brain (e.g., Bennett, Murawski, & Bode, 2015), constitutes another exiting strand of research strongly related to the questions discussed here.

7. References

Akaishi, R., Umeda, K., Nagase, A., & Sakai, K. (2014). Autonomous mechanism of internal choice estimate underlies decision inertia. *Neuron, 81*(1), 195-206.

Batthyany, A. (2009). Mental causation and free will after Libet and Soon: reclaiming conscious agency. In: A. Batthyany, & A. C. Elitzur (Eds.), *Irreducibly Conscious. Selected Papers on Consciousness*. Heidelberg: Winter.

Bennett, D., Murawski, C., & Bode, S. (2015). Single-trial event-related potential correlates of belief updating. *eNeuro, 2*(5).

Berns, G. S., Laibson, D., & Loewenstein, G. (2007). Intertemporal choice – toward an integrative framework. *Trends in Cognitive Science, 11*(11), 482-488.

Bode, S., Bennett, D., Feuerriegel, D., & Alday, P. (2016). The Decision Decoding Toolbox – a multivariate pattern analysis toolbox for event-related potentials. DOI: 10.5281/zenodo.48143.

Bode, S., Bennett, D., Stahl, J., & Murawski, C. (2014). Distributed patterns of event-related potentials predict subsequent ratings of abstract stimulus attributes. *PLoS ONE, 9*(10): e109070.

Bode, S., Bogler, C., & Haynes, J. D. (2013). Similar neural mechanisms for guesses and free decisions. *Neuroimage, 65*(2), 456-465.

Bode, S., Bogler, C., Soon, C. S., & Haynes, J. D. (2012). The neural encoding of guesses in the human brain. *Neuroimage, 59*(2), 1924-1931.

Bode, S., & Haynes, J. D. (2009). Decoding sequential stages of task preparation in the human brain. *Neuroimage, 45*(2), 606-613.

Bode, S., He, A. H., Soon, C. S., Trampel, R., Turner, R., & Haynes, J. D. (2011). Tracking the unconscious generation of free decisions using ultra-high field fMRI. *PLoS ONE, 6*(6): e21612.

Bode, S., Murawski, C., Soon, C. S., Bode, P., Stahl, J., & Smith, P. L. (2014). Demystifying "free will": The role of contextual information and evidence accumulation for predictive brain activity. *Neuroscience & Biobehavioral Reviews, 47,* 636-645.

Bode, S., Sewell, D., Lilburn, S., Forte, J., Smith, P. L., & Stahl, J. (2012). Predicting perceptual decision biases from early brain activity. *The Journal of Neuroscience, 32*(36), 12488-12498.

Bode, S., & Stahl, J. (2014). Predicting errors from patterns of event-related potentials preceding an overt response. *Biological Psychology, 103,* 357-369.

Bogler, C., Bode, S., & Haynes, J. D. (2011). Decoding successive computational stages of saliency processing. *Current Biology, 21*(19), 1667-1671.

Brass, M., & Haggard, P. (2007). To do or not to do: the neural signature of self-control. *The Journal of Neuroscience, 27*, 9141-9145.

Brass, M., Lynn, M. T., Demanet, J., & Rigoni, D. (2013). Imaging volition: what the brain can tell us about the will. *Experimental Brain Research, 299*, 301-312.

Brembs, B. (2012). Towards a scientific concept of free will as a biological trait: spontaneous actions and decision-making in invertebrates. *Proceedings of the Royal Society of London. Series B, Biological Sciences, 278*, 930-939.

Bronfman, Z. Z., Brezis, N., Moran, R., Tsetsos, K., Donner, T., & Usher, M. (2015). Decisions reduce sensitivity to subsequent information. *Proceedings of the Royal Society B: Biological Sciences, 282*: 20150228.

Bunge, S. A. (2004). How we use rules to select actions: a review of evidence from cognitive neuroscience. *Cognitive, Affective, & Behavioral Neuroscience, 4*(4), 564-679.

Burgess, P. W., Dumontheil, I., & Gilbert, S. J. (2007). The gateway hypothesis of rostral prefrontal cortex (area 10) function. *Trends in Cognitive Science, 11*, 290-298.

Carlson, T. A., Schrater, P., & He, S. (2003). Patterns of activity in the categorical representation of objects. *Journal of Cognitive Neuroscience, 15,* 704-717.

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and its behavioural correlates. *Brain, 129*(3), 564-583.

Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on intelligent Systems and Technology, 2*(3), 27.

Charles, L., King, J. R., & Dehaene, S. (2014). Decoding the dynamics of action, intention, and error detection for conscious and subliminal stimuli. *The Journal of Neuroscience, 34*(4), 1158-1170.

Chen, M. Y., Jimura, K., White, C. N., Maddox, W. T., & Poldrack, R. A. (2015). Multiple brain networks contribute to the acquisition of bias in perceptual decision-making. *Frontiers in Neuroscience, 9*: 63.

Cheng, Y. Y., Shein, P. P., & Chiou, W. B. (2012). Escaping the impulse to immediate gratification: the prospect concept promotes a future-oriented mindset, prompting an inclination towards delayed gratification. *British Journal of Psychology, 103*, 129-141.

Clithero, J. A., Carter, R. M., & Huettel, S. A. (2009). Local pattern classification differentiates processes of economic valuation. *Neuroimage, 45*, 1329-1338.

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging fMRI "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage, 19*(2 pt 1), 261-270.

Custers, R., & Aarts, H. (2010). The unconscious will: how the pursuit of goals operates outside of conscious awareness. *Science, 329*, 47-50.

Das, K., Giesbrecht, B., & Eckstein, M. P. (2010). Predicting variations of perceptual performance across individuals from neural activity using pattern classifiers. *Neuroimage, 51*, 1425-1437.

Davatzikos, C., Ruparel., K., Fan, Y., Shen, D. G., Acharyya, M., Loughead, J. W., … Langleben, D. D. (2005). Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. *Neuroimage, 28*(3), 663-668.

Deco, G., & Romo, R. (2008). The role of fluctuations in perception. *Trends in Neuroscience, 31*(11), 591-598.

Dehaene, S., Artiges, E., Naccache, L., Martelli, C., Viard, A., Schürhoff, F., … Martinot, J. L. (2003). Conscious and subliminal conflicts in normal subjects and patients with schizophrenia: the role of the anterior cingulate. *Proceedings of the National Academy of Sciences of the United States of America, 100*, 13722-13727.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods, 134*, 9-21.

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components: II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology, 78*, 447-455.

Felsen, G., & Reiner, P. B. (2015). What can neuroscience contribute to the debate over nudging? *Review of Philosophy and Psychology, 6*(3), 469-479.

Fried, I., Mukamel, R., & Kreiman, G. (2011). Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron, 69*, 548-562.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., & Frackowiak, R.S. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping, 2*, 189-210.

Fründ, I., Wichmann, F. A., & Macke, J. H. (2014). Quantifying the effect of intertrial dependence on perceptual decisions. *Journal of Vision, 14*(7), 9.

Gehring, W. J., Goss, B., Coles, M. G., Meyer, D. E., & Donchin, E., (1993). A neural system for error detection and compensation. *Psychological Science, 4*(6), 385–390.

Gilbert, S. J. (2011). Decoding the content of delayed intentions. *The Journal of Neuroscience, 31*(8), 2888-94.

Glimcher, P. W. (2004). *Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics*. Cambridge (Massachusetts): MIT Press.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience, 30*, 535-374.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology, 55*, 468-484.

Haggard, P. (2008). Human volition: towards a neuroscience of will. *Nature Reviews Neuroscience, 9*, 934-946.

Hare, T. A., Schultz, W., Camerer, C. F., O'Doherty, J. P., & Rangel, A. (2011). Transformation of stimulus value signals into motor commands during simple choice. *Proceedings of the National Academy of Sciences of the United States of America, 108*(44), 18120-18125.

Hawkins, G., Brown, S. D., Steyvers, M., & Wagenmakers, E. J. (2012). Decision speed induces context effects in choice. *Experimental Psychology, 59*(4), 206-215.

Haynes, J. D. (2011). Decoding and predicting intentions. *Annals of the New York Academy of Sciences, 1224*, 9-21.

Haynes, J. D. (2015). A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. *Neuron, 87*(2), 257-270.

Haynes, J. D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience, 8*(5), 686-691.

Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience, 7*(7), 523-534.

Haynes, J. D., Sakai, K., Rees, G., Gibert, S., Frith, C., & Passingham, R. E. (2007). Reading hidden intentions in the human brain. *Current Biology, 17*(4), 323-328.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430.

Hebart, M. N., Schriever, Y., Donner, T. H., & Haynes, J. D. (2016). The relationship between perceptual decision variables and confidence in the human brain. *Cerebral Cortex, 26*(1), 118-30.

Heekeren, H. R., Marrett, S., & Ungerleider, L.G. (2008). The neural systems that mediate human perceptual decision making. *Nature Reviews Neuroscience, 9*(6), 467-479.

Heinzle, J., Anders, S., Bode, S., Bogler, C., Chen, Y., Cichy, R. M., … Haynes, J. D. (2012). Multivariate decoding of fMRI data – Towards a content-based cognitive neuroscience. *e-Neuroforum, 3*(1)*,* 1-16.

Hesselmann, G., Kell, C. A., Eger, E., & Kleinschmidt, A. (2008). Spontaneous local variations in ongoing neural activity biases perceptual decisions. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 10984-10989.

Howard, J. D., Plailly, J., Grueschow, M., Haynes, J. D., & Gottfried, J. A. (2009). Odor quality coding and categorization in human posterior piriform cortex. *Nature Neuroscience, 12*(7), 932-940.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2,* 194-203.

Joordens, S., van Duijn, M., & Spalek, T. M. (2002). When timing the mind one should also mind the timing: biases in the measurement of voluntary actions. *Consciousness & Cognition, 11*, 231-240.

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience, 10*, 1625-1633.

Kamitani, Y., & Sawahata, Y. (2010). Spatial smoothing hurts localization but not information: Pittfalls for brain mappers. *Neuroimage, 49*(3), 1949-1952.

Katznelson, R. D. (1981). EEG recording, electrode placement, and aspects of generator localization. In: P. L. Nunez (Ed.), *Electric fields of the brain.* Oxford: Oxford University Press.

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature, 452*(7185), 352-356.

King, J. R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Science, 18*(4), 203-210.

Koechlin, E., & Hyafil, A. (2007). Anterior prefrontal function and the limits of human decision-making. *Science, 318*, 594-598.

Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in Psychology, 3,* 193.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America, 103*(10), 3863-3868.

Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S., & Baker, C.I. (2009). Circular analysis in systems neuroscience: the danger of double dipping. *Nature Neuroscience, 12*(5), 535-540.

Krüger, D., Klapötke, S., Bode, S., & Mattler, U. (2013). Cognitive control systems in the human cortex are susceptible to unconscious priming. *Neuroimage, 64*(1), 197-208.

Lages, M., Boyle, S. C., & Jaworska, K. (2013). Flipping a coin in your head without monitoring outcomes? Comments on predicting free choices and a demo program. *Frontiers in Psychology, 4,* 925.

Lages, M., & Jaworska, K. (2012). How predictable are "spontaneous decisions" and "hidden intentions"? Comparing classification results based on previous responses with multivariate pattern analysis of fMRI BOLD signals. *Frontiers in Psychology, 3*, 56.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* Technical Report A-8. Gainesville, FL: University of Florida.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences, 8*, 529-566.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D.K. (1983). Time of conscious intention to act in relation to onset of cerebral activities (readiness-potential): the unconscious initiation of a freely voluntary act. *Brain, 106*(Pt 3), 623-642.

Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the BOLD signal. *Annual Review of Physiology, 66*, 735- 769.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge: MIT Press.

Margulies, D. S., Vincent, J. L., Kelly, C., Lohmann, G., Uddin, L.Q., Biswal, B. B., … Petrides, M. (2009). Precuneus shares intrinsic functional architecture in humans and monkeys. *Proceedings of the National Academy of Sciences of the United States of America, 106*(47), 20069-20074.

Mattler, U., & Palmer, S. (2012). Time course of free-choice priming effects explained by a simple accumulator model. *Cognition, 123*, 347–-360.

Mochizuki, K., & Funahashi, S. (2014). Opposing history effect of preceding decision and action in the free choice of saccade direction. *Journal of Neurophysiology, 112*, 923-932.

Morawetz, C., Bode, S., Baudewig, J., Jacobs, A. M., & Heekeren, H. R. (2015). Neural representations of emotion regulation strategies. *Human Brain Mapping, 37(2)*, 600-620.

Mueller, V. A., Brass, M., Waszak, F., & Prinz, W. (2007). The role of the preSMA and the rostral cingulate zone in initially selected actions. *Neuroimage, 37*(4), 1354-1361.

Mulder, M.J., Wagenmakers, E.J., Ratcliff, R., Boekel, W., & Forstmann, B.U. (2012). Bias in the brain: a diffusion model analysis of prior probability and potential payoff. *The Journal of Neuroscience, 32*, 2335-2343.

Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2009). Revealing representational content with pattern- information fMRI – an introductory guide. *Social Cognitive and Affective Neuroscience, 4*(1), 101-109.

Murawski, C., Harris, P. G., Bode, S., Domínguez D., J. F., & Egan, G. F. (2012). Led into temptation? Rewarding brand logos bias the neural encoding of economic decisions. *PLoS ONE, 7*(3): e34155.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences, 10*(9), 424-430.

O'Doherty, J., Winston, J., Critchley, H., Perrett, D., Burt, D. M., & Dolan, R. J. (2003). Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia, 1*, 147-155.

Op de Beeck, H. P. (2010). Against hyperacuity in brain reading: Spatial smoothing does not hurt multivariate fMRI analyses? *Neuroimage, 49*(3), 1943-1948.

Perrin, F., Bertrand, O., & Pernier, J. (1987). Scalp current density mapping: Value and estimation from potential data. *IEEE Transactions on Biomedical Engineering, 34*, 283-288.

Peters, J., & Büchel, C. (2010). Episodic future thinking reduces reward delay discounting through an enhancement of prefrontal-meditotemporal interactions. *Neuron, 66*, 138-148.

Peters, J., & Büchel, C. (2011). The neural mechanisms of inter-temporal decision-making: understanding variability. *Trends in Cognitive Science, 15*, 227-239.

Philiastides, M. G., & Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral Cortex, 16*, 509-518.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences, 10*(2), 59-63.

Poldrack, R. A., & Farah, M. J. (2015). Progress and challenges in probing the human brain. *Nature, 526*(7573), 371-379.

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience, 9*, 545-556.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59-108.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current issues and history. *Trends in Cognitive Science, 20*(4), 260-281.

Roskies, A. L. (2010). How does neuroscience affect our conception of volition? *Annual Review of Neuroscience, 33*, 109-130.

Romo, R., & Salinas, E. (2003). Flutter discrimination: neural codes, perception, memory and decision making. *Nature Reviews Neuroscience, 4*(3), 203-218.

Schneider, L., Houdayer, E., Bai, O., & Hallett, M. (2013). What we think before a voluntary movement. *Journal of Cognitive Neuroscience, 25*, 822-829.

Schonberg, T., Bakkour, A., Hover, A. M., Mumford, J. A., & Poldrack, R. A. (2014). Influencing food choices by training: evidence for modulation of frontoparietal control signals. *Journal of Cognitive Neuroscience, 26*(2), 247-268.

Schurger, A., Sitt, J. D., & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Sciences of the United States of America*, 109(42), E2904-E2913.

Schüür, F., & Haggard, P. (2011), What are self-generated actions? *Consciousness & Cognition, 20*, 1697-1704.

Schwarzlose, R. F., Swisher, J. D., Dang, S., & Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 105*(11), 4447-4452.

Serences, J. T., & Boynton, G. M. (2007). The representation of behavioural choice for motion in human visual cortex. *The Journal of Neuroscience, 27*(47), 12893-12899.

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology, 86*, 1916-1936.

Simmank, J., Murawski, C., Bode, S., & Horstmann, A. (2015). Incidental rewarding cues influence financial decision-making in obesity. *Frontiers in Behavioral Neuroscience, 9,* 278.

Singh-Curry, V., & Husain, M. (2009). The functional role of the inferior parietal lobe in the dorsal and ventral stream dichotomy. *Neuropsychologia, 47*(6), 1434-1448.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences, 27*(3), 161-168.

Soon, C. S., Allefeld, C., Bogler, C., Heinzle, J., & Haynes, J. D. (2014). Predictive brain signals best predict upcoming and not previous choices. *Frontiers in Psychology, 5*: 406.

Soon, C. S., Brass, M., Heinze, H. J. & Haynes, J. D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience, 11*(5), 543-545.

Soon, C. S., He, A. H., Bode, S., & Haynes, J. D. (2013). Predicting free choices for abstract intentions. *Proceedings of the National Academy of Sciences of the United States of America, 110*(15), 6217-6222.

Swisher, J. D., Gatenby, J. C., Gore, J. C., Wolfe, B. A., Moon, C. H., Kim, S. G., & Tong, F. (2010). Multiscale pattern analysis of orientation-selective activity in the primary visual cortex. *The Journal of Neuroscience, 30*(1), 325-330.

Thorndike, E. L. (1911). *Animal Intelligence: Experimental Studies*. New York: The Macmillan Company.

Todd, M. T., Nystrom, L. E., & Cohen, J.D. (2013). Confounds in multivariate pattern analysis: theory and rule representation case study. *Neuroimage, 77*, 157-165.

Tong, F., & Pratte, M. S. (2012). Decoding patterns of human brain activity. *Annual Review of Psychology, 63*, 483-509.

Trevena, J. A., & Miller, J. (2002). Cortical movement preparation before and after a decision to move. *Consciousness & Cognition, 11*, 162-190.

Trueblood, J.S., Brown, S., Heathcote, A., & Busemeyer, J. R. (2013). Not just for consumers: context effects are fundamental to decision making. *Psychological Science, 24*(6), 901-908.

Tusche, A., Bode, S., & Haynes, J. D. (2010). Neural responses to unattended products predict later consumer choices. *The Journal of Neuroscience, 30*(23), 8024-8031.

Tusche, A., Kahnt, T., Wisniewski, D., & Haynes, J. D. (2013). Automatic processing of political preferences in the human brain. *Neuroimage, 72*, 174-182.

van Ravenzwaaij, D., Mulder, M. J., Tuerlinckx, F., & Wagenmakers, E. J. (2012). Do the dynamics of prior information depend on task context? An analysis of optimal performance and an empirical test. *Frontiers in Psychology, 3*, 132.

van Vugt, M. K., Simen, P., Nystrom, L. E., Holmes, P., & Cohen, J. D. (2012). EEG oscillations reveal neural correlates of evidence accumulation. *Frontiers in Neuroscience, 6,* 106.

Vorberg, D., Mattler, U., Heinecke, A., Schmidt, T., & Schwarzbach, J. (2003). Different time courses for visual perception and action priming. *Proceedings of the National Academy of Sciences of the United States of America, 100,* 6275-6280.

Williams, M. A., Dang, S., & Kanwisher, N. G. (2007). Only some spatial patterns of fMRI response are read out in task performance. *Nature Neuroscience, 10*(6), 685-686.

Wilson, M., & Daly, M. (2004). Do pretty women inspire men to discount the future? *Proceedings of The Royal Society: Biological Sciences, 271*(Suppl 4), S177-S179.

Woelbert, E., & Goebel, R. (2013). Temptation in economic decision making: effects of immediate reward and reward-cues. *Neuroscience & Neuroeconomics, 2,* 11-19.

Woolgar, A., Golland, P., & Bode, S. (2014). Coping with confounds in multi-voxel analysis: what should we do about reaction time differences? A comment on Todd, Nystrom & Cohen 2013. *Neuroimage, 98,* 73-80.

Woolgar, A., Thompson, R., Bor, D., & Duncan, J. (2011). Multi-voxel coding of stimuli, rules, and responses in human frontoparietal cortex. *Neuroimage, 56,* 744-752.

Zhong, C. B., & DeVoe, S. E. (2010). You are how you eat: fast food and impatience. *Psychological Science, 21,* 619-622.

Appendix I

Bode, S., & Haynes, J. D. (2009). Decoding sequential stages of task preparation in the human brain. *Neuroimage, 45*(2), 606-613.

https://doi.org/10.1016/j.neuroimage.2008.11.031

Appendix II

Bode, S., & Stahl, J. (2014). Predicting errors from patterns of event-related potentials preceding an overt response. *Biological Psychology, 103,* 357-369.

https://doi.org/10.1016/j.biopsycho.2014.10.002

Appendix III

Bode, S., Bogler, C., Soon, C. S., & Haynes, J. D. (2012). The neural encoding of guesses in the human brain. *Neuroimage*, *59*(2), 1924-1931.

https://doi.org/10.1016/j.neuroimage.2011.08.106

Appendix IV

Bode, S., Bogler, C., & Haynes, J. D. (2013). Similar neural mechanisms for guesses and free decisions. *Neuroimage*, *65*(2), 456-465.

https://doi.org/10.1016/j.neuroimage.2012.09.064

Appendix V

Bode, S., Sewell, D. K., Lilburn, S., Forte, J., Smith, P. L., & Stahl, J. (2012). Predicting perceptual decision biases from early brain activity. *The Journal of Neuroscience, 32*(36): 12488-1249.

https://doi.org/10.1523/JNEUROSCI.1708-12.2012

Appendix VI

Murawski, C., Harris, P. G., **Bode, S.**, Domínguez D., J. F., & Egan, G. F. (2012). Led into temptation? Rewarding brand logos bias the neural encoding of economic decisions. *PLoS ONE, 7*(3): e34155.

https://doi.org/10.1371/journal.pone.0034155

Appendix VII

Bode, S., Bennett, D., Stahl, J., & Murawski, C. (2014). Distributed patterns of event-related potentials predict subsequent ratings of abstract stimulus attributes. *PLoS ONE, 9*(10): e109070.

https://doi.org/10.1371/journal.pone.0109070

Appendix VIII

Soon, C. S., He, A. H., **Bode, S.**, & Haynes, J. D. (2013). Predicting free choices for abstract intentions. *Proceedings of the National Academy of Sciences of the United States of America, 110*(15), 6217-6222.

https://doi.org/10.1073/pnas.1212218110

Appendix IX

Bode, S., He, A. H., Soon, C. S., Trampel, R., Turner, R., & Haynes, J. D. (2011). Tracking the unconscious generation of free decisions using ultra-high field fMRI. *PLoS ONE, 6*(6): e21612.

https://doi.org/10.1371/journal.pone.0021612

Appendix X

Bode, S., Murawski, C., Soon, C. S., Bode, P., Stahl, J., & Smith, P. L. (2014). Demystifying "free will": The role of contextual information and evidence accumulation for predictive brain activity. *Neuroscience & Biobehavioral Reviews*, 47, 636-645.

https://doi.org/10.1016/j.neubiorev.2014.10.017