MAX PLANCK INSTITUTE FOR HUMAN COGNITIVE AND BRAIN SCIENCES LEIPZIG

Andrea Reiter: Out of control behaviors? : Investigating mechanisms of behavioral control in alcohol addiction, binge eating disorder, and associated risk factors. Leipzig: Max Planck Institute for Human Cognitive and Brain Sciences, 2016 (MPI Series in Human Cognitive and Brain Sciences; 177)

Out of control behaviors?
Investigating mechanisms of behavioral control in alcohol addiction, binge eating disorder, and associated risk factors

Impressum

Max Planck Institute for Human Cognitive and Brain Sciences, 2016

Druck: Sächsisches Druck- und Verlagshaus Direct World, Dresden

Titelbild: © Andrea Reiter, 2016

ISBN 978-3-941504-63-9

Out of control behaviors?
Investigating mechanisms of behavioral control in alcohol addiction, binge eating disorder, and associated risk factors

Der Fakultät für Biowissenschaften, Pharmazie und Psychologie

der Universität Leipzig

eingereichte

D I S S E R T A T I O N

zur Erlangung des akademischen Grades
Doktor rerum naturalium

Dr. rer. nat.

vorgelegt

von Dipl. Psych. Andrea Maria Franziska Reiter

geboren am 17.05.1986 in Augsburg

Leipzig, den 16.10.2015

Tag der Verteidigung: 02.06.2016

BIBLIOGRAPHIC DETAILS
Andrea Maria Franziska Reiter
Out of control behaviors? Investigating mechanisms of behavioral control in alcohol addiction, binge eating disorder, and associated risk factors
Fakultät für Biowissenschaften, Pharmazie und Psychologie
Universität Leipzig
Dissertation
256 pages, 383 references, 32 figures, 12 tables

Some psychiatric conditions share as a hallmark the loss over behavioral control: for instance, in substance addiction, individuals continue drug consumption, despite negative consequences. One prominent hypothesis regarding the psychological processes that give rise to addiction is a shift from goal-directed towards habitual behavioral control: formerly reinforcing substance consumption is continued, even though it is at odds with current goals.

In this thesis, this hypothesis was investigated in light of two recently developing research paradigms: (1) A *Dimensional Psychiatry* account was employed: First, goal-directed control was investigated in individuals suffering from alcohol dependence. Second, the research question was extended towards patients suffering from binge eating disorder, an addiction-like disorder characterized by the loss of control over eating behavior. Third, it was asked whether populations at risk for addiction are similarily characterized by a shift from goal-directed towards habitual control. (2) Adopting a *Computational Psychiatry* approach, the current work combined computational modeling of behavior with functional Magnetic Resonance Tomography and Electroencephalography. Flexible goal-directed control was revealed to be impaired in both patient groups. By the use of computational modeling, differential behavioral pathways leading to similar impairments could be dissected: alcohol-dependent patients displayed a failure to integrate alternative choice options in their decisions after punishment whereas binge eating patients were characterized by a bias towards explorative choices. Medial prefrontal cortex learning signals promoting flexible decision-making were found reduced in both patient groups. This suggests the medial prefrontal cortex as a transdiagnostic convergence point essential for monitoring behavioral control. Interestingly, in both groups simpler signatures of decision-making were found to be preserved. Regarding risk factors of addiction, findings were qualitatively different from observations in populations suffering from addiction. Whereas none of the risk factors impeded goal-directed control per se, I observed interaction effects between cognition and impulsivity, as well as acute and chronic stress.

In sum, this thesis constitutes a step towards a mechanistic understanding of psychiatric conditions characterized by a failure of behavioral control. The findings motivate a combination of computational approaches to cognitive neuroscience with longitudinal clinical designs. This could enable definition of biologically informed subgroups and guide new treatment and prevention developments, independent of the current symptom-based classification schemes.

Acknowledgements

These were exciting 3 years in which I have learnt more than I could have imagined.
I am very grateful to Florian Schlagenhauf for the excellent opportunity to work on this thesis in his group in a kind, open-minded and laid-back atmosphere. I would also like to thank all the past and present members of the research group *Cognitive and Affective Control of Behavioral Adaptation*.

I thank Erich Schröger for supervision at the University of Leipzig and for reviewing this work.

I am very grateful to the IMPRS NeuroCom, for funding, but also because it gathered together so many great people; I became friends with many of you and it was beautiful 3 years. I thank the coordinators, Katja Kirsche and Antje Niven for their support.

I would like to thank the members of the Leipzig Mind Body Emotion Cohort, and especially Anahit Babayan and Arno Villringer, for welcoming me as an external member.

Recruitment and data acquisition efforts were enormous for this thesis, and I would like to thank the interns, student assistants and MTAs who made it possible: Mandy Jochemko, Robert Kratzer, Antje Kummer, Mathias Kerkemeyer, Lennart Lüttgau, Karoline Hudl, Tatjana Schmidt and Sylvia Stasch. I am also grateful to the physicians at Soteria Clinic Leipzig, for their support during recruitment. I am particularly thankful to all the patients and relatives for their openness to participate in the studies.

I thank Heike Schmidt-Duderstedt for her help with graphics.

Christoph and Tilmann, it was great fun working with you – such a big chance!

I am happy enough to have a whole bunch of wonderful friends who are there for me, in sunny and in stormy times (of which both there were plenty, in the last three years). I name precious Anna-Sophia, Katharina, Martina and Sofie as representatives here because they also supported the stressful final spurt of this thesis by proofreading and layouting. Thank you, Janis, for our PhD-WG home including hot orange and 'Fuchs du hast die Gans gestohlen'.

I thank my family for having provided pretty optimal grounds to develop, and especially my mother Angelika, who has always promoted and encouraged me during this long route of (academic) education.

Dear Lorenz, I have been thinking about this sentence for four weeks now and still haven't come up with anything which would possibly meet the gratitude I feel. That's not so often that I am missing words; let's take it as a sign. Thank you very much!

Contents

I was out in the city
I was out in the rain
I was feeling down hearted
I was drinking again

[…]

I was young
I was foolish
I was angry
I was vain
I was charming
I was lucky
Tell me how have I changed

Now I'm out
Oh out of control
Now I'm out
Oh out of control
Oh help me now
[…]

(Jagger&Richards, The Rolling Stones, 1998)

1 Introduction

Why do we repeat behaviors we know are bad for us? Imagine, it is new years and you know you have gone overboard with turkey, Christmas cookies, fudge and their equals. Your jeans have got significantly too tight and the new-year's resolution is crystal clear: you should really, really lose some weight. Why, for heaven's sake, do you find yourself now sitting in front of the TV, munching a bag of chips? Obviously, your new goal – losing weight – has failed to override behaviors that have been reinforced in the past: a couch-and-chips evening has probably proven to be quite cozy and tasty before.

Conflicts of this kind have sparked great interest in experimental psychology and led to the reasoning that in situations as described above multiple decision-making systems are competing for control over behavior. A key division spans from reflexive or habitual towards reflective or goal-directed control (Balleine & Dickinson, 1998; Dolan & Dayan, 2013). Whereas habitual control is seen as a retrospective strategy, influenced by rewards achieved in the past, goal-directed control involves prospective planning and considers potential future outcomes of the choice options at hand. Decision-making in healthy individuals seems to be driven by a balance between prospective goals and retrospective habits (Daw *et al.*, 2005; Daw *et al.*, 2011); in other words: sometimes you spend your evening going for a run, and have a green smoothie afterwards, sometimes you fall back upon the couch.

A grave pathological variation of the everyday life conflict sketched in the introductory example is addiction, where maladaptive behaviors are upheld in the face of often devastating consequences. One influential hypothesis on the psychological processes that give rise to addiction is a shift from goal-directed towards habitual control. Putatively, this leads to one dominant habitual mode when making decisions (Robbins & Everitt, 1999; Everitt & Robbins, 2005; Redish *et al.*, 2008; Dayan, 2009a). Formerly reinforcing drug consumption is continued, even though it is at odds with current goals.

This thesis is centered on the latter hypothesis of addiction as a bias towards habitual behavioral control. This research question is explored in light of two recently developing paradigms: (1) a

Computational Psychiatry (Maia & Frank, 2011; Montague *et al.*, 2012; Stephan & Mathys, 2014; Wang & Krystal, 2014) approach is adopted. Computational Psychiatry targets psychopathology by using formal, quantitative models of cognitive and brain function. In this vein, the current work combines computational modeling of behavior with neural measurements, aiming to shed light on "addiction as a computational process gone awry" (Redish, 2004). (2) A *Dimensional Psychiatry* account (Buckholtz & Meyer-Lindenberg, 2012; Robbins *et al.*, 2012) is employed. Dimensional Psychiatry investigates mechanisms of psychiatric conditions independent of diagnostic boundaries. In this thesis, the dimensional approach is realized in two ways. First, a transdiagnostic approach is adopted as behavioral control is studied across two different diagnostic entities. Both share as a clinical feature the loss of control over behavior: alcohol addiction as a prototypical and highly prevalent form of substance dependence (Kraus & Augustin, 2001; Grant *et al.*, 2004) and binge eating disorder (i.e., pathological overeating), a newly defined diagnosis. The latter is sometimes referred to as *food addiction* (Smith & Robbins, 2013; Robbins & Clark, 2015) but understudied to date. Second, recognized risk populations and risk factors for addiction are investigated, to explore whether habitual behavioral control not only is a state marker of mental illness, but also extends as a potential vulnerability factor to healthy individuals at risk for addiction.

After briefly introducing methods and the general theoretical framework of this thesis, six empirical studies on behavioral control will be presented. These studies surround the question of inter-individual differences in behavioral control from a clinically motivated, cognitive neuroscience perspective enriched by computational modeling techniques. This will be followed by a more general discussion of results, including limitations, and the attempt to place the presented studies in a broader picture.

2 Principles of cognitive neuroscience

The first section serves as a primer for those principles of cognitive neuroscience that are employed in the current thesis. First, I will present an introduction on models and computational modeling in cognitive science. I will discuss the application of such computational models in the field of cognitive neuroscience before I continue with a general introduction to measuring brain functions via (functional) Magnetic Resonance Tomography (fMRI) and Electroencephalography (EEG).

2.1 Models and computational modeling in cognitive science

The following outlines the approach of cognitive modeling and its implementation in cognitive science. It follows and summarizes recent comprehensive overviews on the topic, which cover this important field in a more detailed way (Lewandowsky & Farrell, 2011; Forstmann & Wagenmakers, 2015).

Models are broadly defined as an abstract structure that itself captures structure in the data (Luce, 1995). They are at the core of scientific practice of any kind as data are not self-explanatory, but require a model to be interpreted. In the case of psychology and the cognitive sciences, this is most commonly realized via verbal models. These are used to describe psychological phenomena qualitatively, in order to then test predictions of these models empirically. Computational (or mathematical) modeling accounts are different: predictions are formalized in a quantitative computational model, which mathematically aims to describe the relationship between inputs (e.g., a stimulus) and output (e.g. subjects' behavior). One prime advantage of quantitative computational models as compared to the commonly used verbal models is a precise implementation of a theoretical framework: when formulating a computational model, the scientist is forced to precisely specify all parts of a theory. A rigid specification of a theory's components and their interaction with each other enhances the efficiency a theory can be communicated. For a detailed discussion how this might foster reproducibility of results, see Farrell and Lewandowsky, 2010. Further, and importantly, this ensures the testability of predictions, as mathematical models generate quantitative predictions,

which can be directly linked to empirical data (compare Figure 2-1). Quantitative models thus render a theory falsifiable (Popper, 1982). In the same vein, computational models advise the researcher on experimental manipulations which are informative (Myung & Pitt, 2009; Cavagnaro *et al.*, 2013; Kim *et al.*, 2014). Finally, mathematical models also offer a particularly stringent and coherent framework for the interpretation of observed data.

Other than purely descriptive or predictive models, many computational models are constructed to be explanatory, that is, meant to formalize how psychological processes operate and how the observed data were generated. In this sense, models are constructed as models of mechanisms that lead to observable behavior. This comes along with the so-called Bonini's paradox: to be explanatory, models necessarily have to be less complex than the process they aim to capture (Norris, 2005). Farell and Lewandowsky summarize this as retaining "the essential features of the system while discarding unnecessary details" - as the model would not contribute to understanding if the model itself cannot be fully understood (Farrell & Lewandowsky, 2015, p. 10). The general principle of cognitive modeling is illustrated in Figure 2-1.

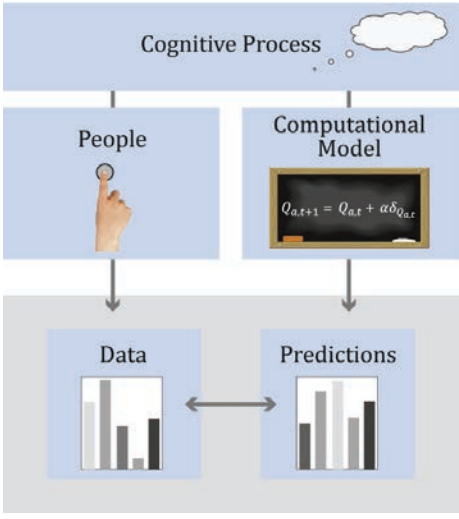


Figure 2-1. General principle of computational modeling in the cognitive sciences. The general aim is to mirror a specific cognitive process (e.g., in this thesis, behavioral control). This is done by constructing an experiment in which participants' behavior (the empirical data) reflects the process of interest as purely as possible (left part of the figure). Computational modeling aims to describe the cognitive process of interest via mathematical models. The core of computational modeling is to link model predictions with empirically observed data. By fitting the model to the empirically observed data, i.e., minimizing the difference between observed and predicted data, individual parameters can be estimated. Parts of the figure are inspired by Lewandowsky & Farrell, 2011.

2.2 Modeling-based cognitive neuroscience[1]

The relationship between experimental cognitive psychology and cognitive neuroscience has been ambivalent. Much of the concerns expressed by cognitive psychologists surround the notion that brain measurements alone are not meaningful in a theoretical sense and that localizing functions in the brain can be irrelevant for understanding how psychological processes operate (Coltheart, 2006; Forstmann *et al.*, 2011; Coltheart, 2013). Some years ago, it was stated that "no functional neuroimaging research to date has yielded data that can be used to distinguish between competing psychological theories" (Coltheart, 2006, p. 423) and neuroimaging research has been repeatedly accused for being the *new phrenology* (Kennedy, 2005; Diener, 2010). These critiques warrant attention, particularly because neuroscientific explanations have been claimed to exert a "seductive allure" (Weisberg *et al.*, 2008, p. 470): for instance, Weisberg and colleagues showed that the mere mention of neuroscience enhanced the credibility of a statement.

Modeling-based cognitive neuroscience promises remedies for these critiques, by leveraging a reciprocal relationship between cognitive modeling and cognitive neuroscience (Friston, 2009; Forstmann *et al.*, 2011): formal and mechanistically informative models of cognition are developed and used to amend cognitive neuroscience. Behavioral data are decomposed by means of computational modeling into multiple latent psychological processes, which are subsequently related to neural data (Figure 2-2). The goal is to derive more precise information on the implementation of the specific process in the brain, also in experimental situations where the manipulation is not process-pure. For instance, in studies of behavioral adaptation, often neural responses to feedback stimuli are the dependent variable of interest. Neural responses to feedback could however represent multiple latent variables: among others, value of feedback, individual sensitivity to feedback, learning or also more unspecific processes like attention. Thus, analyses or interpretation of these results much depend on the experimenters' definitions. Modeling-based analyses on the other hand allow to investigate

[1] To describe an analysis of neural data which is informed via computational models, often-times the term *model-based* analysis is used. In this thesis, the term *modeling-based* is used in order to avoid any confusion with a so-called *model-based* behavioral control strategy (compare section 3.3).

neural activation to feedback in light of one specific cognitive sub-process, e.g. prediction error coding (O'Doherty *et al.*, 2003; O'Doherty *et al.*, 2004). This goes beyond the mere localization of a psychological process in standard fMRI analyses and promises a paradigm shift from the somehow limited focus on localization of function (*where*) to more *how* type of research questions by showing how a certain computational process is indeed represented in the brain (Dolan, 2008). Further, this is a promising route to move from umbrella terms like *Cognitive Control* towards a parcellation of the mechanistic bases (Verbruggen *et al.*, 2014). See Forstmann *et al.*, 2011 as well as Forstmann and Wagenmakers, 2015 for an in-depth discussion of advantages of the modeling-based approach. Of particular importance for the current work is that modeling-based approaches enable to account for inter-individual differences by incorporating these individual differences in the latent cognitive process into the analysis of neural data (e.g., Forstmann *et al.*, 2008). This is especially valuable for patient studies (e.g. Schlagenhauf *et al.*, 2014; for a more detailed outline on the use of modeling-based neuroscience approaches for psychiatry see section 3.6.2). It is noteworthy that not only neuroscience can benefit from implementing computational models in analyses. Vice versa, neural measurements can inform computational models (especially for complex phenomena) as they lead to more dependent measures which help to constrain model testing as compared to behavioral data alone, where sometimes similar behavioral observations can be predicted by models with different assumptions (White & Poldrack, 2013; O'Doherty *et al.*, 2007).

2.2.1 Modeling-based analysis of neural data: general approach

This paragraph outlines general steps which are pursued during a modeling-based analysis of neural data – from finding an appropriate model to fitting the behavioral data to applying the derived parameters to the neural signal (Figure 2-2 for an overview). In the current framework of a general introduction, this needs to be kept on a rather superficial level. Please refer to Lewandowsky and Farrell, 2011 for a thorough introduction on parameter estimation techniques, model comparison and interpretation of modeling results, and to O'Doherty *et al.*,

2007; Daw, 2011 and Mars *et al.*, 2012, for a more detailed description of application to neural, i.e. fMRI and EEG data, specifically in the context of behavioral adaptation.

The first step is to formulate or use a set of quantitative models which describe the observed psychological phenomenon. Typically, these models include fixed and free (i.e., estimated to optimize model-fit to the observed data) parameters that represent psychological constructs or processes. This requires a model describing a hidden state, e.g. a neuronal, perceptual or a learning process. As this latent process is not directly observable in the measured data, the modelled hidden state needs to be transformed into the actually measured response, e.g. choice behavior or reaction times, via another so-called observation model. Next, a-priori simulations are conducted to verify that the model can recover known parameters of behavior. Only if this is assured, the model can be applied to actually observed data in a meaningful way. The models are then fit to the empirical data: free parameters of the models are estimated to optimize the fit of the model to the empirical data. Model fit, or more specifically, model misfit, is quantified by a function of the likelihood L of the data under each model m (the probability that the data is given by the parameters), such as in maximum likelihood estimation or as part of Bayesian techniques for maximum a posteriori estimation (Heathcote *et al.*, 2015, See a more thorough description of model fitting in the empirical part of the thesis, chapter 5-10).

Next, model selection should be used to determine the most parsimonious model which best explains the data (Pitt & Myung, 2002). Often times, this is done by comparing criteria which combine the maximized likelihood with a penalty term for the number of parameters included in the model (to account for the risk of overfitting). One popular example is the Bayesian Information Criterion (Schwarz, 1978):

$$(1)\ -2L_m + k_m \ln(n)$$

where n denotes the sample size, L_m is the maximized (log)-likelihood of the model and k_m is the number of parameters of the model. Essentially, the Bayesian Information Criterion represents one approximation of the so-called model evidence, which refers to integrating out

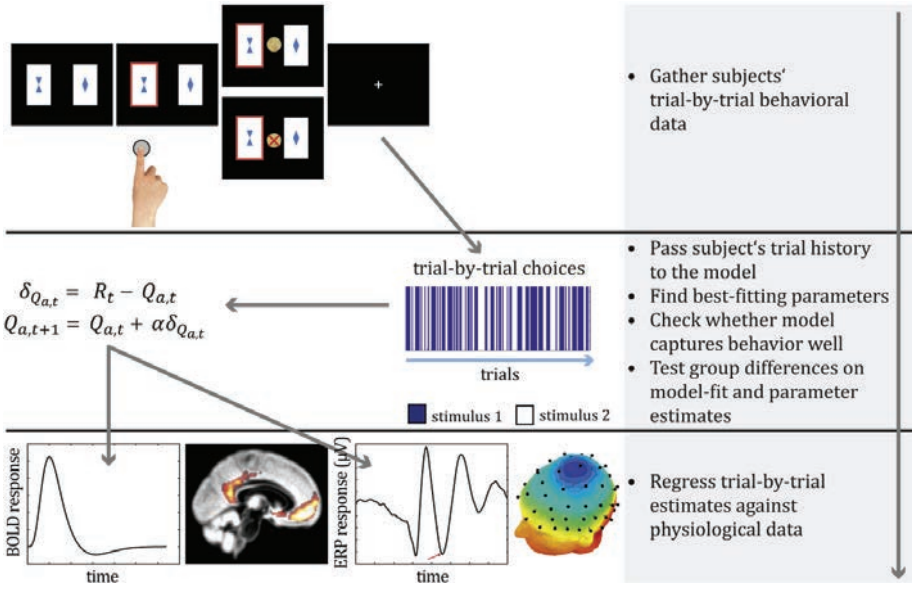


Figure 2-2. The figure illustrates general steps when performing modeling-based analysis of neural data. First, empirical data (e.g. participants' choices in a learning task) is collected. Next, formal models are fit to these data in order to minimize the difference between the predictions of the model and the empirical measure. Following quality checks regarding how well the model captures the actually observed data, best-fitting parameters derived from the model are regressed against physiological data, e.g. the hemodynamic response in fMRI data or event-related potentials in EEG data. Thus, the neural implementation of mechanistically informed parameters can be tested. The illustration is inspired by a figure in O'Doherty et al., 2007.

(or marginalizing) the dependency of the likelihood on the number of parameters (e.g. Stephan *et al.*, 2009). In between-group studies, this is repeated for all experimental groups, to account for the interesting possibility that the groups differ in which model explains their behavior best. Subsequently, model-components are regressed against the neural data on a trial-by trial level: for fMRI, trial-by trial time-series are generated on the basis of best-fitting parameters and these are than convolved with the hemodynamic response (O'Doherty *et al.*, 2004; O'Doherty *et al.*, 2007). In EEG, trial-by-trial values of the model are regressed against trial-by-trial ERP

responses (Mars *et al.*, 2008; Hauser *et al.*, 2014b; Reiter *et al.*, in revision) or trial-by-trial oscillatory activity (Cavanagh *et al.*, 2010). The approach is illustrated in Figure 2-2.

2.2.2 Caveats

2.2.2.1 Goodness of fit vs. generalizability of a model

Additional complexity of a model (i.e., adding more parameters to a model) will improve the fit of the model. The relationship of generalizability and complexity, however, can rather be described as an inverted U-shaped function: while more parameters will lead to an increase of generalizability up to a certain point, adding too many parameters will increase the risk of fitting noise ('overfitting'), and the generalizability of the model decreases (for a comprehensive illustration on the basis of a simulation study, refer to Pitt and Myung, 2002). Model selection thus is key as it aims at identifying the most parsimonious model in the tested model set, which can be further verified using cross-validation techniques. After identifying the model which best accounts for the observed data, an experimental psychology approach would include to ask *why* this is the case and to derive qualitative interpretations on the captured latent processes.

2.2.2.2 Absolute model fit

Despite the obvious importance of model selection, it is to note that model selection is relative in nature and will always come up with one best-fitting model (Lewandowsky & Farell, 2011). Relative model selection however does not guarantee that the preferred model within the tested set of models actually accounts for the empirically observed data. To ensure that the model captures the observed behavior is essential for any meaningful analysis based on the modeling-derived parameters, as well as for its interpretation. Post-hoc simulations on the basis of individual estimated parameters and a comparison of simulated results to the pattern of observed results provides evidence whether the model indeed reproduces the data. For binary choice data, it is possible and important to identify individuals which are not fitted better than chance by the model as their parameter estimates cannot be validly used for further analyses (note however that the identification of a group which is not fit better than chance can be informative in itself, e.g. when this relates to clinical characteristics, Schlagenhauf *et al.*, 2014).

For other data, R^2 and pseudo R^2 can be useful as measures of absolute model fit but it is less easy to define non-arbitrary cut-offs for such measures.

2.3 Methods to study brain function

The following section aims to provide a brief overview over the basic principles of fMRI and EEG. Please refer to Huettel *et al.*, 2004 and Buxton, 2009 for a thorough introduction to (f)MRI, and to Kappenman & Luck, 2010 as well as Pizzagalli, 2007 for an introduction into EEG and ERPs. An introduction into preprocessing and statistical analysis of both measures goes beyond the scope of this short introduction. A detailed description of these procedures as applied in the studies conducted is provided in the empirical section (chapter 5-10). For a comprehensive general overview, please refer to Penny *et al.*, 2011 for (f)MRI and to Nidal and Malik, 2014 for EEG.

2.3.1 Functional and structural magnetic resonance imaging

2.3.1.1 Basic physics of MRI

MRI makes use of the fact that protons in atom nuclei rotate randomly around themselves, which in uneven amounts of protons leads to a magnetic moment (Figure 2-3A; Buxton, 2009). In the absence of an external magnetic field, the spins are oriented in a random fashion, there is no net magnetization. MRI applies strong static magnetic fields to align spins of hydrogen nuclei of the biological tissue with the direction of the magnetic field and to cause them to precess around the field direction (so-called Lamor precession with Lamor frequency, Figure 2-3B). This creates a net magnetization pointing into the direction of the magnetic field (Figure 2-3C; Buxton, 2009). MRI manipulates this by employing radio-frequency (RF-) pulses which match the Lamor frequency: the magnetization vector becomes tilted (Figure 2-3D). MRI measures relaxation time, i.e., the time spins take to align to the strong external magnetic field again as soon as the RF-pulse is turned off (Figure 2-3E, Buxton, 2009): T1 relaxation, as the longitudinal relaxation time, distinguishes tissue types and is thus most commonly employed for structural MRI. T2/T2* refers to the transverse relaxation time, i.e., the decay of phase coherence between the spins. See Figure 2-3 for an illustration. T2* relaxation is exploited in functional fMRI (further discussed in

the next paragraph). Echotime (TE, read-out time after a pulse) and repetition time (TR, time between to pulses) are manipulated by the experimenter to gain optimal T1 (short TR and short TE) and T2 (long TR and long TE) resolutions.

2.3.1.2 fMRI

Mostly in fMRI studies, the so-called blood oxygenation level dependent (BOLD) effect is used to image the functional organization of the brain. BOLD relies on the different magnetic properties of oxygenated and deoxygenated hemoglobin (Hb) in the blood (Ogawa *et al.*, 1990): oxygenated Hb is diamagnetic while deoxygenated Hb is paramagnetic (i.e., has magnetic susceptibility, Pauling & Coryell, 1936). Paramagnetic deoxygenated Hb influences the static magnetic field leading to local distortions. This results in an accelerated decay of transverse magnetization, and a shortened time constant T2* (which is susceptible to field inhomogeneities). This can be captured via T2*-weighted images.

How is this exploited to measure brain activation? Crucially for fMRI, the working brain needs a continuous supply of oxygen, such that oxygen metabolism, blood flow and blood volume are higher in particularly activated brain areas. Thus, in activated areas, the level of deoxygenated blood first rises due to enhanced consumption of oxygen in these areas. This is then compensated by an increase in cerebral blood flow (so-called neurovascular coupling) and thus, the supply with fresh oxygenated hemoglobin. Consequently, the concentration of paramagnetic deoxygenated Hb decreases, local field inhomogeneities become cancelled out, and the transverse relaxation time (and thus the T2*-weighted signal) increases. This change in the MR signal evoked by neuronal activity is referred to as hemodynamic response (see Figure 2-3F). Peak BOLD activity evolves at 5-10s after stimulus presentation with a post-stimulus undershoot of up to 20s, thus returns to baseline in 12 to 30s. BOLD signal can be predicted by neurophysiologically recorded local field potentials and the relationship is nonlinear. BOLD has been shown to mirror predominantly a neuronal populations' input and intrinsic processing rather than output and measures primarily postsynaptic activity (Logothetis *et al.*, 2001; Logothetis, 2007).

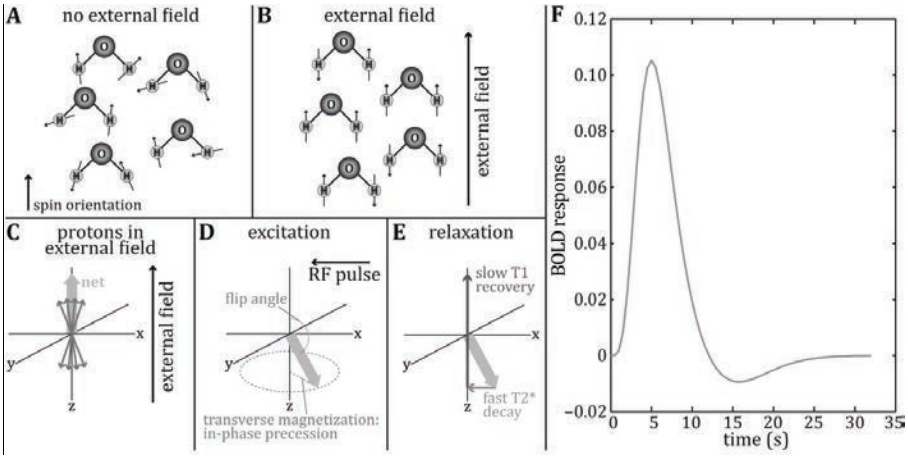


Figure 2-3. MRI exploits magnetic properties of hydrogen atoms (H) in water molecules (H_2O) of biological tissue. A) In the absence of a magnetic field, the spins are oriented in a random fashion. There is no net magnetization. B) As soon as the strong external magnetic field is applied, the spins of hydrogen nuclei are forced to align with the direction of the magnetic field. C) The net magnetization evoked by the application of an external magnetic field points into the direction of the magnetic field. D) By employing a RF-pulse, the magnetization vector becomes tilted. The pulse at the Lamor frequency causes in-phase precession, which leads to transverse magnetization. E) As soon as the RF-pulse is turned off, the spins align to the strong external magnetic field again. These so-called relaxation times are registered by MRI: T1 relaxation describes the longitudinal relaxation time and T2/T2* refers to the transverse relaxation time, that is, the decay of phase coherence between the spins. F) Hemodynamic response function of the BOLD signal. Plotted is the rise to peak after 6 seconds, the undershoot (delay relative to onset 16 seconds) and the return to baseline (total length: 32s). Plotted with the function spm_hrf contained in spm 8 http://www.fil.ion.ucl.ac.uk/spm/). Parts of the figure are inspired by a figure in Smittenaar, 2015.

In a nutshell, BOLD fMRI assesses changes in the quantity of deoxygenated Hb in a voxel (smallest observable element at a certain anatomical location) over time. It builds on the fact that in so-called *activated* brain regions, blood flow increases, which reduces the quantity of deoxygenated Hb in these areas. Hence, BOLD is an indirect measure of neuronal activity. Research utilizing task-based fMRI exploits this indirect measure by locking changes in BOLD to internal or external events of the participants.

2.3.1.3 Advantages and limitations of fMRI

Limitations. The low temporal resolution (several seconds) of fMRI precludes the investigation of time-sensitive topics. fMRI measurements take place under rather unnatural conditions: participants are lying in the scanner and noise produced by the scanner is high, which has been shown to significantly influence behavioral performance (van Maanen *et al.*, 2015) and hampers external validity. MRI is costly to achieve and maintain, which might restrict the resources to acquire large sample-sizes. In the same vein, for the participation in an MRI study, several exclusion criteria hold (e.g., claustrophobia or metal substances in the body), such that it is less feasible for special populations such as children or psychiatric patients.

Advantages. MRI is a non-invasive technique, operating without radiation (compared to Position Emissions Tomography [PET] and Computer Tomography). MRI allows imaging of the whole brain, including cortical and subcortical structures with a relatively high spatial resolution (< 3mm). These advantages have rendered fMRI the predominant technique in cognitive neuroscience to date (Friston, 2009).

2.3.2 Electroencephalography (EEG) and event-related potentials (ERPs)

2.3.2.1 Basic principle: linking voltage changes to psychological processes

ERPs are defined as voltage changes time-locked to an internal or external event (e.g., preparation of a motor response or onset of a stimulus, e.g. in the current work: onset of a feedback stimulus; Kappenman & Luck, 2010). ERPs, defined as neural activity reflecting a particular psychological process, are used to unravel the underlying neural mechanisms of this particular process.

2.3.2.2 Neurophysiological basis of ERP recordings

ERPs are scalp-recorded non-invasively by EEG. These voltage fluctuations are suggested to mirror the summation of excitatory and inhibitory post-synaptic potentials. Post-synaptic potentials occur at the cortical pyramidal neurons as a response to opening or closing ion channels, which are stimulated by neurotransmitter binding at postsynaptic receptors. If such activity occurs simultaneously in a sufficiently large cluster (i.e., tens of thousands) of spatially

aligned pyramidal neurons, it forms an equivalent current dipole which can be measured by EEG (Kappenman & Luck, 2012, but see e.g. Klimesch *et al.*, 2004; Klimesch *et al.*, 2007 for an alternative account on how ERPs emerge). Dendritic trunks of pyramidal cortical neurons are coherently oriented, that is, aligned in parallel to each other and perpendicular to the cortex, which is prerequisite for the summation of the potentials, and thus also for their provability via EEG. Hence, signals recorded by EEG predominantly reflect cortical activity, although subcortical contributions are discussed (for an interesting, controversially discussed example in the current theoretical framework of feedback-related potentials, please refer to Cohen *et al.*, 2011a; Foti *et al.*, 2011b; a).

Among ongoing raw EEG activity, ERPs are comparably small. To overcome this issue, the most common approach – next to appropriate filtering – has been to first identify the stimulus-dependent amplitude and latency, e.g., of the peak of activation within a predefined time window on a trial-by-trial basis, and then to average these trial-by-trial measures as a function of the experimental condition (*Grand Averages*, Makeig & Onton, 2012). This isolates stimulus-driven activity, which should rather be similar over trials under the same experimental condition (but see McGee *et al.*, 2001), from spontaneous EEG activity, unrelated to the time-locked events of interest (Makeig & Onton, 2012). This *Grand Average* technique indeed has the advantage of improving the signal-to-noise ratio, but is suboptimal when it comes to studying dynamic processes which evolve over time, such as Reinforcement Learning (RL) and behavioral adaptation: by averaging over several trials, information on the neural implementation of the psychological process' trial-by-trial dynamics is lost. Finally, a mean (or an Grand Average) constitutes one statistical measure of a distribution, which, if reported detached from other statistics, can be opaque or even misleading, especially when it tries to summarize distinct time serial activities of several areas characterized by spatial and temporal trial-by-trial variability (Makeig & Onton, 2012). The current work tries to overcome this by a modeling-based trial-by-trial analysis of ERPs.

2.3.2.3 The Feedback-Related Negativity as an ERP correlate of behavioral control processes

An event-related potential which has been frequently linked to reward-based learning and behavioral control is the Feedback-Related Negativity (FRN). The FRN peaks 200-350ms after the presentation of an outcome stimulus at fronto-central electrodes (Miltner *et al.*, 1997). In several previous studies, the anterior cingulate cortex was suggested as its neural generator (Miltner *et al.*, 1997; Gehring & Willoughby, 2002; Hewig *et al.*, 2007). The FRN is evoked by outcome stimuli as well as stimuli that predict outcome. It has frequently been hypothesized to encode a reward prediction error (Holroyd & Coles, 2002; Chase *et al.*, 2011; Walsh & Anderson, 2012). Importantly, the latter makes it a candidate ERP for studying behavioral control processes. For a comprehensive review on the FRN, please refer to Holroyd & Coles, 2002 and Walsh & Anderson, 2012.

2.3.2.4 Advantages and limitations of EEG/ERP

Limitations. In terms of limitations of EEG/ERP, most commonly brought forward is the restricted spatial resolution of EEG as the electrical field spreads out through the brain and the high resistance of the skull leads to further smearing of the spatial distribution, which is even aggravated when measuring signals from deeper brain regions (Luck, 2012). As detailed above, activity recorded by EEG reliably only measures cortical activity. This constitutes a further drawback – particularly for the research questions at hand concerning behavioral adaptation, for which deeper brain structures like striatum play a key role (Kelley, 2004; O'Doherty *et al.*, 2004; Daw *et al.*, 2005).

Advantages. EEG shows precise temporal resolution, reflecting ongoing activity with very little delay. This renders EEG an excellent tool for investigating research questions regarding the timing of psychological processes. Further, and this makes the technique valuable in the context of patient studies, EEG is a feasible technique as compared to brain imaging methods like fMRI or PET: procurement costs are comparably low, exclusion criteria are minimal, almost no side effects are reported and tolerance of the laboratory conditions is comparably high. Newest

developments promise portability and even more flexibility in various contexts (de Lissa *et al.*, 2015).

3 Theoretical framework and development of research questions

3.1 Goal-directed vs. habitual modes of behavioral control

Individuals are constantly faced with a complex and dynamically changing environment, which requires them to continuously evaluate and select among many decision options and their potential consequences – and to shift gears towards alternative options as soon as environmental conditions have changed. It is well accepted that to master this taxing challenge, more than one strategy is available: the key distinction between two classes of instrumental behavior, today commonly referred to as *goal-directed* and *habitual*, has a long-standing history – involving controversial discussions during the mid 20th century – in experimental psychology (for an overview see Dolan & Dayan, 2013). On the one hand, behaviorist stimulus-response based theories date back to Thorndike's *Law of Effect*, which was derived from the observation that hungry cats in a puzzle box would, following a trial-and-error, stimulus-response (S-R) principle, repeat rewarded behaviors in the future and give up those that lead to undesired outcomes (Thorndike, 1911). On the other hand, Tolman, as a counter-proposal to the then-popular behaviorist theories, promoted the idea of a cognitive map, originally to describe rats' capability of latent (i.e., without direct reinforcement) spatial learning in mazes (Tolman, 1948; Jensen, 2006). Over the subsequent decades, in the non-spatial domain, the former control mechanism which is based on S-R associations, has been reformulated as *habitual* behavior, whereas the latter control mechanism has been operationalized as *goal-directed* behavior (Dickinson, 1985; Balleine & Dickinson, 1998; Dickinson & Balleine, 2002; Dolan & Dayan, 2013).

Habitual control relies on values stamped-in by past experience during a trial-and-error process. In situations of a stable environment and where practice is assured, these values map the actual reward-value in an appropriate manner. However, as it relies on a trial-and-error type of information accumulation, habitual control is slow in adapting outcome values. In dynamically changing environments, where rapid changes of reward contingencies frequently happen, these values can thus be disparate from the current outcome value of a decision. Thus, this automated

strategy might, curiously enough, lead to an agent choosing an action although it does not desire this action's outcome (anymore). This automated type of control however has the advantage of saving cognitive resources (Balleine & Dickinson, 1998; Rangel *et al.*, 2008; Dayan, 2009b; Dolan & Dayan, 2013).

In contrast, goal-directed control is driven by knowledge of action-outcome contingencies, including the associations of multistep action sequences to outcomes. This enables the agent to evaluate the outcome value of all decision options at hand in a context-dependent way and to flexibly adapt decisions when environmental contingencies have changed. Ideally, this flexible strategy should lead to an agent choosing an action that evokes a presently desired outcome. With respect to cognitive resources, goal-directed control is more costly (Balleine & Dickinson, 1998; Rangel *et al.*, 2008; Dayan, 2009b; Dolan & Dayan, 2013). Thus, goal-directed control is efficient in new environmental situations, when behavior needs to be adapted; as soon as the environment is stable, the agent will ideally employ habitual strategies, as they are computationally less demanding.

3.2 Devaluation paradigms: behavioral and neural correlates of habitual and goal-directed control

Traditionally, and firstly in animal models, devaluation paradigms have been a prominent paradigm to assess goal-directed vs. habitual behavioral control: an animal is trained to press a lever to receive food. This type of food is subsequently devalued (e.g., by feeding the animal to satiety). Goal-directed vs. habitual control is assessed in a subsequent extinction phase, without delivery of feedback. The experimenter makes use of different predictions of the two systems regarding behavior during extinction: goal-directed control can rely on a representation of the outcome when pressing the lever (i.e. representation of an action-outcome association) and would therefore refrain from pressing the lever, as the outcome is no longer desired. In contrast, habitual control, driven by an automated stimulus-response representation, would continue pressing the lever as it used to be reinforced – despite a no longer desired outcome. These studies provided evidence in rodents and humans that actions can indeed be controlled by a

goal-directed or a habitual mode (Adams & Dickinson, 1981). The influence of devaluation, and thus, the degree of control exerted by the goal-directed system, depends on the amount of training before devaluation (Adams, 1981): the longer the training, the more the animal becomes insensitive to devaluation. Over the last decade, this paradigm was employed to unravel the neural correlates of habitual and goal-directed behavior: regarding habit formation, in lesion studies, rats were found to remain goal-directed after selective lesion to the dorsolateral striatum, implying its important role for habit formation (Featherstone & McDonald, 2004; Yin *et al.*, 2004; Balleine & O'Doherty, 2010). Using a devaluation paradigm combined with fMRI in human participants, overtraining was shown to lead to increased activation in the lateral striatum (Tricomi *et al.*, 2009). Employing devaluation tasks in combination with lesion studies in rodents to identify the neural correlates of goal-directed control, rats' prelimbic cortex and dorsomedial striatum were implicated in goal-directed control. The prelimbic cortex was rather deemed responsible for the initial acquisition phase, whereas the dorsomedial striatum seems to be important for both, acquisition and expression of goal-directed behavior (Balleine & Dickinson, 1998; Corbit & Balleine, 2003; Ostlund & Balleine, 2005; Yin *et al.*, 2005; Balleine & O'Doherty, 2010). In human participants investigated during a selective outcome paradigm via fMRI, the ventromedial prefrontal cortex (vmPFC) showed a response profile consistent with the goal-directed system: activation in the vmPFC towards the devalued action (compared to the valued action) was reduced when comparing the pre- to post devaluation phase of the experiment (Valentin *et al.*, 2007).

3.3 Computational formulations: model-free vs. model-based modes of control

Recently, computational accounts, inspired by artificial intelligence and machine learning (Bellman, 1957; Sutton & Barto, 1998), have amended these theories by formalizing habitual and goal-directed control with the help of algorithms derived from Reinforcement Learning (RL, Doya *et al.*, 2002; Daw *et al.*, 2005). The computational reformulation builds on the same ideas as sketched in the last paragraph, but uses mathematical models to describe and test these. The computational approach goes beyond the before-described verbal theory of goal-directed and

habitual control (Anselme *et al.*, 2013). This thesis builds on these computational re-formulations, and their principles are summarized in the following. I present the computational architecture in rather limited mathematical detail, roughly following the description and notation of Balleine *et al.*, 2008b, Daw and O'Doherty, 2013 and Huys *et al.*, 2014. The interested reader is referred to this work and the empirical part of this thesis (chapter 5-10) for more details.

In the computational framework, goal-directed control arises from so-called *model-based* computations: at one point in time t, the environment is in a specific state s_t and an agent chooses an action a_t in order to get a reward r_t. In a psychological experiment, states s_t could for example be a specific set of stimuli and action a_t would refer to the agent choosing one of the stimuli. The combination of s_t and a_t at one point in time determines the subsequent state s' based on a transition function T. T defines the probability distribution over this new state s':

$$(2)\quad T(s, a, s') = P(s_t + 1 = s' | s_t = s, a_t = a)$$

The transition function gives the probability of s' given the previous state-action combination. In an instrumental learning experiment, this would reflect the probability (or contingency) by which one choice leads to an event like monetary gain. When playing a game, knowing T would correspond to knowing the rules of the game.

The reward function is the average reward as a function of state

$$(3)\quad R(s) = E[r_t | s_t = s]$$

When playing a game, this would reflect the goal of a game. Importantly, a model of the world consists of the combination of transition function T and reward function R.

Stepping back to the previous section which introduced key characteristics of *goal-directed* behavior, it is particularly noteworthy that T formally incorporates response-outcome associations which is also the key criterion of goal-directed behavior (compare section 3.1) Considering the devaluation paradigms described above, a model-based agent embodying T and R would withdraw from choosing the devalued option and thus show goal-directed behavior.

In real life environments, an agent is often faced with a multitude of actions and states leading to rewards, which themselves lead again to different states associated with rewards. How does a model-based agent master such sequential decision processes? Generally speaking, model-based control would mean to sequentially evaluate all action sequences, according to T and R. Formally speaking, the agent aims at choosing actions in a way that maximizes future reward, summed over all possible future states, while accounting for the probabilities of the state transitions in his computations. This requires evaluating all branches of a decision tree: mapping potential state trajectories as consequences of potential actions. Thus, the value of taking action a in state s is evaluated by summing up future rewards R depending on the sequence of states, s, s_{t+1}, s_{t+2} and averaging these over different trajectories, given the state transition probability function T. This can be formalized in the state-action-value function:

$$(4)\ Q(s,a) = E[r_t + r_{t+1} + r_{t+2} + \cdots | s_t = s, a_t = a]$$

Where $E[\cdot]$ is the average expectation of an agent, averaging over the probabilities given by T, the state transitions. The state-action-value-function thus sums up a particular reward that follows a particular action given a particular state.

This function is often rewritten to account for the fact that the $r_t + r_{t+1} + r_{t+2}$ partition of the formula above equals the value of the respective successor state Q_{t-1} / Q_t / Q_{t+1}. Thus, the long-term value Q can be defined in terms of itself:

$$(5)\ Q(s,a) = R(s) + \sum_{s'} T(s,a,s') \max_{a'} [Q(s',a')]$$

By $\max_{a'} [Q(s',a')]$, we assume that the agent takes the best (i.e. the one with maximum value), action available.

In brief, equation 5 provides the key computations the model-based system works out to come to a decision value. An essential characteristic of model-based control is that it takes into account both T and R. By building a decision-tree based on the world model of the environment, model-based control represents all possible actions and simulates all potential outcomes online. This enables prospective planning and thus flexible behavioral adaptation. It is sometimes

compared with a chess-player who takes into account a sequence of multiple states (board positions) and actions (moves and countermoves, Balleine *et al.*, 2008a). From equations 4 and 5, it becomes clear that model-based control, i.e., computing the whole decision tree, is computationally very laborious.

In contrast, the computational counterpart to habitual control, namely *model-free* control substitutes these (costly) computations by prediction errors based on stored or so-called *cached* values, e.g. the *temporal difference prediction error* (Sutton & Barto, 1998), which approximates current value estimates using previously learnt estimates. One version of model-free learning is *Q-learning* (Watkins & Dayan, 1992). Here, the new value is updated based on the old one:

$$(6)\ Q(s_{t+1}, a_{t+1}) = Q(s_t, a_t) + \alpha\delta_t$$

where δ_t is the reward prediction error in trial t, defined as follows:

$$(7)\ \delta_t = r_t + \alpha \max_{a'}[Q(s_{t+1}, a) - Q(s_t, a_t)]$$

δ indicates the discrepancy between the old prediction and the new outcome (i.e. the *temporal difference*, TD). Note that this algorithm does not represent the world model (like above defined as the combination of the transition function T with the reward function R). It becomes clear that this is computationally less demanding as, e.g., solving equation (5). However, a drawback of this strategy is that new values can only be acquired via direct new experience; the model-free learner only knows that a hot stove should not be touched after he has actually tried it out, experienced pain as a form of punishment, and updated the value accordingly. Hence, estimates are retrospective as derived from past utilities and might not appropriately map on the value of the actual outcome.

The trade-off between the two systems can thus be seen as a trade-off between computational and statistical efficiency (Dayan, 2009b; Simon & Daw, 2011; Dolan & Dayan, 2013; Gershman *et al.*, 2014): while predictions of the model-based system are more flexible and more precise, a computational device like the brain is restricted in resources and will not possibly be able to scour the entire decision-tree at any point in time. For this thesis, the trade-

off is interesting: it is conceivable that inter-individual differences exist in how individuals arrive at the compromise between statistical and computational efficiency.

The computational formulation of the longstanding psychological ideas of behavioral control enables us to investigate behavioral control with all benefits of a modeling-based analysis as sketched in the General Principles section (section 2.1 and 2.2). This includes the development of experimental paradigms specifically tailored to measure the dichotomy between model-free and model-based control. Further, it opens the doors to testing inter-individual differences on mechanistically informed parameters and their implementation in the brain. In the next two sections, I will give an overview on how this has stimulated recent cognitive neuroscientific investigations (section 3.4), and finally, first attempts to translate this to clinical populations (section 3.5). Regarding terminology, whenever a specific analysis is influenced by the computational approach, the terms *model-free* for habitual control and *model-based* for goal-directed control will be used. See Figure 3-1 for an illustration of the two control systems.

3.4 Behavioral and neural correlates of model-free and model-based control

3.4.1 Paradigms to test model-free and model-based control

The reformulation of habitual and goal-directed modes as the computationally informed concepts of model-free and model-based control came along with the development of several elegant experimental and computational approaches to test both modes of control in a computational modeling framework and to gain insight in their neural implementation. These approaches can be roughly divided into two types of instrumental learning paradigms: counterfactual serial reversal learning and sequential decision-making (Doll *et al.*, 2012).

Counterfactual learning tasks implement an implicit or explicit counterfactual task structure (Hampton *et al.*, 2006; Bromberg-Martin *et al.*, 2010; Wunderlich *et al.*, 2011; Wimmer *et al.*, 2012). An agent can derive hypothetical information from abstract inference on this task structure, i.e. infer *what might have happened* if it had taken another course of action – although it has not actually experienced the outcome. Probabilistic Reversal Learning tasks with anti-

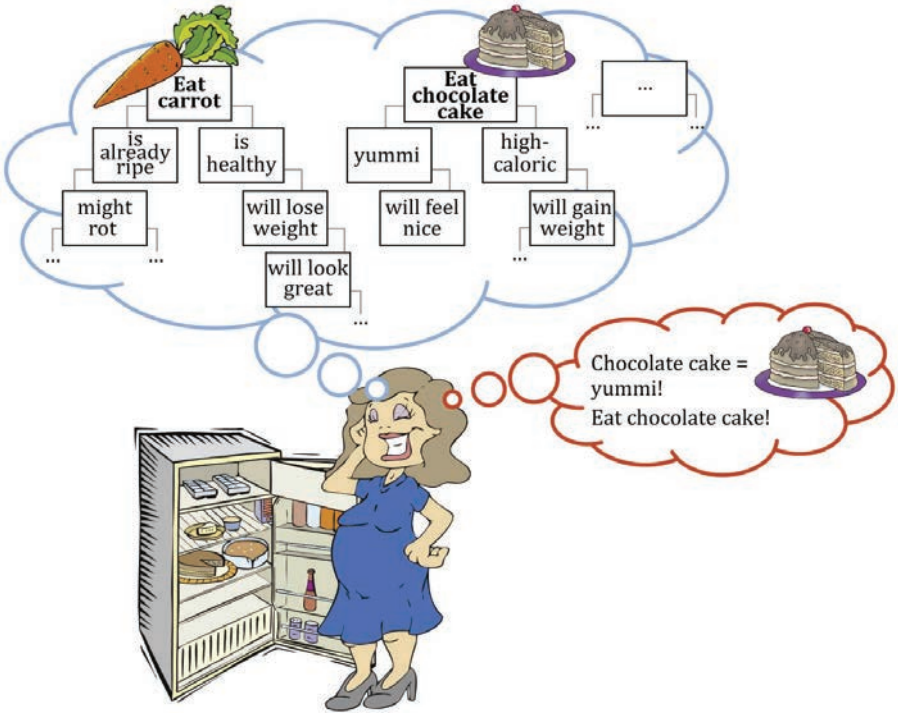


Figure 3-1. Model-Based and Model-Free Decision-Making in the context of deciding what to eat. Model-based computation (left bubble) uses a forward model of the environmental conditions. This model involves characteristics of the environment, including different transitions (e.g., high-caloric → gain weight) and different outcomes (e.g., look great, feel nice, gain weight). Model-based control searches the mental map to estimate the long-run value of each action (eating cake vs. carrot vs. any other type of food vs. eating nothing…) at the current state with respect to the expected reward. The action with the highest overall reward is then chosen. Model-free control (right bubble) does not search through such a model. Instead, model-free control draws on cached values, values which store past experience. Experience might have shown that eating a chocolate cake is an extremely rewarding action. Under the influence of model-free control, the agent will thus decide to eat the cake, even though this action might be at odds with current goals, e.g., losing weight (Dayan & Niv, 2008).

correlated reward distributions (i.e, whenever one stimulus is good, the other one is bad) are frequently used to operationalize counterfactual learning (Doll *et al.*, 2012). Sequential decision-making describes types of tasks in which the agent can make use of a transition structure in order to solve multi-step decisions (e.g. situations in which decision options at a start lead to a specific subsequent state, which requires the next decision from which a reward can be obtained, Glascher *et al.*, 2010; van der Meer *et al.*, 2010; Daw *et al.*, 2011; Simon & Daw, 2011).

As these are the experimental paradigms which are applied in the empirical studies of the current thesis, exemplary previous results on behavioral control and its neural substrates derived from these types of tasks will be reviewed in the following section. For detailed description of tasks and computational modeling, including algorithms, please refer to the methods section in the chapters reporting the empirical studies (chapters 5-10).

3.4.2 Previous cognitive neuroscience findings in healthy individuals

The first influential human computational imaging studies that used RL-algorithms exclusively concentrated on model-free TD prediction errors and their representation in the brain (O'Doherty *et al.*, 2003; O'Doherty *et al.*, 2004; Dolan & Dayan, 2013), without dissociating these from model-based learning signals. These studies (and a wealth of replications) have shown coding of model-free prediction errors in striatal BOLD activity (Balleine & O'Doherty, 2010). This activation has for long been theoretically linked to phasic releases of dopamine carrying a model-free prediction error (Schultz *et al.*, 1997; Schultz, 2013). In line with the earlier animal studies touched upon before, there is evidence that it is rather the putamen, and not the caudate, that is involved with model-free predictions (Wunderlich *et al.*, 2012a; Lee *et al.*, 2014).

In recent years, attention has also shifted towards model-based learning signatures in the brain. Using modeling-based fMRI combined with a counterfactual reversal learning task as introduced above, Hampton and colleagues revealed that BOLD activity in the ventromedial prefrontal cortex was consistent with a model incorporating an abstract model of the counterfactual task-structure than by model-free RL, suggesting the vmPFC as a key region for model-based evaluation (Hampton *et al.*, 2006). Sequential decision-making tasks were applied

to dissociate neural signatures of model-free vs. model-based prediction errors. In an echo of the latent learning paradigm used by Tolman (Tolman, 1948), Gläscher and colleagues investigated the acquisition of a mental model as a key characteristic of model-based control using fMRI. In a latent learning period (i.e., without delivery of reward), participants acquired the transition structure of a two-step sequential decision problem, showing model-based state prediction errors in lateral prefrontal cortex, whereas model-free prediction-errors were observed again in ventral striatum (Glascher *et al.*, 2010). In a subsequent phase of the experiment where rewards were delivered, subjects used the latently acquired model of the task to gain these rewards. In a recent study using sequential decision-making, Doll and colleagues have tested another key aspect of model-based control: the forward-planning nature, or prospective evaluation, of model-based decision-making. The authors indeed could show that BOLD-activation in brain regions that represented prospectively chosen categories correlated with the degree of model-based behavior and to other neural signatures of model-based choice in the dorsomedial prefrontal cortex and the fronto-polar cortex (Doll *et al.*, 2015). To summarize, up to this point, a clear picture seems to emerge: model-free, habitual learning, across species, is best explained by signaling in striatum, most specifically putamen (or in rodents the dorsolateral, but not the dorsomedial striatum), whereas model-based, goal-directed learning can be localized in prefrontal regions, namely medial and lateral prefrontal cortex.

In an attempt to probe this putatively clear-cut anatomical distinction, Daw and colleagues used a sequential decision making variant designed to explicitly dissociate the two systems (Daw *et al*., 2011). This task makes qualitatively different predictions for a model-free vs. model-based behavior which are not necessarily mutually exclusive. This enabled to measure a balance between model-free and model-based control. An interesting (and up to now several times replicated, see Wunderlich *et al.*, 2012b; Eppinger *et al.*, 2013; Smittenaar *et al.*, 2013; Deserno *et al.*, 2015b), finding was that individuals use a hybrid model: their behavior showed evidence for elements of both, model-free and model-based systems. It came as a surprise that TD-prediction errors modified by the model-based system correlated with (until then)

supposedly model-free BOLD activation in the ventral striatum (Daw *et al.*, 2011; for a replication see Deserno *et al.*, 2015b). This brain-behavior correlation was associated with subjects' degree of model-based behavior (Daw *et al.*, 2011). Similarly, both model-free and model-based contributions could be detected in the medial prefrontal cortex. These findings led other authors to conclude a "ubiquity of model based reinforcement-learning" (Doll *et al.*, 2012, p. 1075), rather than a clear-cut distinction of both systems. Thus, the coherent computational framework allows testing the hypothesis that both types of control represent two ends of a continuum, rather than two distinct modes of control. It is from this strand of work that we now can build on accumulating evidence for a continuous spectrum of control with decision-making in healthy individuals being influenced by both systems (Daw *et al.*, 2011; Doll *et al.*, 2012). Some authors have argued that both control systems compete for behavioral control (Lee *et al.*, 2014). On the other hand, some recent evidence suggests that both systems of behavioral control work together in a cooperative way (e.g., with the model-based system teaching the model-free system) to manage behavioral control challenges in multiple environmental situations (Keramati *et al.*, 2011; Gershman *et al.*, 2014).

The proposal of a continuum has interesting implications for an inter-individual perspective as adopted in this thesis. Important open questions arise: can the balance between model-free and model-based control be shifted towards one end of the continuum? What are situational or trait factors associated with a bias towards one or the other pole? Are some individuals more located towards one or the other end of the continuum? Can certain psychiatric disorders be characterized as extreme examples for an imbalance of the two systems?

3.5 Shifting the balance: previous research on inter-individual differences in behavioral control

3.5.1 Behavioral control going awry – previous findings in addictive disorders

Not only in clinical psychology and in cognitive neuroscience, also amongst the general public, addiction has for a long time been conceptualized as an inability to learn from one's errors, or stigmatized as *failure of the will* (Heinz & Batra, 2003). Indeed, the inability to stop substance

consume in the face of negative consequences is at the core of current diagnostic criteria for substance addiction (American Psychiatric Association, 2013) and also for related disorders like binge eating disorder.

In resonance with these clinical observations, an earlier strand of research in cognitive neuroscience has demonstrated blunted neural responses to performance errors and reduced error awareness in addiction (Garavan & Stout, 2005; Paulus *et al.*, 2008; Goldstein *et al.*, 2009). Further, deficits in cognitive flexibility and associated neural correlates in addiction have been frequently reported in animal (Jentsch *et al.*, 2002; Schoenbaum & Shaham, 2008) as well as human studies (Bechara & Damasio, 2002; Ersche *et al.*, 2008; Park *et al.*, 2010; Ersche *et al.*, 2011; Deserno *et al.*, 2015a). The hypothesis of a shift from goal-directed towards habitual behavior in addiction has become particularly prominent over the last decade (Everitt & Robbins, 2005; Dayan, 2009a; Lucantonio *et al.*, 2012). Considering the clinical picture, this hypothesis indeed seems tantalizing: addicted patients in fact make very poor choices; instead of striving for reasonable goals like maintaining their job or a social network, they habitually continue substance consumption, even though it might not even be pleasant anymore, just because it used to be rewarding sometime in the past. This hypothesis has received first empirical confirmation using devaluation tasks in rodents (Schoenbaum & Setlow, 2005) and variations of such in human beings (Sjoerds *et al.*, 2013). On the theoretical level, addiction has however also for long been described from a computational perspective (Redish, 2004; Redish *et al.*, 2008; Dayan, 2009a), offering mechanistic hypotheses. Addiction was conceptualized as a "computational process gone awry" (Redish, 2004, p. 1944) resulting in a bias towards the model-free system. Until very recently, empirical computational neuroscience investigations to test these hypotheses in addictive disorders have been surprisingly scarce. Regarding model-free prediction errors, Park and colleagues could not find evidence for altered neural prediction error coding in alcohol-dependent patients - despite a clearly observed behavioral learning deficit (Park *et al.*, 2010, but also note Tanabe *et al.*, 2013, who found reduced prediction error coding in substance-dependent participants). This discrepancy might, in light of the previously

discussed continuum hypothesis of behavioral control, invite speculations about a deficit in the model-based system in alcohol dependence (which was not investigated in the latter study) while the model-free system is preserved. In an overlapping sample with Park and colleagues, Deserno and colleagues reported that ventral striatal coding of model-free prediction errors correlated inversely with craving in patients (Deserno *et al.*, 2015a), confirming the clinical relevance of model-free prediction error coding. Only very recently, two behavioral studies have explored addiction-related alterations in the balance between the model-free and the model-based system in humans. Applying a sequential decision-making task in alcohol-dependent patients, Sebold and colleagues found group differences in model-based behavioral control, specifically after negative feedback, which were attenuated when adjusting for cognitive functioning (Sebold *et al.*, 2014). Using sequential decision making combined with computational modeling in a transdiagnostic study, Voon and colleagues did not observe group differences in model-based decision-making between healthy controls and alcohol-dependent patients who had remained abstinent for 17 weeks on average (Voon *et al.*, 2015). However, a correlation of model-based control and duration of abstinence was observed. This finding could point towards the role of model-based control in successful abstinence. Interestingly, the same study showed changes in the model-based system for other diseases within the addictive spectrum: stimulants dependence was associated with a reduced degree of model-based control. Further, binge eating disorder was related to less model-based behavior. Thus, it appears as an interesting question whether different diagnostic entities, which share as a common ground a loss of control over certain behaviors, are reduced in model-based control or biased towards model-free control. The underlying common and differential cognitive and neural mechanisms have however as yet not been investigated.

To this end, in the empirical part, I present a study on behavioral adaptation in alcohol-addicted patients as a prototypical substance addiction. In a further study, I adopted the same study design to a group of binge eating patients. Binge eating disorder is a newly defined disorder, first described in DSM-5. The clinical picture is characterized by recurrent episodes of

overeating. During these binge eating episodes, patients have the feeling that they cannot stop eating, in the face of negative consequences like feelings of guilt and shame and the high risk for obesity. According to DSM-5, it is classified as an eating disorder. It however shares with substance addiction as a common clinical feature the experienced loss of control over behavior despite negative consequences and is by some authors referred to as *food addiction* (Smith & Robbins, 2013; Robbins & Clark, 2015). Compared to alcohol addiction, binge eating disorder involves loss of control regarding a natural reinforcer (food) which lacks neurotoxic effects. The latter is advantageous when it comes to measuring neural correlates.

3.5.2 Failure of abstract inference on alternative choices in addictive disorders

What exactly is impaired about model-based control in addiction? On the theoretical level, disturbed mechanisms of inference in addiction have been discussed (Huys *et al.*, 2015b). The rationale is that inference on higher-order environmental structure, e.g., on the interdependencies of values is important when constructing and thus also using a model of the world. In fact, generalizing from explicitly experienced reward relationships to those that were never directly learned has been demonstrated to be a feature of model-based decision-making (Wunderlich *et al.*, 2011; Doll *et al.*, 2015). If these mechanisms are impaired, this should result in less accurate models of the environment and consequently to reduced goal-directed behavior. In line with this view, recent evidence from animal models for addictive behaviors points toward a deficit to mentally simulate outcomes of one's behavior which are not directly experienced (Lucantonio *et al.*, 2014). Deficits in using fictive (i.e., not directly experienced) learning signals which code 'what might have happened' have also been described in human smokers before (Chiu *et al.*, 2008).

3.5.3 Risk factors for addiction: habitual behavior as a vulnerability marker for addiction?

A shift from model-based to model-free behavior has not only been proposed for addiction itself, but mostly on the theoretical level also as a vulnerability factor for addictive behaviors (Huys *et al.*, 2014). In the following, I discuss findings in the field of behavioral control in selected risk factors for addictive disorders (Merikangas *et al.*, 1998; Sinha, 2008; Verdejo-Garcia *et al.*, 2008;

Robbins *et al.*, 2012). Risk factors investigated in this thesis are 1) impulsivity 2) positive family history 3) cognitive functioning 4) acute and chronic stress. Although it is conceivable that combinations of and interactions between these factors contribute to vulnerability for addiction, the subsequent passage mainly follows an isolated description of risk factors.

3.5.3.1 Impulsivity

From a personality psychology perspective, impulsivity has for long been discussed as closely associated with addictive behavior (Verdejo-Garcia *et al.*, 2008) and has indeed been termed an endophenotype of addiction (Ersche *et al.*, 2010; Robbins *et al.*, 2012). Within the pathogenesis of addiction, impulsivity has also been associated with habitual behavior: based on animal work, it is suggested that impaired inhibitory control – by definition a hallmark of high impulsivity – increases the risk of initial voluntary substance use, but also speeds up the transition into habitual drug consumption (Everitt *et al.*, 2008). In human beings, Ersche and colleagues found higher impulsivity levels in addicted individuals and intermediate impulsivity levels in their 1st degree relatives (Ersche *et al.*, 2012a). Low ventral striatal activation towards anticipation of rewards was shown to be associated with increased impulsivity in alcohol-dependent patients (Beck *et al.*, 2009). Regarding behavioral control, first support for the hypothesis of an overreliance on the habitual system in high-impulsiveness comes from devaluation paradigms, demonstrating reduced goal-directed control in high-impulsive smokers (Hogarth *et al.*, 2012b). However, to the best of my knowledge, no previous studies involving a neuro-computational approach to tackle the association of impulsivity with model-free vs. model-based control in non-addicted, healthy, but high-impulsive individuals have been reported.

3.5.3.2 Positive family history of addiction

Addiction runs in families, such that relatives of individuals suffering from substance addiction have an eight-fold increased risk of substance abuse disorder, rendering it as one of the most prominent risk factors for substance-related disorders (Merikangas *et al.*, 1998). Studying relatives of drug-dependent individuals enables to identify well-defined cognitive or behavioral processes as (familial) vulnerability factors, i.e. stemming from genetic or shared environmental

causes: if such alterations in a certain cognitive or behavioral process can not only be found in patients themselves, but also in unaffected siblings, it seems reasonable that they constitute a risk factor of the disease rather than a consequence. These cognitive or behavioral alterations have been termed *endophenotype* or *intermediate phenotype* (Gottesman & Shields, 1973; Ersche *et al.*, 2010; Robbins *et al.*, 2012). In the case of alcohol-addiction, where the abused drug has strong neurotoxic effects, the endophenotype approach enables to disentangle cognitive, behavioral and neural factors related to addictive behavior per se from more general deficits, emerging as a consequence of drug consumption, e.g. due to grey matter loss (Beck *et al.*, 2012; Ersche *et al.*, 2012a) or drug-induced cognitive impairment (Beveridge *et al.*, 2008). However, to date, behavioral control as defined in this thesis has not been studied in unaffected relatives of psychiatric patients.

3.5.3.3 Low cognitive function

In studies involving individuals suffering from addiction and their siblings, low cognitive function has been claimed a further endophenotype for addiction. Low cognitive function, specifically executive function, was not only found impaired in drug users but also in their relatives (Ersche *et al.*, 2012b). On the other hand, neuro-computational investigations of the balance between model-free vs. model-based control have provided evidence for the theoretically well-accepted fact that the model-based system strongly relies on cognitive capacities (Otto *et al.*, 2013a; Otto *et al.*, 2013b; Schad *et al.*, 2014; Otto *et al.*, 2015). This complicates the interpretation of studies in addicted individuals as observed differences in behavioral control might be attributed either to a disease-specific mechanism or to a general cognitive decline (Sebold *et al.*, 2014). As described above, studies in at-risk populations are a promising venue to assess the association of cognitive capacities and behavioral control as vulnerability for the development of addiction beyond general cognitive decline.

3.5.3.4 Stress

Stress is a well-established risk factor for addiction development and addiction relapse, as evidenced by rodent as well as human epidemiological studies (Sinha, 2008). Specifically, the

role of chronic stress for addiction onset is emphasized Interestingly, for acute stress there is also evidence derived from self-control (Maier *et al.*, 2015) or devaluation paradigms (Schwabe & Wolf, 2009; 2011; 2013) that stress impairs goal-directed behavior or biases behavioral control towards habitual decision-making. Adopting a computational approach, a recent study has however not found between-group differences in acutely stressed vs. control participants regarding the balance of model-free vs. model-based behavior during sequential decision-making (Otto *et al.*, 2013b), but an association with individual physiological stress responsivity. Noteworthy, stress influences cognitive processes differently, depending on the timing and duration of exposure to stress (Lupien *et al.*, 2009). Thus, it seems surprising that effects of chronic stress on behavioral control have scarcely been investigated in human beings; in a rodent model, Dias-Ferreira *et al*. (2009) combined a devaluation procedure with chronic stress exposure. They could show that chronic stress triggered rats to become insensitive to changes in outcome value and thus more habitual. The interaction of chronic and acute stress on habitual vs. goal-directed control in humans has, to my knowledge, not been assessed in previous studies. This might however be an important influence factor for onset of and relapse to substance abuse.

3.6 Current challenges in psychiatric research and emerging paradigms

Progress in psychiatric research has been deemed "disappointing" (Stephan & Mathys, 2014, p. 86) or as "a field in crisis" (Wiecki *et al.*, 2015, p. 378; Hyman, 2012; Montague *et al.*, 2012). At the core of critique is an "explanatory gap" (Kapur *et al.*, 2012; Montague *et al.*, 2012, p.73). This terms the lack of an intermediate level connecting explanations, e.g. on the molecular level (where in fact psychopharmacology exerts its impact) or other possible causes to clinical phenotypes. Due to this gap, psychiatry mainly relies on oversimplified explanations (Maia, 2015) – schizophrenia arises from dopamine, depression involves serotonin etc., although, arguably, psychiatric disorders affect an extremely complex and dynamic biological system – the brain. Further, the lack of any mechanistically informed explanations for individual disease processes leads the field to cling to a purely symptom-based description of psychiatric states, as

operationalized in the current classification systems. These classifications rather rely on clinical plausibility and utility than on scientifically justified criteria and seem far removed from biological plausibility (Buckholtz & Meyer-Lindenberg, 2012). Critically, interrater-reliability of classifications based on the Diagnostic Statistical Manual (DSM), which had originally been the cause for its development, has been disparaged: the newly developed DSM-5 shows moderate to low reliability estimates for the majority of diseases (Freedman *et al.*, 2013). One major problem contributing to deficient current classification is heterogeneity: patients can get the same diagnoses, even if their symptoms are widely disparate (as in the case of schizophrenia), or even opposite in nature (compare psychomotor retardation and agitation, which are both criteria for depression). On the other hand, similar symptoms are present in different diagnostic entities, e.g. compare loss of control over certain behaviors in substance addiction as well as eating disorders like binge eating or bulimia nervosa (Robbins *et al.*, 2012; Wiecki *et al.*, 2015). A related problem is comorbidity, or so-called *artificial comorbidity*, which means that currently separate diagnoses are given for symptoms that could have a common cause or pathogenetic pathway – thus belong together for treatment or outcome prediction (Buckholtz & Meyer-Lindenberg, 2012; Wiecki *et al.*, 2015). As a consequence, current psychiatric practice is accused of treatment on a trial-and error-basis (Wiecki *et al.*, 2015). To overcome these issues, over the recent years, the field has pinned its hopes on two emerging paradigms, namely *Dimensional Psychiatry* and *Computational Psychiatry* that complement each other rather than being mutually exclusive.

3.6.1 Dimensional Psychiatry

It has been criticized that the categorical definition according to DSM leads to the classifications being treated as if they were "natural kinds – inherently meaningful, ontologically (biologically) valid taxons." (Buckholtz & Meyer-Lindenberg, 2012, p. 993). Consequently, the focus of psychiatric research is shifted towards unique and entity-specific pathogenetic processes. This seems off-target: on the clinical level, as outlined above, symptoms overlap, there is frequent crossover from one disorder to another, and comorbidity is pervasive. In addition, current

treatment does not follow these categorical boundaries, as similar psychopharmacological agents are used to treat different diagnostic entities (Stahl, 2008). Likewise, similar or non-specific psychotherapeutic factors are effective across diagnoses (Grawe, 2005). Furthermore, current biological accounts of mental illness, e.g. genetic or neurobiological explanations, rather speak in favor of a continuous and dimensional model of mental illness (Heinz, 2002; Buckholtz & Meyer-Lindenberg, 2012). From a dimensional perspective, different symptoms of multiple domains are seen as more or less expressed in the population, which can fall within the range of normal behavior, but in the case of extreme hypo- or hyper-expression manifests itself in psychopathology. Expression of one or the other symptom can also be common to a variety of diagnoses rather than specific for one categorical disorder. Figure 3-2 illustrates this conception.

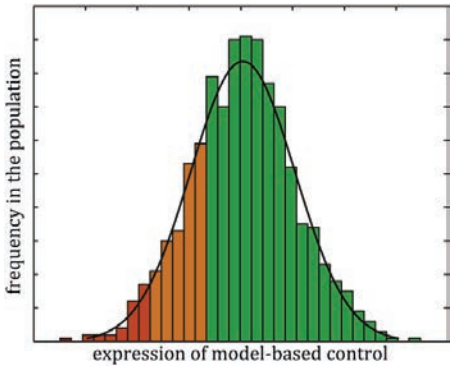


Figure 3-2. Hypothesis of expression of model-based behavior in the population according to a Dimensional Psychiatry approach. Marked in red would be the putative expression in the disease group, marked in orange would be the putative expression in the risk group. Note that the other end of the continuum is not marked in red, which mirrors the current research focus that particularly a reduction in model-based control leads to harmful decisions. However, hypotheses on too much model-based control begin to arise, with speculations that an overdominance might be related to psychotic phenomena (Dolan & Dayan, 2013).

Dimensional Psychiatry builds on this notion and aims at investigating psychiatric disorders from this transdiagnostic perspective, across nosological boundaries. In an influential article, Robbins and colleagues suggested neurocognitive (endo-)phenotyping of multiple symptom dimensions (e.g. compulsivity and impulsivity) across disorders (e.g. substance abuse, eating disorders, Obsessive Compulsive Disorder), but also independent of diagnosis, in healthy but at-risk populations, e.g. in first-degree relatives (Robbins *et al.*, 2012). The approach relies on behavioral tasks combined with neural measures to find alternative ways of classifying. In line with this, a recent initiative by the United States' National Institute of Mental Health, the *Research Domain Criteria* project, aims at a new description of mental disorders, agnostic to DSM-criteria, by combining information of genetics, neuroimaging, behavior and self-report (Insel *et al.*, 2010).

This new way of classification is aimed to be quantitatively measurable, putatively closer to the underlying neural mechanism. Further, it acknowledges heterogeneity in the general population (Wiecki *et al.*, 2015). However, the suggested phenotypes might again constitute umbrella terms composed of multiple poorly defined underlying basic mechanisms. This reformulates behavioral impairments in patients but might still leave the *how* question unanswered (Verbruggen *et al.*, 2014).

3.6.2 Computational Psychiatry

The nascent field of Computational Psychiatry is concerned with translating the modeling-based cognitive neuroscience approach outlined in section 2.2 and concepts of theoretical computational neuroscience in general, to the field of psychiatric research. Hence, much as described in the General Principles section, Computational Psychiatry uses quantitative mathematical models to unravel behavioral alterations, neural characteristics and brain-behavior relations in psychiatric disorders and symptoms. Building on this, Computational Psychiatry targets at answering specific questions concerned with a mechanistic explanation of psychiatric disorders and developing improved theories of mental illness, and works out custom-made techniques for data analysis to achieve this. Specific goals include (Maia, 2015):

classifying patients, e.g. based on neuroimaging or EEG data (Kloppel *et al.*, 2012), clustering clinically relevant subgroups based on modeling techniques agnostic to standard diagnostic criteria (Brodersen *et al.*, 2014), and, investigating group differences (e.g. patients vs. control group, patients with different diagnoses, subgroups of patients) regarding models or model parameters which best fit behavioral or physiological data (e.g. fMRI, EEG, Maia & Frank, 2011). The latter goal is pursued in this thesis. It involves questions such as: do patients follow the same model to solve an experimental task as controls or as patients with another diagnosis? Do groups differ in model parameters that represent specific neural or cognitive processes? Is the latter related to clinical measures, like symptom severity or treatment outcome? In sum, the overarching aim of Computational Psychiatry relates back to shrinking the explanatory gap between cause and phenotype by tracing a causal pathway from a change in one variable to the effect onto another variable (Maia, 2015) with the final goal to predict outcome and to guide treatment.

4 Research questions and study design

Building on the previous chapters, the next paragraph amalgamates the methodological and theoretical aspects introduced to develop the rationale and hypotheses underlying the current empirical investigation. Figure 4-1 and 4-2 are designed to illustrate the current framework. Adopting a dimensional psychiatry approach (see Figure 4-1), the empirical section of the thesis focusses on RL-mechanisms of behavioral control in different psychiatric patient groups and associated risk factors in a series of six studies. To this end, I employed two different experimental paradigms, sequential decision-making and counterfactual decision-making, both designed to measure flexible behavioral adaptation. Following the Computational Psychiatry approach, all studies applied computational RL-models to the measured behavioral data. In studies 1, 2 and 4, this is combined with a modeling-informed fMRI analysis. In study 3, this is combined with a modeling-based ERP analysis. Study 5 and 6 are behavioral studies (see Figure 4-2).

The empirical studies revolve around the following general research questions:

1) Is substance addiction characterized by impaired mechanisms of flexible goal-directed behavioral adaptation? Can this deficit be explained by reduced abstract inference on un-chosen choice options ("what might have happened")? What are neural correlates of this deficit? (study 1)
2) Do potential impairments of this kind transdiagnostically extend to binge eating disorder, a nosologically distinct diagnosis, which shares clinical features with substance addiction? What are shared and differential behavioral mechanisms and neural correlates? (study 2)
3) Does a shift from model-based to model-free behavioral control extend to recognized risk factors of addiction and might thus be seen as a vulnerability factor for addiction? (studies 4-6)

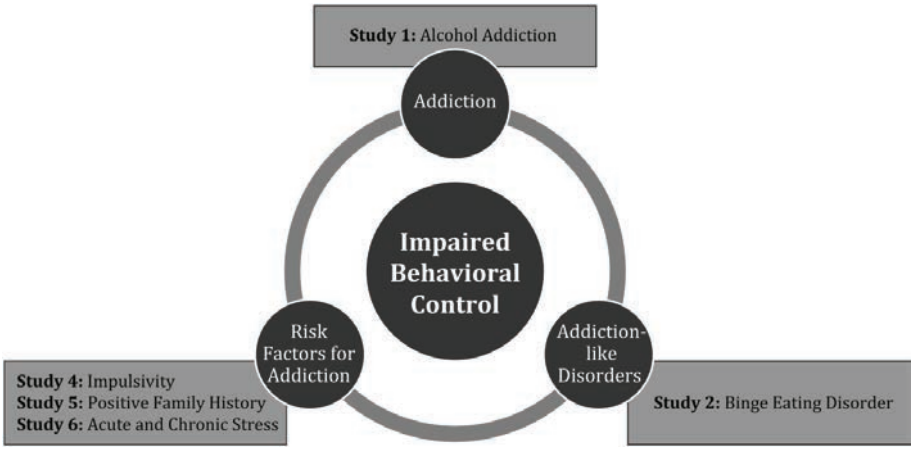


Figure 4-1. Research questions and design. This thesis was motivated by the question of mechanisms underlying impaired behavioral control in substance addiction (alcohol addiction, study 1). Adopting a dimensional psychiatry approach, I aimed to investigate whether and how impaired mechanisms of behavioral adaptation extend to so-called addiction-like disorders (here using the example of binge eating disorder, study 2), and several risk factors of addiction (study 4: impulsivity, study 5: positive family history, study 6: acute and chronic stress).

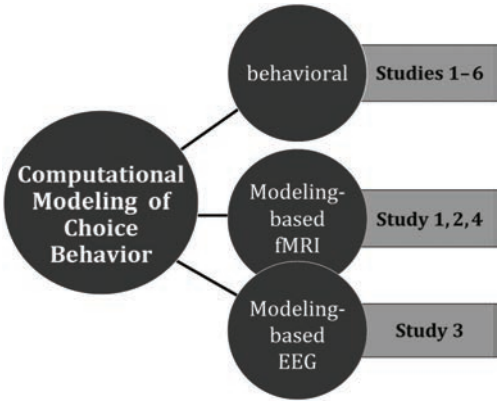


Figure 4-2. Methods. Behavioral computational modeling was applied in all studies to analyze choice behavior in reward-based decision-making tasks. In studies 1, 2 and 4, parameters derived from the computational model were combined with neural signals measured via BOLD-fMRI. Study 3 aimed to identify modeling-derived signals in the EEG.

According to research question 1, mechanisms of abstract inference on unchosen alternative options were investigated in alcohol-dependent patients by using counterfactual decision-making during fMRI (study 1). Inspired by the transdiagnostic framework, I extended the same study design to binge eating disorder. Study 2 was thus inspired by research question 2 whether aspects of altered behavioral control found in substance addiction translate to a nosologically different disorder which shares as a common clinical feature the loss over behavioral control.

Study 3 extended the study design of the fMRI studies 1 and 2 towards EEG – due to enhanced feasibility, this might enable studying research questions as in study 1 and 2 also in a broader range of patient populations. I ask whether, using a modeling-based approach combined with EEG, we can define electrophysiological correlates of abstract inference during flexible behavioral adaptation.

Studies 4-6 use sequential decision-making, all designed to tackle research question 3. They build on previous findings as reviewed above which indicate reduced model-based behavior in addiction. In the sense of a dimensional approach, sequential decision-making was applied to the following risk factors of addiction: impulsivity (study 4) positive family history of addiction, impulsivity and cognitive function (study 5) and acute and chronic stress (study 6).

5 Study 1: Neglecting what might have happened – disturbed inference on alternative choices in alcohol-dependent patients[2]

5.1 Introduction

A key characteristic of addictive disorders is the fact that addicted individuals continue substance use despite evident harmful consequences. Addicted individuals regularly report *having no choice* but to consume. This suggests an impairment of integrating different choice options and their potential consequences. Thus, neglecting 'what might have happened' may rigidly bias decision-making towards choice options that have been proven to be rewarding in the past (Chiu *et al.*, 2008; Dayan, 2009a). *Computational Psychiatry* accounts (Montague *et al.*, 2012) have theoretically linked these maladaptive decision-making processes to disturbed mechanisms of inference (Huys *et al.*, 2015b), e.g., difficulties to use statistical regularities of the environmental structure to guide decisions, and in the same vein, to the dichotomy of habitual and goal-directed behavioral control (Dickinson & Balleine, 2002; Dolan & Dayan, 2013). Habitual control arises from retrospective 'model-free' learning via past rewards and neglects environmental structure with regard to alternative choice options. In contrast, goal-directed control enables rapid behavioral adaptation due to planning action-outcome contingencies involving a mental model of the environment. When inferring such a mental model, an individual must rely on environmental statistics, e.g., on how different decision options relate to each other. Thus, concurrent tracking of multiple decision possibilities, including the incorporation of abstract inference about 'what might have happened', promotes a mental map of the environment and advances flexible and goal-directed adaptation in dynamic environments (Hampton *et al.*, 2006; Li & Daw, 2011). Neural signatures of habitual and goal-directed control have been demonstrated by fMRI in prefrontal cortex and striatum (Balleine & O'Doherty, 2010; Daw *et al.*, 2011). Thus, intact decision-making relies on the integration of both aspects of

[2] This chapter corresponds to the following article: Reiter, A.M.F., Deserno, L., Kallert, T., Heinze, H.J., Heinz, A. & Schlagenhauf, F. (under review). Neglecting what might have happened - reduced inference on alternative choices in alcohol-dependent patients.

behavioral control (Daw *et al.*, 2005) with specific involvement of lateral and medial prefrontal regions in an integration process between both systems (Lee *et al.*, 2014).

Addiction has been theorized as one prime example for a breakdown of behavioral control in favor of model-free control (Everitt & Robbins, 2005; Dayan, 2009a; Lucantonio *et al.*, 2012) with support from first behavioral studies (Sebold *et al.*, 2014; Voon *et al.*, 2015). A common factor underlying observed impairments in behavioral adaptation might be a deficit in abstract inference on higher-order environmental structure, e.g., on the interdependencies of values, which might hamper the addicted individual in constructing and using a model of the environment. Interestingly, on the neural level, intact model-free learning signals have been demonstrated in alcohol addiction (Park *et al.*, 2010). Here, we probe if the modification of basic model-free neural learning signals by goal-directed inference regarding the environmental structure is disturbed in alcohol addiction and whether this relates to the clinical feature of obsessive drinking.

To address this, we employed functional Magnetic Resonance Imaging (fMRI) during decision-making in a dynamic environment to examine flexible behavioral adaptation. Importantly, reward contingencies of different options were anti-correlated: whenever one stimulus was a good choice, the other one would be the worse choice and vice versa. When confronted with options as in this task, individuals make choices based on decision values computed for the options at hand (e.g. Rangel *et al.*, 2008). These values can either be deduced by action-reward pairings or abstract inference on the task structure, i.e. the anti-correlated reward probabilities (Hampton *et al.*, 2006; Bromberg-Martin *et al.*, 2010). This allows to test whether an agent infers the correct structure of the task and learns accordingly. We hypothesized that alcohol-dependent patients fail to integrate this abstract inference, 'what might have happened', into the value of chosen options. To this end, we compared computational models of learning that differ in the degree of updating alternative choice options. As a neural substrate, we predicted prefrontal signatures reflecting abstract inference on alternative options to be reduced in alcohol-dependent patients.

5.2 Materials and methods

Participants. 43 alcohol-dependent patients and 35 healthy participants were included. fMRI data were available of 35 healthy participants and 34 patients. Patients had abstained from alcohol for at least 8 days (range: 8-56 days, *m*=28.80, *SD*=11.85) and were diagnosed as alcohol-dependent (DSM-V and ICD-10). See supplementary information (SI-1) and S-Table 1-1 for demographic, neuropsychological and clinical characteristics and for details on the recruitment strategy, diagnostics and exclusion. The local Ethics committee approved the study. Participants gave written informed consent and were reimbursed for participation.

Task. Participants performed counterfactual decision-making in a dynamic environment that requires flexible behavioral adaptation (Figure 5-2A for illustration). Importantly, the task incorporated a simple higher-order structure: task-structure was counterfactual by a perfect anti-correlation of the reward probabilities associated with the two choice options; whenever stimulus A was a good choice, stimulus B would be the worse choice and vice versa. Even though the outcome for the alternative option is never shown, the agent can infer from the anti-correlation of the options 'what might have happened' if he had taken the other stimulus (Figure 5-1). Reward contingencies remained stable for the first 55 trials (first, 'pre-reversal' phase) and also for the last 35 trials (last, 'post-reversal' phase). During the second, 'reversal', phase, reward contingencies changed (four changes in total, after 15 or 20 trials, see Figure 5-2B). This required participants to flexibly adapt their behavior. For details on timing, training and instruction see SI-1.

Analysis of choice behavior. Behavioral performance was quantified as percentage of correct choices (choices of the stimulus with 80% reward probability), and was analyzed using repeated-measures ANOVA including the between-subject factor *group* (patients vs. controls) and the within-subject factor *phase* (pre-reversal: first 35 trials, reversal: intermediate 90 trials, post-reversal: last 35 trials). We investigated the effect of previous feedback on subsequent decisions, i.e. repeating choices after reward, 'win-staying', and shifting responses after losses, 'lose-shifting' (den Ouden *et al.*, 2013).

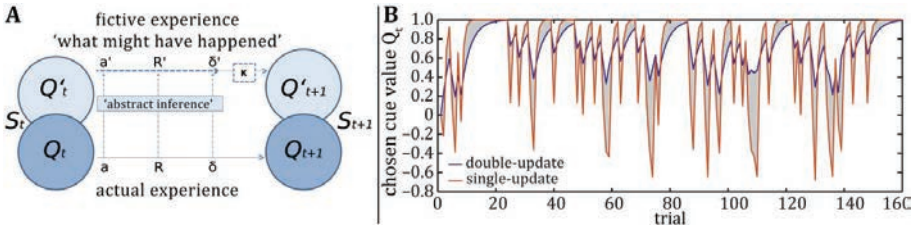


Figure 5-1. Schematic: parallel *double-updating* of chosen and unchosen choice values. A) At time t, an agent in state S_t passes to a new state S_{t+1} by the action a, observing the outcome R which leads to the reward prediction error δ as the difference between expected and actually gained reward. Accordingly, the agent updates the chosen value for the next trial, Q_{t+1}. Although not explicitly observed, the agent can conclude from the counterfactual task-structure *what might have happened* (R') if it had chosen an alternative action a', resulting in a fictive prediction error δ'. Thus, by abstract inference on the counterfactual task-structure and parallel to updating chosen values, the agent additionally *double-updates* unchosen values Q'_{t+1}. Individuals might differ in their degree of abstract inference on the environmental structure. The individual degree of *double-updating* is therefore weighted by the parameter κ. B) Effect of abstract inference, *double-updating,* on chosen values. For one exemplary participant, values of the respective chosen value are plotted per trial, as a function of the two alternative control strategies: pure *single-updating* (κ=0, neglecting 'what might have happened', red) versus pure *double-updating* (κ=1, full abstract inference on the task-structure, blue). Hence, the difference of both (here, highlighted in gray) represents an estimate of the degree of abstract inference on the counterfactual task-structure. In our analysis of functional imaging data, we probe how this difference in choice values with respect to abstract inference modulates the coding of the core teaching signal, the reward prediction error δ for chosen values.

We furthermore quantified how often participants repeated a choice despite two consecutive losses for the same choice in the preceding two trials, relative to all loss trials.

Computational Modeling: Different reinforcement learning (RL-) models were fitted to the data. All models learn values of choice options via reward prediction errors (RPEs), a teaching signal that compares received rewards and expected values. In essence, the first three RL-models differ in the degree of updating both, the chosen and alternative, decision options.

I) A model-free learner who only updates values for the chosen stimulus and therefore neglects the counterfactual task-structure. We refer to this as the single-update (SU) model;

II) A learner who updates values of chosen and unchosen stimuli equally using abstract inference on the task-structure. We refer to this as the double-update (DU) model;

III) A hybrid model connecting the models described above by individually weighting the degree of double-update learning, accounting for individual variability regarding this type of abstract inference. This is given by the weighting parameter κ.

Figure 5-1 provides a schematic of abstract inference on the task-structure with respect to unchosen choice values ("double-updating"). In the task at hand, as double-updating depends on abstract inference derived from actually experienced feedback, updating of the unchosen stimulus always relies on learning from feedback for the chosen stimulus, i.e. is rather unlikely to be a process independent from updating the chosen stimulus; compare Li et al., 2011 for an identical implementation. In all three models, we let free parameters differ for reward and punishment $\alpha_{rew,}\, \alpha_{pun}$ *and* $\beta_{rew,}\, \beta_{pun}$ as we had verified that they outperform models without this distinction (see SI-1).

As suggested by our reviewers, we additionally included a model with an adaptive learning rate (Sutton-K1), in which the learning rate is dynamically updated as a function of the change in prediction errors encountered (Sutton, 1992). It was previously discussed and used as a non-hierarchical approximation of a dynamic learning rate (Chumbley *et al.*, 2012; Kepecs & Mainen, 2012; Landy *et al.*, 2012; Iglesias *et al.*, 2013). By including this model, we tested whether a model with a dynamic learning rate captures the observed behavior better than algorithms with a fixed learning rate. In SI-1, we describe each model accompanied by equations and provide details on simulations, model-fitting and model selection.

Analysis of fMRI data. See SI-1 for functional and structural MRI data acquisition and preprocessing. The aim of the statistical analysis was to elucidate neural signatures of RPEs for chosen values as a function of SU- versus DU-learning and potential group-differences. Onsets of feedback delivery were modulated parametrically by two trial-by-trial regressors from computational modeling: first, individual RPEs for chosen values were computed on basis of the SU-Model with $\kappa = 0$ (RPE_{SU}). Second, individual RPEs for chosen values were computed based on the DU-Model with $\kappa = 1$ (RPE_{DU}). We computed the difference of RPE_{DU} *minus* RPE_{SU}. This reflects the difference in chosen values from the DU- and SU-algorithms, which, in the hybrid model, is provided by the estimate of κ (illustrated in Figure 5-1B). Thus, throughout the manuscript, the second parametric modulator, the difference regressor, is referred to as RPE_{DU}. This procedure accounts for co-linearity between the regressors (Daw *et al.*, 2011). At the second level, contrast images for RPE_{SU} and RPE_{DU} were taken to a random-effects analysis. A full-factorial ANOVA contained the type of RPEs (RPE_{SU} /RPE_{DU}) as within-subject factor and group as between-subject factor. For details on the statistical analysis and Voxel-Based-Morphometry (VBM), see SI-1

5.3 Results

Correct choices. An ANOVA revealed a significant effect of *phase* ($F(2, 75)=21.76$, $p<.001$), *group* ($F(1,76)=19.97$, $p<.001$) and a significant *group* x *phase* interaction ($F(2,75)=3.27$, $p=.04$, Figure 5-2C).

Win-staying and lose-shifting. We further explored patients' deficit in correct choices by analyzing how often participants repeated choices after reward, 'win-staying', and shifted after losses, 'lose-shifting'. A between-group difference was found on win-staying ($t(76)=2.23$, $p=.03$) with patients showing less stay behavior after wins ($m_{controls}=.93$, $SD_{controls}=.06$; $m_{patients}=.87$, $SD_{patients}=.14$). There was no difference in lose-shifting ($t(76)=.25$, $p=.80$).

Repeating choices despite recurrent negative consequences. We found a significant between-group difference ($t(76)=2.63$, $p=.01$) in repetition behavior after two successive losses ($m_{controls}=.11$,

$SD_{controls}$=.08; $m_{patients}$=.18, $SD_{patients}$=.14); patients reiterated disadvantageous choices more often, despite negative consequences in preceding trials.

Computational modeling: model comparison. Using Bayesian Model Selection (Stephan *et al.*, 2009), we identified a best-fitting model in controls and patients. Exceedance probabilities *XP* show which model most likely accounts for behavior in each population. Across both groups, this revealed that the hybrid model outperformed the other models ($XP_{hybrid,}$=.89, XP_{DU} =.00, XP_{SU}=.11, $XP_{Sutton\text{-}K1}$ =.00). Groups differed regarding the model which explained their behavior best, indicating different control strategies to solve the task (Figure 5-3A): controls were best

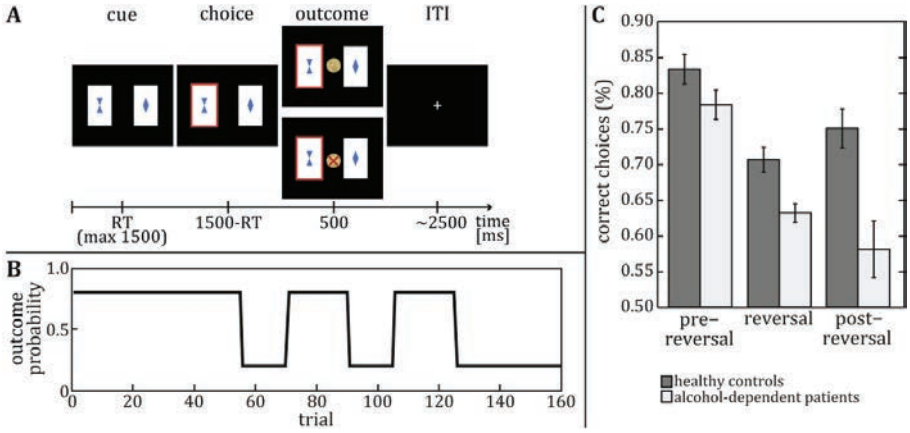


Figure 5-2. Counterfactual decision-making task. A) Exemplary trial sequence. In a total of 160 trials, participants were instructed to select between two cards each showing a different geometric stimulus, the card which they believed to be associated with reward. After the participant had chosen one stimulus by button press (left vs. right button) the selected stimulus was highlighted. Subsequently, feedback was shown: outcome stimuli were a 10 Eurocent coin (in the case of a win) and a crossed 10 Eurocent coin (in the case of a loss). B) One of the stimuli was assigned with a reward probability of 80% and a punishment probability of 20% (vice versa for the other stimulus). Reward contingencies remained stable for the first 55 trials (pre-reversal block) and also for the last 35 trials (post-reversal block). In between, reward contingencies changed four times (reversal block). C) Raw Data Results. Correct choices differed significantly as a function of phase (pre-reversal, reversal, post-reversal, $F(2,75)=21.78$ $p<.001$). We observed a main effect group and a significant interaction of phase by group ($F(2,75)=3.27$, $p=.04$). Between-group post-hoc t-tests revealed that group differences were present in the reversal phase ($t(76)=3.48$, $p=.001$) and in the post-reversal phase ($t(76)=3.36$, $p=.001$) but not in the initial stable pre-reversal phase ($t(76)=1.69$, $p=.10$).

explained by the hybrid model which includes abstract inference on the task-structure by considering individual variability given by the parameter κ, an individual weight of the degree of DU-learning models ($XP_{\text{hybrid,}}$=.99, XP_{DU} =.00, XP_{SU}=.003, $XP_{\text{Sutton-K1}}$ =.00). Patients were best explained by the model-free SU-Model which neglects an update of the alternative choice option ($XP_{\text{hybrid,}}$=.06, XP_{DU} =.00, XP_{SU}=.93, $XP_{\text{Sutton-K1}}$ =.01). Model selection results were robust against the exclusion of participants who were not fitted better than chance by any model (n=7, for details see SI-1). Measures of absolute model fit indicated that the best-fitting model captured behavior well (adjusted Mc Fadden's pseudo R^2 =.60, see SI-1 and S-Figure 1-3 for details).

Table 5-1. Descriptives of best-fitting parameters (hybrid model). The multiplication of κ with the learning rate α corresponds to the double-update learning rate for the unchosen stimulus.

	β_{reward}	β_{punish}	α_{reward}	α_{punish}	$\kappa^*\alpha_{\text{reward}}$	$\kappa^*\alpha_{\text{punish}}$
Healthy controls	*M* = 4.28 *SD* = 1.16	*M* = 1.99 *SD* = 1.45	*M* = 0.56 *SD* = 0.23	*M* = 0.48 *SD* = 0.25	*M* = 0.10 *SD* = 0.11	*M* = 0.09 *SD* = 0.12
Alcohol-Dependent Patients	*M* = 4.14 *SD* = 0.96	*M* = 1.41 *SD* = 1.21	*M* = 0.49 *SD* = 0.32	*M* = 0.45 *SD* = 0.31	*M* = 0.08 *SD* = 0.12	*M* = 0.04 *SD* = 0.05

Computational Modeling: model parameters. We tested for between-group differences by subjecting the inferred parameters of the hybrid model, the best-fitting model across both groups (Table 5-1), to a multivariate ANOVA (MANOVA) with group as between-subject factor (patients vs. controls). This MANOVA contained the following parameters each separately for reward and punishment: stochasticity parameters (β_{rew}, β_{pun}), learning rates for the update of chosen (α_{rew_c}, α_{pun_c}), and unchosen values (α_{rew_uc}, α_{pun_uc}, products of the weighting factor κ with α_{rew_c} and α_{pun_c}). This revealed a significant effect of the between-subject factor group (F(7,70)=2.92, p=.01) due to a significantly lower DU-weighted learning rate after having received punishment α_{pun_uc} (F(1,76)=6.35, p=.01, Figure 5-3B), whereas none of the other parameters differed between groups (group differences regarding learning rates of the simpler

model-free SU algorithm, all p>.66). Results were robust against the exclusion of participants who were not fitted better than chance by any model (n=7, see SI-1). Using these inferred parameters to simulate choices (S-Figure 1-2), and testing between-group effects on simulated choice data in the same manner as in the empirical data, all group differences on choice behavior were recaptured except for the group by phase interaction (see SI-1). Taken together, this indicates a specific learning impairment in the alcohol-dependent group affecting abstract inference on alternative choice options after punishment.

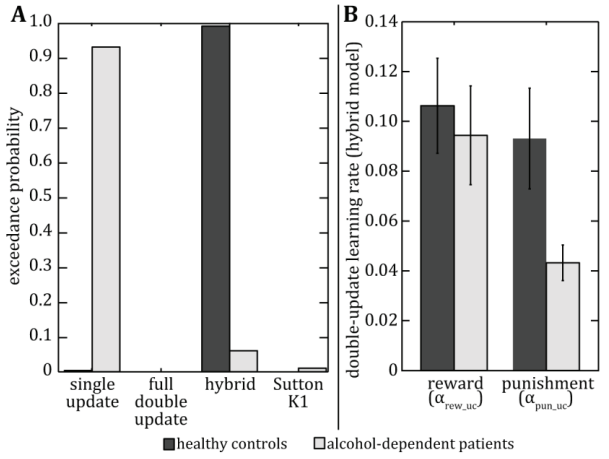


Figure 5-3. Computational Modeling Results. A) Bayesian Model Selection revealed that healthy controls were best explained by the hybrid model, including a factor which weights the individual degree of abstract inference ('double-updating'), whereas for alcohol-dependent patients, model evidence was maximal in favor of the model-free Single-Update Model. B) Between-group comparisons on the inferred parameters revealed that alcohol-dependent patients specifically differed in the DU-learning rate for punishments compared to healthy controls (t(76)=2.52, p=.01), whereas there was neither a group difference in the reward learning rate nor in any of the other inferred parameters of the model.

Neural signatures. To explore neural signatures of this behavioral deficit, we analyzed encoding of two types of RPEs for the chosen option, i.e. RPEs derived from the SU-Model (RPE_{SU}) versus RPEs derived from the DU-Model (RPE_{DU}). Main effects for both types of learning signatures for both groups taken together are illustrated in Figure 5-4 and Table 5-2. To investigate between-

group differences, we tested for a type of RPE (RPE_{SU}/RPE_{DU}) x group (patients/controls) interaction. For between-group comparisons, the conjunction of both RPEs across the entire sample (thresholded at *p*-FWE<.05 for the whole brain, Figure 5-4, Table 5-2) was used to correct for multiple comparisons (at *p*-FWE<.05 based on this search volume). The RPE type x group interaction reached significance in medial prefrontal cortex (mPFC, X=-10, Y=62, Z=12, *t*=3.98, FWE-corrected for the conjunction *p*=.01) and posterior cingulate cortex (X=0, Y=-40, Z=32, *t*=3.72, FWE-corrected for the conjunction *p*=.03). As post-hoc contrast, we compared RPE_{SU} and RPE_{DU} between groups.

Table 5-2. fMRI whole-brain results for the conjunction of single-update and double-update learning signals across both groups. Only clusters k>15 are depicted. FWE-cor: family-wise-error corrected, MNI: Montreal Neurological Institute, K: cluster size.

Region	MNI Coordinate	K	p (FWE-cor) Cluster Level	t-value	p (FWE-cor) Peak-Level
Conjunction: single-update and double-update learning signals					
Superior Medial Gyrus	-10 64 12	79	<0.001	6	0.001
Middle Orbital Gyrus	-6 54 2			5.85	0.002
Middle Orbital Gyrus	6 42 -8	174	<0.001	7.46	<0.001
Middle Orbital Gyrus	8 60 4			6.18	<0.001
Middle Orbital Gyrus	-24 32 -16	19	0.001	5.55	0.006
Ventral Striatum	-8 8 -10	43	<0.001	7.39	<0.001
Ventral Striatum	10 12 -8	66	<0.001	7.01	<0.001
Posterior Cingulate Gyrus	0 -34 34	56	<0.001	6.30	<0.001
Precuneus	-4 -50 16	25	<0.001	5.55	0.006

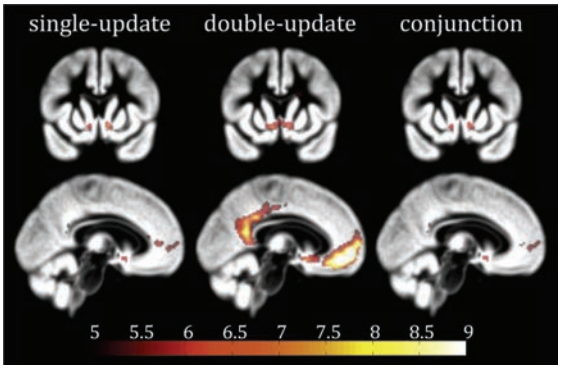


Figure 5-4. Neural coding of single-update vs. double-update signals across the entire sample. Across all participants (patients and controls), we observed model-free RPE_{SU} in bilateral ventral striatum, and medial and lateral prefrontal cortex (FWE-corrected for the whole brain p<.05, S-Table 1-5). For the difference regressor RPE_{DU}, we found effects in overlapping regions (bilateral ventral striatum, medial and lateral prefrontal cortex) and additionally in hippocampus and insula (FWE-corrected for the whole brain p<.05, S-Table 1-6). The conjunction of both contrasts revealed overlapping effects of RPE_{SU} and RPE_{DU}, in bilateral ventral striatum, medial and lateral prefrontal cortex, and posterior cingulate cortex (FWE-corrected for the whole brain p<.05, Table 5.2). The latter was used as a search volume for small-volume-correction of group differences. Effects are reported using a significance level of p<.05 FWE-corrected for the whole brain. Activations are shown superimposed on an averaged gray-matter mask of the entire sample. For display purposes, threshold is set at t>5.

This confirmed significantly reduced coding of RPE_{DU} signatures in patients in the mPFC (X=-8, Y=62, Z=12, t=4.36, FWE-corrected for the conjunction p=.003; X=-6, Y=56, Z=12, t=3.68, FWE-corrected for the conjunction p=.02, Figure 5-5) and posterior cingulate cortex (X=-2, Y=-42, Z=32, t=3.72, FWE-corrected for the conjunction p=.03) but no significant between-group differences in activation elicited by model-free RPE_{SU}. Modeling-based fMRI results were robust against the exclusion of participants not fitted better than chance by any model (n=7, see SI-1). Further, we explored an association of the behavioral deficit in DU-punishment learning and the observed reduced neural representation of RPE_{DU} in mPFC. Mean parameter estimates at the peak of the between-group difference (X=-8, Y=62, Z=12, surrounded with an 8mm sphere) were extracted and, for both groups separately, correlated with the DU-punishment learning rate α_{pun_uc}. In patients, this revealed a positive association indicating that an attenuated mPFC

double-update learning signature was related to lower DU-punishment learning rates (r(33)=.36, p=.04, Figure 5-5C). No significant correlation was found in controls (r(34)=.002, p=.99). This confirms an association of the observed behavioral deficit in updating alternative options after punishment and the reported reduction of DU-signatures in mPFC in patients

Analyses of covariance. To adjust for possible confounding influences, the following variables were included as covariates in all behavioral and fMRI analyses reported: smoking status, depression score (BDI), composite measure of neurocognitive functioning as well as gray matter density (VBM). All reported results remained significant when adjusting for these possible

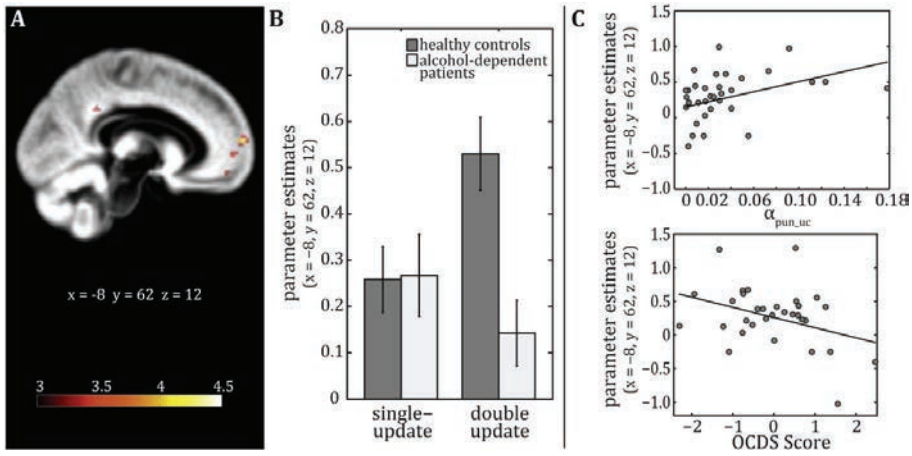


Figure 5-5. Group differences in the neural coding of single-update vs. double-update signals. A) Reduced abstract inference signatures were found in the medial prefrontal cortex (mPFC) in alcohol-dependent patients compared to healthy controls (X=-8, Y=62, Z=12, t=4.36, FWE-corrected for the conjunction p=.003, X=-6, Y=56, Z=12, t=3.68, FWE-corrected for the conjunction p=.02) and posterior cingulate cortex (X=-2, Y=-42, Z=32, FWE-corrected for the conjunction p=.03, t=3.72). No group difference regarding model-free signatures was found. For display purposes, thresholded at t>3. B) Plot of parameter estimates at the peaks of the group difference in the mPFC. C) In patients, parameter estimates from an 8mm-sphere around the peak coordinate (X=-8, Y=62, Z=12) of the group difference correlated with the behavioral deficit in double-update learning after punishments (left panel, r(33)=.36, p=.04), and were negatively associated with the sum score of the obsessive compulsive drinking scale (OCDS, *beta*=-.64 t=2.64, p=.014, right panel). We plot studentized residuals of the OCDS with respect to other disease severity measures.

confounders (see SI-1) suggesting that the behavioral and neural between-group differences are more than epiphenomena of well-known general impairments.

Relationship with symptom severity. We tested for an association of the reported behavioral and neural alterations with symptom severity in alcohol-dependent patients. We performed a linear regression analysis with mean parameter estimates of the global maximum of the observed group-difference in mPFC (at the peak-voxel with an 8mm sphere) as dependent variable and the applied self-rating measurements of addiction severity (S-Table 1-1) as predictor variables: (1) Units of alcohol consumed within 4 weeks before treatment commenced (TLFB), (2) Obsessive Compulsive Drinking score (OCDS), (3) Craving score (ACQ), (4) Alcohol Use Disorder score (AUDIT). This revealed the OCDS-Score as having a significant negative association with the neural DU-learning signature (*beta*=-.64, *t*=2.64, *p*=.01; Figure 5-5C). Patients reporting a higher level of obsessive-compulsive drinking habits showed, on the neural level, lower coding of abstract inference components.

5.4 Discussion

We provide novel insight into mechanisms of maladaptive decision-making and behavioral adaptation in alcohol-dependent patients and its underlying neural substrates. Our results support the view of a predominance of model-free control in addiction associated with disturbed mechanisms of inference on environmental structure: choice behavior in patients was best explained by a model-free RL-algorithm, which neglects updating of alternative choice options. This was due to a specific reduction of abstract inference after punishments. On the neural level, representation of abstract inference in patients' mPFC was reduced and correlated with the observed behavioral deficit in updating alternative choices as well as obsessive-compulsive drinking habits.

Disrupted behavioral adaptation in addiction. Deficits in cognitive flexibility are known in patients suffering from addiction (Bechara & Damasio, 2002; Garavan & Stout, 2005; Ersche *et al.*, 2011; Goldstein & Volkow, 2011). In line, we demonstrate that alcohol-dependent patients show diminished behavioral adaptation in a dynamic environment. Crucially, by using

computational modeling, we provide a mechanistic account for this deficit: alcohol-dependent patients are specifically impaired in the capacity to integrate alternative choice options and their fictive outcomes – 'what might have happened' – in the decision-making process after having received punishment. To put this differently, patients show less consideration of 'what might have been good instead': formally, after patients had received punishment for the chosen option they did not increase the values of the alternative option as would have been appropriate according to the anti-correlated task-structure, which was captured by a significantly lower double-update punishment learning rate in patients. This specific finding derived from computational modeling can account for the overall impairment in correct decisions, reduced win-staying and the repetition of choices despite successive punishment as suggested by our simulation analysis. Our observation suggests that simpler, model-free single-update learning is intact in addiction (such that updating of chosen values after rewards and punishments remains relatively unaffected) but that updating of alternative, un-chosen values is abolished after punishment. Abstract inference on 'what might have happened' goes awry when values need adjustment after negative feedback and thus potentially advantageous alternative choice options are neglected when making decisions. This provides translational evidence for recent animal models of addiction suggesting a specific deficit in mentally simulating outcomes not directly experienced and a disturbed integration of multiple predictions (Lucantonio *et al.*, 2014). Intriguingly, this behavioral deficit resonates well with clinical observations and diagnostic criteria of addiction describing the maintenance of disadvantageous behaviors despite negative consequences. Importantly, our finding goes beyond previous studies on behavioral adaptation linking addiction to blunted neural responses to performance errors and reduced error awareness (Paulus *et al.*, 2008; Goldstein *et al.*, 2009): a disturbed capacity to integrate abstract inference constitutes one plausible explanation for these deficits.

Abstract inference in decision-making has been previously linked to a goal-directed or model-based control system (Hampton *et al.*, 2006; Bromberg-Martin *et al.*, 2010). An alternative explanation includes that abstract double-update inference does not arise from a full

model-based system but rather reflects temporal-difference learning about the relationship of choice values (Shohamy & Wagner, 2008; Wimmer *et al.*, 2012; Doll *et al.*, 2015). In this framework, our results could be interpreted as an impairment in generalizing from one stimulus to another. Either way, the capacity to simultaneously update multiple decision values including those of unobserved outcomes might be regarded as *sine qua non* for building and using an internal model of the environment, which is important for goal-directed or model-based control. Using sequential decision-making, reduced model-based behavioral control was observed in alcohol-dependent patients (Sebold *et al.*, 2014), although this impairment was attenuated when adjusting for cognitive deficits. In the present study, the impairment in abstract inference remained significant when adjusting for cognitive capacities suggesting a specific characteristic for alcohol dependence rather than an epiphenomenon of a global impairment. Thus, our finding of reduced inferential capacities appends prominent theories proposing a shift from goal-directed to habitual behavioral control in addiction (Everitt & Robbins, 2005; Dayan, 2009a; Lucantonio *et al.*, 2012).

Blunted abstract inference mPFC signatures in alcohol-dependent patients. Patients were characterized by reduced coding of double-update RPE signals in mPFC. Reduced representation of these abstract inference signatures in patients' mPFC was related to the observed behavioral deficit and to obsessive-compulsive drinking habits. In line with our findings, alcohol-dependent patients showed hypo-activation in a similar region for a contrast assessing goal-directed learning during a different instrumental learning task (Sjoerds *et al.*, 2013). In healthy individuals, the medial prefrontal and orbitofrontal cortex is known to encode model-based values computed 'on the fly', which allows behavioral flexibility (Haber & Behrens, 2014). In consonance, the mPFC has been identified as a key region for flexible behavioral adaptation and model-based evaluation (Hampton *et al.*, 2006; Daw *et al.*, 2011). Specifically, this region has been linked to the integration of computations from habitual and goal-directed systems (Lee *et al.*, 2014): interestingly Lee et al. identified computational signals for the reliability of both systems in the mPFC. Reliability signals are likely used by an arbitration mechanism in order to

allocate the degree of control exerted by one of the two systems at a given point in time. Our observation of reduced abstract inference signatures at nearby coordinates may support a view on behavioral control in addiction that Lee et al. invite in their discussion: a failure of the arbitration process, i.e., the ability to appropriately parse behavioral control between different modes. Remarkably, reduced coding of abstract inference components in mPFC of alcohol-dependent patients remained significant when adjusting for reductions in gray matter density supporting the view of a specific neural signature of abolished abstract inference. This interpretation is strengthened by correlations of mPFC signatures with reduced double-update learning rates after punishment and obsessive drinking habits in patients. Taken together, reduced coding of abstract inference regarding alternative options in alcohol-dependent patients' mPFC may indeed account for their decreased behavioral flexibility and constitute one piece in the puzzle of obsessive alcohol consumption despite negative consequences.

Neurochemical considerations. Blunted presynaptic dopamine function was found in alcohol-addicted patients (Martinez *et al.*, 2005) and lower levels of ventral striatal presynaptic dopamine were demonstrated to be associated with a lower degree of model-based behavioral control and diminished coding of model-based prefrontal signatures during sequential decision making (Deserno *et al.*, 2015b). Thus, low levels of presynaptic dopamine could hypothetically explain the reported findings to some extent. Further, reduced dopamine D2 receptor availability is among the best-established findings in addiction (Volkow *et al.*, 1990; Heinz *et al.*, 2004). Low levels of D2 receptors were linked to an impairment of re-evaluating decisions via the prefrontal cortex after negative feedback (Frank *et al.*, 2004; Goto & Grace, 2005). Recent evidence from an animal model indicates that chronic alcohol-induced malfunction of specifically mPFC D2/D4 receptors disrupts flexible behavioral adaptation (Trantham-Davidson *et al.*, 2014), which fits neatly to the presented findings. Interestingly, a behavioral study in humans showed that genetic variability in dopaminergic neurotransmission relates to perseveration during reversal learning (den Ouden *et al.*, 2013) also supporting the view that

dopamine could at least partially account for the behavior observed in alcohol-dependent patients.

Limitations. Whether diminished use of abstract inference arises as a consequence of chronic alcohol consumption or reflects a predisposition factor for the development of addictive behavior cannot be elucidated by a cross-sectional design. Groups differed in terms of general cognition, smoking status and grey matter density, even though our results were robust when adjusting for these variables. Cross-sectional studies in at-risk-populations (Ersche *et al.*, 2010) and longitudinal designs are warranted to track the balance of behavioral control systems across different stages in the development of addiction. It is to be noted that our model was not able to capture one specific aspect of the observed choice behavior; whereas we observed a group by phase interaction effect on actual correct choices, model simulations did not show this effect. Remarkably, however, apart from this aspect, all choice data effects could be replicated on the simulated data and measures of absolute model fit indicated that the models applied in this study provided a good account for the observed behavior.

Summary. In conclusion, alcohol-dependent patients showed a deficit to integrate abstract inference regarding alternative choice options after punishment in their decisions. Our data provide first neuroimaging support for reduced coding of abstract inference in the mPFC - a key region for flexible behavioral control - underlying this deficit. The same mPFC signatures were negatively related to obsessive-compulsive drinking habits. The *computational psychiatry* account applied here improves our understanding of the perplexing question of why addicted individuals continue drug consumption despite negative consequences.

6 Study 2: Impaired flexible reward-based decision-making in binge eating disorder: evidence from computational modeling and functional neuroimaging[3]

6.1 Introduction

A hallmark of binge eating disorder (BED), recently recognized in DSM-5 (American Psychiatric Association, 2013), is a lack of behavioral control during recurrent binge eating episodes, i.e., the feeling that one cannot stop eating. As propounded by the diagnostic criteria, excessive food intake takes place despite psychological and physical consequences such as feelings of guilt, shame or remorse, and high risk for obesity and associated maladies. To put it differently, patients suffering from BED make disadvantageous decisions and fail to adapt their behavior in the face of negative consequences. The first few behavioral studies are in accordance with the idea of decision-making impairments in BED (Davis et al., 2010; Svaldi et al., 2010; Danner et al., 2012; Voon et al., 2015).

Healthy individuals guide their decisions by prediction errors generated from observed outcomes, indicating that this outcome is better or worse than expected. These prediction errors were defined as "model-free" because they neglect the structure of the environment and simply lead to a repetition of previously reinforced actions, but render adaptation in dynamic environments slow (Dayan & Daw, 2008). When decisions made in the past have indeed turned out to be rewarding, individuals can exploit this experience for maximal gain. As is in everyday life, environmental conditions are frequently probabilistic and dynamic, challenging the individual to explore new alternatives at the right time (Daw et al., 2006; Cohen et al., 2007; Frank et al., 2009; Hare, 2014). Interestingly, prediction error signals do not only exist for options actually chosen, but can also process information on alternative choice options: this results in more complex prediction errors incorporating "what might have been". That is,

[3] This chapter corresponds to the following article: Reiter, A.M.F., Heinze, H.J., Schlagenhauf, F. & Deserno, L. (under review). Impaired flexible reward-based decision-making in Binge Eating Disorder: evidence from computational modeling and functional neuroimaging.

inference about unchosen options and their fictive outcomes (Lohrenz et al., 2007; Boorman et al., 2009). Thus, concurrent tracking of chosen and unchosen decision options helps to balance exploration and enables flexible goal-directed behavior.

Recent important studies indeed linked BED to impairments in making goal-directed decisions based on an internal mental model of the environment and to biases towards exploratory behavior (Morris *et al.*, 2015; Voon *et al.*, 2015). However, neural learning signatures of impaired flexible behavioral control in BED remain to be elucidated. A key brain region for flexible goal-directed behavior is the medial prefrontal cortex (mPFC). Studies in healthy participants have emphasized its important role in selecting reward goals and monitoring the value of actions (for review see Rushworth *et al.*, 2011). The mPFC is deemed responsible for on-the-fly valuation processes, which rely on the incorporation of environmental structure (Hampton *et al.*, 2006; Glascher *et al.*, 2009; Wunderlich *et al.*, 2012a). To our knowledge, this is the first task-based functional Magnetic Resonance Imaging (fMRI) study investigating neural learning signatures of impaired flexible decision-making in BED.

Therefore, we adopted a Computational Psychiatry (Montague *et al.*, 2012; Stephan & Mathys, 2014) approach to investigate mechanisms of behavioral adaptation and associated neural signatures in BED by combining computational modeling with fMRI during a dynamic counterfactual choice task. We aimed to elucidate the hypothesized impairment in flexible behavioral adaption of BED patients via the application of reinforcement learning (RL-) models. Regarding fMRI data, we first studied neural correlates of model-free prediction errors. Second, we investigated flexible behavioral adaptation via neural signatures incorporating inference about unchosen options and expected these signals, as well as between-group differences, to be coded in mPFC. In addition, we tested neural correlates of exploratory choices inspired by the behavioral finding in BED.

6.2 Materials and methods

Participants. Twenty-two BED patients and 22 healthy control subjects (HC) were recruited. For recruitment, exclusion criteria and neuropsychological testing see SI-2. Group characteristics are

summarized in S-Tables 2-1 and 2-2. Participants provided written informed consent and were paid on an hourly basis. The local Ethics committee approved the study.

Task. During fMRI, participants performed 160 trials of a counterfactual decision-making task designed to examine flexible behavioral adaptation. Participants decided between two cards distinguishable by two different abstract stimuli (Figure 6-1A). One of the cards had a reward probability of 80% and a punishment probability of 20% (vice versa for the other card). In this way, the task implied a simple higher-order structure: task structure was counterfactual as reward probabilities of the two decision options were anti-correlated. Whenever card A was a good choice, card B would be a bad choice and vice versa. Even though the outcome for the unchosen option was never shown, from the experienced value of one stimulus the hypothetical value of the other stimulus ("what might have happened") could be deduced by abstract inference on the anti-correlated task-structure. The task required flexible behavioral adaptation as reward contingencies remained stable for the initial 55 trials (initial block, pre-reversal), then, changed four times after 15 or 20 trials (middle block, reversal) and became stable again for the last 35 trials (last block, post-reversal). Contingency reversals were independent of participants' choices (Figure 6-1B). For details on timing, see SI-2.

Behavioral raw data analysis. First, correct choices were defined as choosing the stimulus with 80% reward probability and analyzed using repeated-measures analysis of variance (rm-ANOVA: within-subjects factor phase [pre-reversal, reversal, post-reversal phase], between-subjects factor group [BED vs. HC]). Second, switching behavior as a function of the outcome in the preceding trial was analyzed using rm-ANOVA (within-subject factor outcome [win vs. loss], between-subjects factor group). Third, we analyzed perseveration, defined as how often participants repeated choices for one stimulus despite two consecutive losses after having chosen this stimulus in the two preceding trials relative to all loss trials. Data analyses were performed using MATLAB R2012**,** IBM SPSS Statistics for Windows, Version 22 and R.

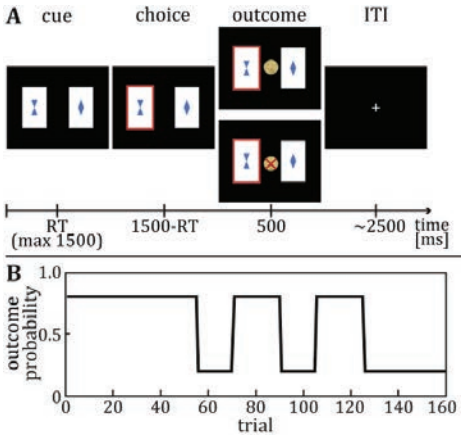


Figure 6-1. Counterfactual decision-making task. A) Exemplary trial sequence. Binary choice task: participants were instructed to choose the card that they thought would lead to a monetary reward. After the participant had selected one stimulus this card was highlighted and feedback was displayed outcome stimuli were a 10 Eurocent coin (in the case of a win) and a crossed 10 Eurocent coin (in the case of a loss). B) One of the stimuli had a reward probability of 80% and a punishment probability of 20% (vice versa for the other stimulus). Reward contingencies were stable for the first 55 trials (pre-reversal block) and also for the last 35 trials (post-reversal block). In the intermediate block, reward contingencies changed four times (reversal block).

Computational Modeling. We used computational modeling to analyze choices and to examine group differences in decision-making. The tested model space included four variations of RL-models. These models update expectations via reward prediction errors (RPEs), which quantify the mismatch between actual outcome and prediction. Model-free RPEs are only computed for chosen stimuli, although RPEs can also be computed for the unchosen stimulus (Lohrenz *et al.*, 2007; Boorman *et al.*, 2009). In line, the first three RL-models applied here differ in the degree of updating values for chosen and unchosen stimuli: I) a model-free learner updating values for the chosen stimulus. This neglects the counterfactual task-structure. We refer to this model as the single-update (SU) model; II) a learner updating values of chosen and unchosen stimuli to the same extent, thus, using abstract inference of the task structure. We refer to this model as the full double-update (DU) model; III) a model that individually weights the degree of double-update learning thereby accounting for inter-individual variability regarding this type of

abstract inference via the parameter κ. We name this the individually weighted double-update (iDU) model. In previous studies, it was suggested that behavior in probabilistic reversal learning tasks might be explained by a RL-model with a dynamic learning rate (estimating learning rates on a trial-by-trial basis, e.g. Krugel et al., 2009; Hauser et al., 2014a). To test this, we additionally included the Sutton-K1 model (Sutton, 1992), which updates the learning rate dynamically as a function of the change in prediction errors.

For all four RL-models, we translate values into actions using a Softmax rule including the parameter β, which estimates how tightly decisions are influenced by the contrast of values between the alternatives. Higher β values indicate that decisions are influenced more by relative value (exploitative), whereas with lower β estimates, decisions are more stochastic, thus explorative (Daw *et al.*, 2006; Cohen *et al.*, 2007). In total, seven models were compared: SU, DU, iDU, each with one learning rate or separate learning rates for rewards and punishments and the Sutton-K1 model. For equations and model fitting, see SI-2 and S-Table 2-3, 2-4, 2-5 and 2-6.

Model selection. The aim of model selection is to define a model that accounts best for the behavior in each group. Log-evidences (see SI-2) for each model and participant were subjected to random-effects Bayesian Model Selection (BMS, spm_BMS in SPM8, www.fil.ion.ucl.ac.uk/spm/ Stephan et al., 2009) to determine Expected Posterior Probability (PP) and Exceedance Probability (XP) for each model. XPs describe the probability that PPs of one model exceed that of another model in the comparison set. BMS was run for all subjects together and for each group separately to account for the possibility that the groups differ in best-fitting models.

Statistical analysis of fMRI data. See SI-2 for fMRI acquisition and preprocessing. We applied the general linear model approach (SPM8) for an event-related analysis. At the first level, onsets of feedback were entered into the model and modulated parametrically by two trial-by-trial regressors, which were constructed by using each individual's set of best-fitting parameters. The following regressors were computed: 1) model-free RPE_{SU}: RPEs for chosen values as computed on basis of the SU-Model with $\kappa = 0$; 2) more complex RPE_{DU}: RPEs for chosen values were

computed based on the DU-Model with $\kappa = 1$. We computed the difference of RPE_{DU} *minus* RPE_{SU} accounting for co-linearity between the regressors; also see Daw *et al.*, 2011 for such an implementation. This difference in chosen values from the DU- and SU-algorithms is given by κ in the iDU-algorithm. Thus, this regressor incorporates inference about alternative choices in RPEs. Throughout the manuscript, this second parametric modulator—the difference regressor—is referred to as RPE_{DU}. Building on the behavioral finding of elevated stochastic behavior in BED, each individual's trial-by-trial choice-probabilities from the decision-model were classified according to whether the actual choice was indeed the one predicted by the model to have the highest choice probability (exploitative), or the one with a lower choice probability (exploratory). Next, we added the onsets of cues to the first-level model of fMRI data described above with trial-type (binary: exploitative vs. exploratory) as first and choice probabilities as second parametric modulators. Compare Daw *et al.*, 2006 for the same implementation of an analysis regarding explorative versus exploitative trials. Onset of outcomes with RPE_{SU} and RPE_{DU} remained in the model to partial out their influence. Missing trials were modeled separately. The six realignment parameters, the first temporal derivative of the translational realignment parameters and a further regressor censoring scan-to-scan movement >1mm were included in the analysis to account for residual effects of motion.

At the second level, the contrast images of RPE_{SU} and RPE_{DU} were entered to a full-factorial ANOVA with the type of RPEs (RPE_{SU}/RPE_{DU}) as within-subject factor and group as between-subject factor. For contrast images regarding exploration, an independent-sample t-test (exploratory vs. exploitative trial type) was calculated. Results are accepted as significant at $p<.05$ using Family-Wise-Error (FWE) correction for the whole brain for task effects across all participants; for between-group comparisons, the peak voxel of the conjunction of both types of RPE (RPE_{SU} /RPE_{DU}) across the entire sample, surrounded by a 15mm sphere, was used to correct for multiple comparisons (significance threshold of p-FWE<.05 based on this search volume, S-Figure 2-1A). A similar search-volume was constructed for the neural analysis of

exploration by surrounding whole-brain peak voxels across all participants in bilateral insulae each with 15mm spheres in one volume (S-Figure 2-1B).

6.3 Results

Choice behavior. Rm-ANOVA on correct choices, including the within-subject factor phase (pre-reversal, reversal, post-reversal) and the between-subject factor group (BED vs. HC) showed main effects of phase ($F(2,42)=34.64$, $p<.001$) and of group ($F(1,43)=5.72$, $p=.02$, Figure 6-2A) but no significant interaction ($F(2,42)=.17$, $p=.79$). Switching behavior as a function of the outcome in the preceding trial was analyzed using rm-ANOVA (within factor outcome (win vs. loss), between-subject factor group (BED vs. HC). This revealed a main effect of outcome ($F(2,42)=288.93$, $p<.001$), and a main effect of group ($F(1,43)=8.75$, $p=.005$, Figure 6-2B), but no interaction ($F(2,42)=.11$, $p=.74$). Irrespective of the outcome in the previous trial, BED patients switched choices more frequently. Further, an independent t-test did not indicate any difference between groups in repeating choices for one stimulus despite two consecutive losses after having chosen this stimulus in the two preceding trials ($mean_{BED}=0.11\pm.07$, $mean_{HC}=0.10\pm.07$, $t(43)=0.18$, $p=0.86$).

Computational Modeling: model selection. BMS across all participants revealed that iDU-models provided the best account for observed choices peaking for iDU with one learning rate (iDU XP=.60, iDU-WL XP=.30, Table 6-1). In all subsequent analyses, we thus use parameters derived from this model. When running BMS for both groups separately, iDU models were clearly winning in the control group, whereas results were more ambiguous in the BED group (Table 6-1) indicating pronounced heterogeneity in the BED group.

Computational Modeling: parameter comparison. Independent sample t-tests with Bonferroni-correction (adjusted $p=.017$) to compare three modeling-derived parameters between groups (decision parameter β, learning rates for chosen values α_c, and learning rates for unchosen values α_{uc}, as product of κ with α_c) revealed a significant group difference for the decision parameter β ($t(43)=2.51$, $p=.016$, Figure 6-2C), but not for any other parameter ($ts<.81$, $ps>.42$).

Table 6-1 Model Selection. Exceedance Probabilities (XP) for all models. SU: single-update, DU: double-update, iDU: inidividually-weighted double-update, WL indicates that the model had separate learning rates for wins and losses. For expected posterior probabilities and XP compare S-Table 2-4.

	SU	SU-WL	DU	DU-WL	iDU	iDU-WL	Sutton K-1
Overall (n=44)	.054	.001	<.001	.039	.601	.303	<.001
HC (n=22)	.035	.005	.003	.112	.321	.522	<.001
BED (n=22)	.210	.057	.030	.033	.426	.150	.094

A lower decision parameter β indicates a higher degree of stochastic choices unrelated to the current choice value, i.e., lower values in BED indicate exploratory rather than exploitative choices. Importantly, when excluding two patients who were not fit better than chance by the model (for definition, see SI-2), the difference remained significant ($t(43)=2.14$, $p=.039$). Thus, the significantly lower decision parameter β did not result from simply poor fit per se in the patient group.

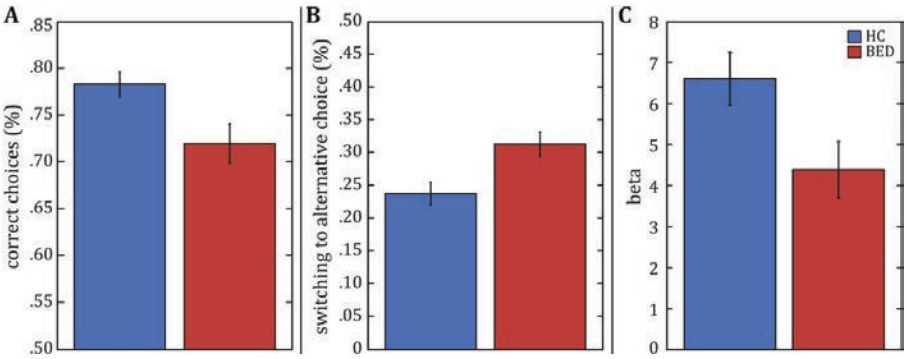


Figure 6-2. Behavioral Results. A) Raw Data Results. Correct choices differed significantly between groups ($t(43)=3.48$, $p=.001$, left panel). B) BED patients showed enhanced switching behavior between the two stimuli. C) Comparison of modeling parameters revealed that BED patients had a lower decision parameter β. Lower values of β indicate a higher degree of stochastic choices unrelated to the current choice values. Hence, lower values in BED indicate enhanced exploratory choices.

Neural RPE processing: entire sample. We aimed to explore neural signatures of simple and more complex (including abstract inference components) RPE processing in BED versus HC. Thus, we analyzed encoding of RPEs for chosen values as a function of SU- versus DU-learning, that is, RPEs derived from the SU-Model (RPE_{SU}) versus RPEs derived from the DU-Model (RPE_{DU}). See S-Table 2-7 and Figure 6-3A for results.

We observed whole-brain correctable activation for model-free RPE_{SU} in, e.g., bilateral ventral striatum as well as orbitofrontal/medial prefrontal cortex, amygdala, right hippocampus, right putamen, and posterior cingulum (p-$FWE_{wholebrain}$<.05). RPE_{DU} elicited activation in similar regions, including bilateral ventral striatum and mPFC (p-$FWE_{wholebrain}$<.05). The conjunction of RPE_{SU} and RPE_{DU} reached significance peaking in the mPFC (X=-6, Y=52, Z=-12, t=5.27, p-$FWE_{wholebrain}$ =.03).

Neural RPE processing: group comparison. Regarding model-free RPE_{SU} processing, we did not observe significant between-group differences in the processing of RPE_{SU} based on the mPFC search-volume (peak-coordinates: X=-10, Y=60, Z=-10, t_{peak}=2.17 p_{peak}= 0.74). There was no significant group difference in other regions at a liberal threshold (cluster level k=10, p<.001 uncorrected). To investigate between-group differences in coding of complex abstract inference RPE signatures, we tested for a type of RPE (RPE_{SU}/RPE_{DU}) x group interaction. This interaction was significant in mPFC (X=-12, Y=40, Z=-6, t=4.00, FWE-corrected for mPFC p=.03). As post-hoc contrast, we compared RPE_{SU} and RPE_{DU} between groups. We observed significantly reduced coding of RPE_{DU} signatures in BED in the mPFC (X= -12, Y=40, Z=-6, t=4.06, FWE-corrected for mPFC p=.02, Figure 6-3B and Figure 6-3C).

Next, we tested for an association of neural coding of RPE_{DU} and choice behavior: parameter estimates at the peak-coordinate of the group difference in mPFC (X= -12, Y=40, Z=-6) for RPE_{DU} were extracted and, for both groups separately, correlated with behavioral performance (percentage of correct choices, percentage of switching). One outlier (z-value of parameter estimates<-2.8) in the BED group was removed beforehand. We found a significant positive association between the neural RPE_{DU} signature and correct choices in BED (r(19)=.60,

p=.005) as well as in HC (r(20)=.53, p=.02). Indeed, with group as a covariate, the correlation between RPE_{DU} and correct choices was significant (r(2,39)=.55, p<.001). The association of the RPE_{DU} signature and switching was significantly negative in HC (r(20)=-.47, p=.03), but non-significant in BED (r(19)=-.24, p=.32). Across both groups, when controlling for group, the negative correlation between RPE_{DU} and switching was significant (r(2,39)=-.35, p=.03). No moderation effect of group on the association of neural signature and behavioral performance was found (t<1.41, p>.17, R^2 change due to moderator<.03).

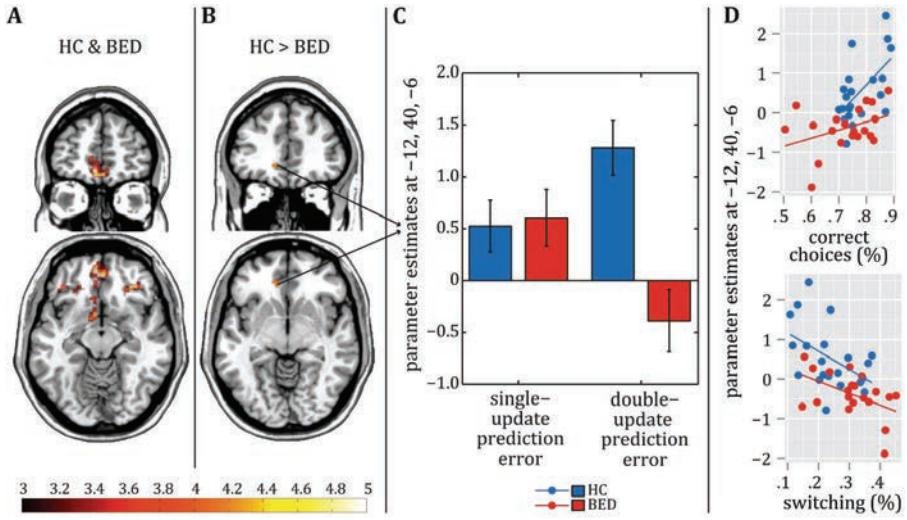


Figure 6-3. Neural correlates of Single-Update and Double-Update Prediction Error Processing. A) Across both groups, peak conjunct activity elicited by RPESU and RPEDU was observed in the medial prefrontal cortex (mPFC, X=-6 Y=52 Z=-12, t=4.86, p-FWE for the whole brain = .03, see S-Table 2-7) B) and C) Comparing RPE_{SU} and RPE_{DU} between groups revealed significantly reduced coding of RPEDU signatures in BED in the medial prefrontal cortex (X= -12, Y=40, Z=-6, t=4.06, FWE-corrected for mPFC p=.02). For display purposes, threshold is set at p<.001, cluster level k=10. D) Parameter estimates at the peak coordinate of the group difference was positively correlated with correct choices in the decision-making task and negatively related to switching behavior in the decision-making task, independent of group.

These findings suggest that neural coding of RPE_{DU} is related to better behavioral performance and less switching, in both healthy individuals and BED (see Figure 6-3D for an illustration).

Neural correlates of the exploration-exploitation trade-off. Building on the observation of enhanced exploration in BEDs' choices, we compared activity elicited in exploratory vs. exploitative trials using an F contrast. This revealed peak activation in bilateral anterior insula/ventro-lateral prefrontal cortex (aI/vlPFC, thresholded at $p\text{-FWE}_{wholebrain}<.05$, Figure 6-4A, S-Table 2-8), due to higher activation for exploratory vs. exploitative trials. A between-group effect was revealed in the right aI/vlPFC. BED showed significantly lower activation for explorative trials compared to HC (X=44, Y=22, Z=-12, t=4.42, FWE-corrected for aI/vlPFC, p=.04, Figure 6-4B and Figure 6-4C). We did not find a significant association of aI/vlPFC activation with behavioral performance (For details, SI-2).

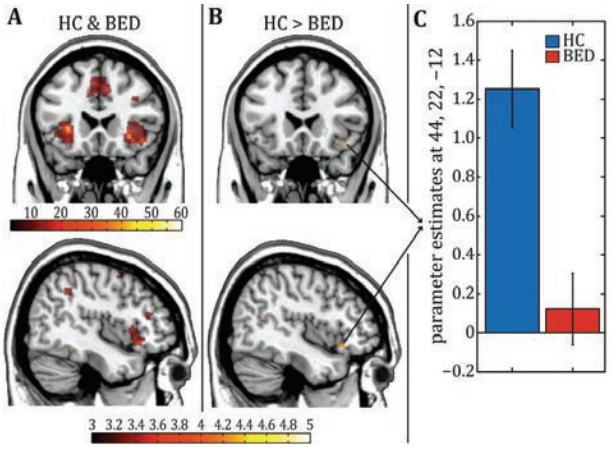


Figure 6-4. Neural correlates of the exploration-exploitation trade-off. A) Across both groups, explorative trials vs. exploitative trials were associated with bilateral activation of anterior insula/ventro-lateral prefrontal cortex (see also S-Table 2-8). B) and C) Comparing activation in explorative vs. exploitative trials between groups demonstrated that BED patients show significantly diminished activity in the aI/vlPFC during explorative trials (X=44, Y=22, Z=-12, t=4.42, FWE-corrected for aI/vlPFC, p=.04). For display purposes, threshold is set at p<.001, cluster level k=10.

6.4 Discussion

This fMRI study combined with computational modeling provides novel insight into the neural learning signatures of maladaptive decision-making in BED, thereby helping to refine a neurocognitive phenotype of the newly classified disorder. Firstly, whereas we found compelling evidence that healthy controls used inference on alternative choices to guide decision-making, this was ambiguous for BED. Relatedly, patients showed reduced encoding of learning signatures incorporating alternative choice options in the mPFC. Secondly, we found decision-making in BED to be characterized by enhanced switching between choices, thus, a bias towards explorative decisions during behavioral adaptation in a dynamic environment. Parallel to this behavioral observation, BED was characterized by less aI/vlPFC activation during exploratory decisions.

Reduced incorporation of abstract inference learning signals. According to BMS, HCs convincingly integrated abstract inference on alternative choices into decision-making, to use "what might have happened" when making choices, whereas this was ambiguous for BED patients. In consonance, coding of RPEs incorporating abstract inference on alternative options was reduced in mPFC of BED patients. In healthy individuals, concurrent tracking of multiple decision possibilities and their potential consequences contributes to a mental map of the environment and thus promotes flexible goal-directed behavior in dynamic environments (Hampton *et al.*, 2006; Lohrenz *et al.*, 2007; Abe & Lee, 2011; Takahashi *et al.*, 2013). In the present study, mPFC learning signatures incorporating inference on alternative options were indeed associated with successful choices and less explorative switching behavior. Thus, the specific reduction in mPFC signaling could be one explanation for the shift from goal-directed towards habitual behavior in BED as reported previously (Voon *et al.*, 2015). Corresponding to this conclusion, the latter study also found an association of reduced goal-directed behavior with reduced gray matter at nearby coordinates in BED (Voon *et al.*, 2015).

Disadvantageous bias towards explorative choices in BED. While clinical characteristics and diagnostic criteria suggest impaired mechanisms of flexible behavioral adaptation as crucial for

BED, systematic investigations of this impairment are scarce. Our findings suggest that BED is not characterized by learning deficits per se, suggested by absent differences between learning rates and intact model-free neural learning signals. Rather, patients suffering from BED did not exploit a relatively good option as consistently as controls but show pronounced exploratory switching behavior as also confirmed by the computational modeling analysis (lower decision parameter β). Although it is obviously advantageous for an individual to explore alternatives in a changing environment, appropriate value tracking would result in choosing, and staying with, the most valuable option, when appropriate. In BED, this type of behavior was accompanied by less correct choices confirming that the amount of exploration was indeed suboptimal. Correspondingly, our observation is consistent with a very recent study, which found obese people with BED to be characterized by enhanced exploration in comparison to obesity not accompanied by BED (Diener, 2010).

In patients, the behavioral tendency to switch was paralleled by reduced activation during such exploratory choices in anterior aI/vlPFC, key regions implicated in reversing behavior (Cools *et al.*, 2002; Menon & Uddin, 2010). Thus, less activation in this region might hinder the individual to get back on the right track after an exploratory try that has not resulted in positive benefits. Complimentary, enhanced anterior aI/vlPFC activation in explorative trials as observed in healthy controls could reflect a potential uncertainty or warning signal for these trials. In line with this idea, prior imaging studies have also reported aI/vlPFC activation during uncertainty prediction and when making a risky compared to a safe decision (Paulus *et al.*, 2003; Preuschoff *et al.*, 2008; Singer *et al.*, 2009). In this framework, aI/vlPFC activation could guide choices in moments of uncertainty. Therefore, the reduction of such signaling during exploratory decisions reduces awareness of the uncertain (or disadvantageous) character of these decisions. This might bias the individual towards more and suboptimal exploratory decisions instead of selecting a relatively good option based on accumulated experience.

Relevance to theories of addiction. A hallmark of BED, the maintenance of maladaptive behaviors despite negative consequences, closely resembles key criteria of substance addiction and a

current debate relates to the classification of BED as an addiction spectrum disorder (Smith & Robbins, 2013; Volkow *et al.*, 2013). However, a noteworthy albeit controversially discussed review (Ziauddeen *et al.*, 2012), issues caveats against a premature adoption of the food addiction model: the article deems functional imaging attempts to profile BED as insufficient to date and calls for task-dependent measurements based on cognitive-neuroscience models in order to relate behavioral and cognitive phenotypes to neuroimaging findings. The study at hand is one step in this direction. The adopted *Computational Psychiatry* approach enables estimation of specific parameters that provide mechanistic accounts of functioning in one or another cognitive domain (Wiecki *et al.*, 2015) and informs the modeling-based fMRI analysis of neural learning signatures (Stephan & Mathys, 2014).

Limitations. The cross-sectional design precludes any conclusions on causality. The study a priori focused on the mPFC. However, the finding of between-group differences in mPFC and ai/vlPFC associated with different decision signatures asks whether it is one of the two regions or their interaction that mediates the observed behavioral changes. Interestingly, mPFC signals were clearly associated with behavioral performance. Lesion studies in animals and translation to humans, e.g. via brain stimulation techniques, could elucidate this question.

Conclusions. Taken together, this study provides insight into specific impairments in reward-guided decision-making in BED. That is, a disadvantageous behavioral bias towards explorative decisions accompanied by less coding of explorative trials in the aI/vlPFC as well as diminished representation of more complex RPEs in mPFC. By adopting a Computational Psychiatry approach combined with modeling-informed fMRI analysis, this study contributes to refining the neurocognitive phenotype of BED in addition to clinical observations and new diagnostic criteria in DSM-5.

7 Study 3: The Feedback-Related Negativity codes components of abstract inference during reward-based decision-making[4]

7.1 Introduction

A core function of human decision-making is to evaluate the motivational significance of ongoing events in order to weigh up different decision options and to guide future decisions accordingly. There is a growing consensus that individuals make decisions by computing decision values of potential options which are then compared in order to make a choice (Rangel & Hare, 2010; Rushworth *et al.*, 2011; Sokol-Hessner *et al.*, 2012). It is thereby indispensable for an agent to keep track of choice values of options that were actually taken. This is achieved via updating chosen values by reward prediction errors (RPE), which result from comparing the observed outcome to what has been expected. However, the idea that an agent also learns from 'what might have happened', that is, abstract counterfactual inference regarding unchosen options and their fictive outcomes, has recently sparked considerable interest (Boorman *et al.*, 2009; Boorman *et al.*, 2011; Fischer & Ullsperger, 2013). Concurrent coding of multiple actions and their outcomes is suggested to enhance the efficiency of reinforcement learning (Abe & Lee, 2011; Boorman *et al.*, 2011; Takahashi *et al.*, 2013) and disturbed abstract inference on 'what might have happened' has been implied in psychopathological states, e.g. in animal models of addiction (Lucantonio *et al.*, 2012; Lucantonio *et al.*, 2014) . In the present study, we employ a probabilistic decision-making task and focus on simultaneous updating of chosen and unchosen decision values: In this task, decision values were anti-correlated such that the drop of one value directly implied the rise of the other. Thus, decision values for the options at hand can be inferred by action-reward pairings but also by abstract inference based on the anti-correlated task structure (Bromberg-Martin *et al.*, 2010).

The standard reinforcement learning account only updates the value of a chosen stimulus via a RPE. This is referred to as a 'model-free' RPE because it neglects the structure of

[4] This chapter corresponds to the following article: Reiter, A.M.F., Koch, S.P., Schröger, E., Hinrichs, H. Heinze, H.J., Deserno, L. & Schlagenhauf, F. (in revision). The Feedback-Related Negativity codes components of abstract inference during reward-based decision-making.

the environment, here, the anti-correlation of the two values. Even though no outcome is delivered for the unchosen option, via abstract counterfactual inference on the reward statistics of the experimental environment, the agent could make use of the task structure and concurrently 'double-update' the values of chosen and unchosen options. The 'model-free' RPE is thereby modified by an abstract inference component. Double-updating optimizes learning even in simple, but dynamic binary choice environments (Hampton *et al.*, 2006; Li & Daw, 2011; Dolan & Dayan, 2013; Schlagenhauf *et al.*, 2013).

In the current study, we aimed to define electrophysiological signatures of incorporating such abstract counterfactual inference on the task structure into the decision-making process. In a pioneering theory, Holroyd and Coles proposed that model-free RPEs linearly scale with the so-called Feedback-Related Negativity (FRN, Holroyd & Coles, 2002). The FRN is an event-related potential peaking at frontocentral electrodes about 200–300ms after the delivery of feedback. While a wealth of studies has tackled the relationship between FRN and feedback learning (for a review see Walsh & Anderson, 2012), most studies relied on cross-trial averages in order to mirror learning processes on an electrophysiological level. As learning is a dynamic process over time, a trial-by-trial approach seems well suited to capture these dynamics. So far, only few EEG studies have adopted such a parametric design by linking modeling-derived trial-by-trial learning signatures to single-trial amplitudes of event-related potentials (ERP; Philiastides *et al.*, 2010; Chase *et al.*, 2011; Fischer & Ullsperger, 2013; Hauser *et al.*, 2014). Chase et al. (2011) reported a correlation of the FRN with model-free RPEs derived from a reinforcement learning model that only updates values of the chosen stimuli ('single-update model'). A recent study by Hauser *et al.* (2014) used a modified algorithm that simultaneously updates values of chosen and unchosen stimuli ('double-update model') and replicated the correlation of RPEs and the single-trial FRN. Additionally, the authors studied the influence of unsigned PEs, a value-unspecific signal reflecting the unexpectedness of events and found that unsigned PEs better accounted for their data. Hauser *et al.* (2014) concluded that the FRN codes surprise rather than a signed learning signal. Notably, both studies have not systematically

separated uniquely explained variance by single-update versus double-update algorithms in the FRN.

Thus, the goal of the present study was to identify correlates of abstract inference on the anti-correlated task structure in the FRN. In order to address this question, we recorded EEG from 21 healthy participants while they performed counterfactual decision-making in a dynamically changing environment with anti-correlated reward probabilities. We evaluated the relation between single-trial FRN responses to feedback and RPEs, respectively. The following hypotheses were probed: (1) The FRN signals a model-free RPE as suggested by prior studies. (2) The FRN additionally signals values estimated through abstract inference on the anti-correlated task structure as derived from the difference between value estimates of double-update and single-update algorithms.

7.2 Materials and methods

Participants. Twenty-one healthy participants (age: 26.20 +/- 3.49 years, range 22–34 years, 11 females) were recruited from the Max Planck Institute's subject database. Participants received remuneration for participation in addition to the money won in the experimental task. The study was approved by the ethics committee of the University of Leipzig. All subjects gave written informed consent to participate prior to the study. One dataset had to be excluded due to technical problems during EEG recording.

*Task***.** During EEG acquisition, subjects performed a probabilistic decision-making task in a dynamic environment with anti-correlated reward probabilities (Figure 7-1A) that requires flexible behavioral adaptation. In a total of 160 trials, participants decided between two cards both showing an abstract geometric stimulus (maximum response time: 1500ms). After the participant had chosen one stimulus using either the left or the right button, the selected stimulus was highlighted and depicted for another 500ms plus reaction time. Following an inter-stimulus interval of 1500ms (presentation of a fixation cross), one of two feedback stimuli was presented for 500ms, indicating either a win (a 10 cent coin) or loss (a crossed 10 cent coin) of money. During the intertrial interval, a fixation cross was shown for a variable duration with a

mean of ~2500ms. Total trial duration was 6500ms on average. The location (right vs. left side of the screen) where each of the stimuli was presented was randomized over trials. Crucially, reward contingencies were perfectly anti-correlated, that is, one of the stimuli was assigned with a reward probability of 80% and a punishment probability of 20%; and vice versa for the other stimulus. This anti-correlation is essential because, if exploited by the agent, it enables the learner to draw inference on the value of the unchosen stimulus even though no outcome is delivered for the unchosen option.

Reward contingencies remained stable for the first 55 trials and also for the last 35 trials. In between, reward contingencies changed four times, after 15 or 20 trials, respectively. This required participants to flexibly adapt their behavior and ensured constant learning (Figure 7-1B). Note that contingency reversals were predefined by the stimulation protocol and did not depend on subjects' performance in the task. Prior to the experiment, participants were instructed that depending on their choice they could either win 0.10€ or lose 0.10€ per trial and that the total amount of money gained would be paid out at the end of the experiment. Additionally, participants were informed that one of the two cards had a superior chance of winning money but that this might change during the task. Participants became familiar with the task prior to the experiment by performing 20 training trials with different stimuli and without any reversal of reward contingencies. Distance between participant and computer screen was kept constant across all subjects.

Implemented models. The focus of the present study was to examine trial-by-trial signatures of each individual's learning process in the electrophysiological recordings. Thus, the actually observed behavioral responses were analyzed in a computational modeling framework. Three different types of learning models were fitted to the data: 1) a single-update model which updates values only for the chosen stimulus, 2) a double-update model which updates values of chosen and unchosen stimuli with equal weight and 3) a hybrid model which individually weights the degree of double-update learning.

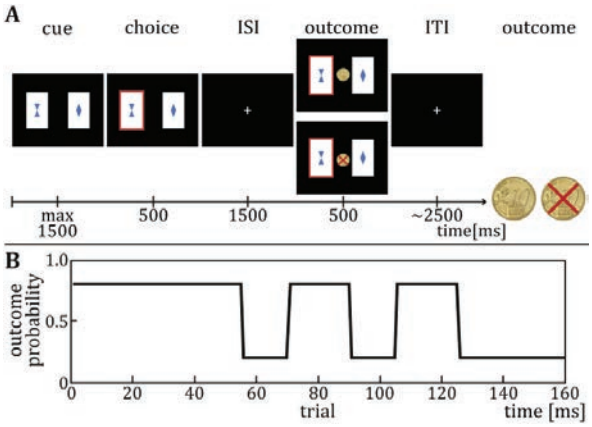


Figure 7-1. A) Trial sequence of the serial reversal task. Participants were instructed to find the card with the superior chance of winning. One of the stimuli was assigned with a reward probability of 80% and a punishment probability of 20% (vice versa for the other stimulus). Outcome stimuli were a 10ct coin (in the case of a win) and a crossed 10ct coin (in the case of a loss). B) Reward contingencies over the course of the experiment. Reward contingencies remained stable for the first 55 trials and also for the last 35 trials. In between, reward contingencies changed four times.

Single-update model. The single-update model updates a decision value $Q_{a,t}$ for the chosen stimulus via the RPE $\delta_{Q_{a,t}}$, which is defined as the difference between the received reward R_t and the expected reward for the chosen stimulus $Q_{a,t}$:

$$(8)\ \delta_{Q_{a,t}} = R_t - Q_{a,t}$$

This teaching signal is then used to iteratively update chosen decision values trial-by-trial:

$$(9)\ Q_{a,t+1} = Q_{a,t} + \alpha\delta_{Q_{a,t}}$$

Here, α represents a learning rate, which weights the influence of reward RPEs on the updated values. This free parameter of the model has natural boundaries between 0 and 1. Note that this model neglects the anti-correlated structure of the task by simply updating decision values for chosen stimulus only while the value of the unchosen stimulus $Q_{ua,t}$remains unchanged:

$$Q_{ua,t+1} = Q_{ua,t} \quad (10)$$

Double-update model. In a next step, we extended this model to more closely match the experimental environment by using abstract inference. In this task, an update of both decision values in each trial is advantageous because it takes into account the anti-correlated structure of the task. This double-update learning has previously been demonstrated to result in improved behavioral adaption to changing reward contingencies (Hampton *et al.*, 2006; Li & Daw, 2011; Schlagenhauf *et al.*, 2013). In order to mirror this strategy in our computational modeling approach, the unchosen decision values are updated using a different error signal. The prediction error for the double-update model is:

$$\delta_{Q_{ua,t}} = -R_t - Q_{ua,t} \quad (11)$$

The same learning rate α is used for updating the unchosen value:

$$Q_{ua,t+1} = Q_{ua,t} + \alpha\delta_{Q_{ua,t}} \quad (12)$$

We refer to this model as the double-update model as it takes into account the values of the chosen and unchosen stimuli and thereby implements abstract inference on the anti-correlated task structure.

Hybrid model. Note that equation 5 gives the same weight to the update of unchosen decision values as to the chosen decision values. However, it is conceivable that the unchosen option is updated with a reduced rate if change compared to the update of chosen values. Moreover, the degree of abstract inference may differ across individuals, as this is computationally more expensive and consequently may be limited depending on the specific situation that challenges the individual. This notion has been emphasized in healthy individuals (Daw *et al.*, 2011; Wunderlich *et al.*, 2012b; Eppinger *et al.*, 2013; Otto *et al.*, 2013b; Smittenaar *et al.*, 2013) and recently also in psychiatric disorders (Sebold *et al.*, 2014; Voon *et al.*, 2015). To account for this potential variability in employing abstract inference, we additionally tested a hybrid model by inferring the individual degree of double-update learning. Therefore, the weighting parameter κ is introduced:

$$(13) \qquad Q_{ua,t+1} = Q_{ua,t} + \kappa\alpha\delta_{Q_{ua,t}}$$

Note that equations 1 to 5 refer to the special cases $\kappa = 0$ or $\kappa = 1$, respectively.

Observation model. Finally, for all models, we transform decisions into action probabilities by applying a softmax equation including the parameter β, which determines the stochasticity of the choices:

$$(14) \qquad p_{(a,t)} = \frac{\exp(\beta Q(a))}{\sum \exp(\beta Q(a'))}$$

The sum of this probability $p_{(a,t)}$ over all trials (and subjects, eventually) is the so-called negative log-likelihood which is the probability of observing the data given the parameters of a model.

A priori model simulation. We simulated choice behavior 1000 times per model (Single-Update Model κ=0, Hybrid Model κ=0.5, Double-Update Model κ=1) by fixing all other free parameters to the mean estimate of an independent sample (n=35) performing the same task. The highest simulated mean of correct choices over the whole course of the experiment was observed for the hybrid model with κ=0.5 (mean simulated correct choices (%): single-update model=0.81, hybrid model=0.82, double-update model=0.79). The advantage of the hybrid model, brought about by implementing abstract inference on the task structure ('double-updating') becomes particularly prominent in trials with frequent reversals of reward contingencies (trials 56-125, compare Figure 7-1B) which require flexible behavioral adaptation (mean simulated correct choices (%): single-update model=0.71, hybrid model=0.77, double-update model=0.77).

Model fitting and model selection. Free parameters that have natural boundaries were fitted after transformation to a logistic (α) or exponential (β) distribution to render normally distributed parameter estimates. A maximum-a-posteriori estimate of each parameter for each subject was found by setting the prior distribution to the maximum-likelihood given the data of all participants and then Expectation-Maximization was used. For an in-depth description please compare Huys *et al.*, 2011; Huys *et al.*, 2012. All modeling analyses were performed using MATLAB R2013a (The MathWorks, Inc., Natick, Massachusetts, United States). For all three models, we first report the negative log-likelihood and the Bayesian Information Criterion (BIC,

Schwarz, 1978) based on the negative log-likelihood (Table 7-1). Note that the BIC includes a penalty term for the number of parameters in the model to account for the risk of overfitting. Second, the model evidence was approximated by integrating out the free parameters. The integral was approximated by sampling from the prior distribution and we therefore add the subscript 'int' to the BIC (Table 7-1; see Huys *et al.*, 2011; Huys *et al*., 2012).

Table 7-1. Model Selection. –LL: negative log-likelihood, BIC:Bayesian Information Criterion, BIC_{int}: Bayesian Information Criterion after integrating out the free parameters via sampling from the prior distribution, XP: Exceedance Probabilities after random-effects Bayesian model selection.

	-LL	BIC	BIC_{int}	XP
Full hybrid model (κ as a free parameter)	849	751	808	0.9872
	Δ-LL hybrid	Δ BIC hybrid	Δ hybrid	BIC_{int}
Single-update model ($\kappa = 0$)	-34	26	12	0.0127
Double-update model ($\kappa = 1$)	-69	57	55	0.0001

Third and reported in the results section, we subjected this integrated likelihood to a random-effects Bayesian model selection procedure, spm_BMS contained in SPM8, http://www.fil.ion.ucl.ac.uk/spm/ (Stephan *et al.*, 2009). After having identified the best-fitting model, we also verified that best-fitting parameters reproduce the observed behavior well.

Modeling-derived EEG analysis. In line with previous reports, we regressed single-update RPEs against single-trial EEG data (Chase *et al.*, 2011). To specifically address the question if the coding of RPEs in the FRN contains additional effects of abstract inference on the task structure as expressed in the double-update model, we computed a difference regressor (Daw et al., 2011), which was defined as the difference between the two error signals:

$$\delta_{a,t_{difference}} = \delta_{a,t_{double-update}} - \delta_{a,t_{single-update}} \quad (15)$$

Note that this regressor reflects differences of chosen decision values estimated by the single-update versus double-update algorithm.

Electrophysiological recording and ERP extraction. Electrophysiological activity was recorded from 60 scalp positions according to the 10-20 EEG system (BrainAmp MR plus, Brainproducts, Gilching, Germany).

Four additional ocular electrodes (vertical and horizontal electrooculogram; EOG) were attached to monitor eye movements and blinks. EEG and EOG, referenced against the linked mastoids, were sampled at 2500 Hz (1-sec low cut-off, 250-Hz high cut-off, Notch off). Electrode impedances were kept below 10 kOhm.

EEG data were preprocessed using EEGLAB 4.515 (Delorme & Makeig, 2004) and MATLAB R2013a. EEG data were down-sampled to 250 Hz and bandpass filtered between 0.5 and 46 Hz (butterworth filter, 3rd order). Trials were segmented from -2 to 6 sec relative to the onset of the outcome stimulus. An independent component analysis (ICA; logistic Infomax ICA, Delorme & Makeig, 2004) was applied to decompose the multivariate EEG signal into statistically independent components. By two independent assessors, movement related ICA sources and frontal sources with ocular artifacts such as blinks and eye movements were visually identified by inspecting the scalp maps, time courses and power spectra in all components and were removed before back-projection of the remaining components onto the EEG channels. All EEG epochs were visually inspected before and after ICA. Thereafter, data were re-referenced to the average (Lehmann & Skrandies, 1980).

For the analysis of the feedback-related electrophysiological responses, we identified the peak negativity at an a priori site of interest (Cz, according to Holroyd & Coles, 2002, Figure 7-2A) and in a predefined time window of 200–300ms after feedback onset (Hauser *et al.*, 2014b). We aimed to account for inter-individual differences in ERP latencies by determining individual peak latencies. The individualized latency of the evoked components was derived from the individual subject's average FRN latency. This individualized time point was then used to extract single-trial amplitudes in all 160 trials. Figure 7-2B shows the averaged waveform at the electrode Cz and the topological distribution of the deflections at the time point of the average peak FRN across subjects.

7.3 Results

7.3.1 Behavioral analyses

Task performance. Subjects chose the correct stimulus, i.e. the stimulus with the higher reward probability, on average in 81% (*SD*=6) of all trials, indicating that participants understood and mastered the task appropriately. A mean of 0.25 (*SD*=0.72) trials per subject had to be excluded for the computational modeling and single-trial EEG analysis due to missing responses.

Computational modeling. Three different computational models were implemented to describe different ways of updating decision values during the learning process. First, a single-update model, which updates chosen values only, and second, a double-update model, which additionally updates unchosen decision values. Compared to the second model, the third model quantified the degree of double-updating individually via the double-update weighting parameter κ. Bayesian model selection demonstrated that the hybrid model, a combination of both strategies quantified by the free parameter κ, explained the data best at the group level (Table 7-1). This suggests considerable inter-individual variability in the extent to which individuals use double-updating regarding the unchosen option. Importantly, choice behavior of all participants was explained better than chance by the best-fitting model (mean explained choices: 78%, *SD*: .093) considering the negative log-likelihood. Choices explained are in a similar range as in prior studies (Daw *et al.*, 2011) and suggest that the winning model accounted well for the observed choice data. A simulation based on each individual's inferred parameters additionally showed that our model captured behavior well. The distribution of the best fitting parameters and the negative log-likelihood is shown in Table 7-2.

Table 7-2. Distribution of best-fitting parameters (hybrid model)

	β	Initial Q	α	κ	-LL
25th percentile	1.96	-.36	.53	.07	-54
Median	2.80	-.24	.57	.10	-33
75th percentile	3.80	-.07	.60	.17	-27

7.3.2 Correlation of FRN and learning signatures

Coding of model-free RPEs in the FRN. In line with previous findings, we evaluated whether FRN amplitudes correlate with model-free RPEs. The relationship between the FRN amplitude and the magnitude of the model-free, single-update RPE was evaluated on a trial-by-trial basis using linear regression analysis. We found a significantly positive correlation between this RPE and FRN amplitudes (mean slope of regression=.086, *SD*=0.116, $t(19)=3.31$, $p=.01$). In line with previous findings, this indicates that RPEs scale linearly with FRN amplitude, i.e. that the FRN codes model-free RPEs

Coding of abstract double-update inference in the FRN. In the next step, we directly aimed to determine the coding of single-update versus double-update components on the FRN-amplitudes. Therefore, in addition to the model-free, single-update RPE regressor, we also entered a difference regressor between single-update and double-update RPEs into the multiple regression analysis. Note that the difference regressor as described in equation (15) describes differences of decision values estimated by double-update minus the single-update algorithm and thus represents the change in values uniquely associated with abstract double-update inference. We found a significant effect for both the single-update RPE regressor ($t(19)=2.32$, $p=.032$) and this difference regressor (mean slope of regression: -.085, *SD*=.110), $t(19)=-2.30$, $p=.033$). Note, that the correlation with the model-free, single-update RPE is positive, indicating that RPEs scale with FRN amplitudes (see Figure 7-3B) whereas the correlation of FRN with the difference regressor is negative. The latter negative correlation reflects two key characteristics of abstract double-update inference:

1) By concurrently updating the unchosen choice option, double-updating maps the anti-correlated environment more precisely. This leads to differences in the size of RPEs from the double-update (RPE_{DU}) vs. the single-update model (RPE_{SU}), as for the double-update learner, in certain cases, feedback is more predictable than for the single-update learner. This leads to a relatively attenuated RPE_{DU} in trials where RPE_{SU} has high absolute values. More detailed, in the

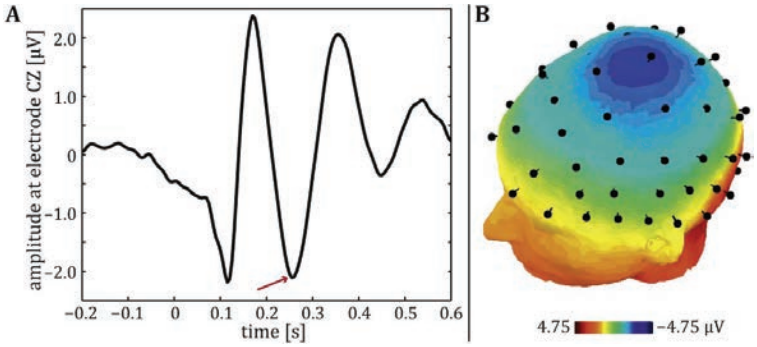


Figure 7-2. A) Grand average waveform of the FRN, revealing the FRN with its maximal amplitude 261ms after feedback onset at the electrode CZ (here indicated by the arrow. B) Topological distribution of the deflections at the time point of the average peak FRN.

case of unexpected punishments (defined by large negative RPEs), the RPE from the double-update model is less negative than the RPE from the single-update model, thus the difference regressor is *positive*. Contrary, in cases of relatively unexpected rewards (defined by large positive RPEs) - often after an agent has switched to the alternative option - the RPE from the double-update model is less positive than the RPE from the single-update model; consequently, the difference regressor becomes *negative*.

2) Double-update learning is smoother as recent events do not impact choices values as strongly as in single-update learning. Thus, after relatively expected punishments (e.g., after a series of punishments, indicative of the necessity to switch to the alternative option), RPE_{SU} approaches zero faster and RPE_{DU} is more negative than RPE_{SU}. This results in a *negative* difference regressor. Contrary, in a rewarded trial, which has already been preceded by a series of rewards (e.g., after having learnt to stay with the better stimulus at one point in time), RPE_{DU} is numerically higher than RPE_{SU}. Thus, the difference regressor is positive in these cases. Figure 7-3 illustrates this description by plotting mean RPE_{SU} and RPE_{DU} as well as difference regressor and mean FRN amplitudes as a function of expectedness and valence of the feedback in the trial.

Our observations suggest that in addition to the model-free, single-update RPE, also the difference regressor, which reflects the degree of abstract double-update inference, uniquely explains variance in the FRN amplitudes. These data indicate that in addition to coding model-

free, single-update RPEs, the FRN also signals values estimated through abstract inference on the anti-correlated task structure.

Influence of signed versus unsigned RPEs. It had been suggested that, rather than a reinforcement learning RPE signal, the FRN might reflect a surprise signal. Unsigned model-derived prediction errors (|PEs|) have been claimed to encode such a surprise signal. For the sake of replication, we repeated the analyses reported by Hauser and colleagues (2014b), and entered signed prediction errors (PEs) derived from the double-update model as well as their unsigned values |PEs| in a multiple regression analysis. The resulting two beta weights were then analyzed in a t-test (Holmes & Friston, 1998). Note that, although signed and unsigned PE values are correlated, this analysis only accounts for uniquely explained variance and betas derived from this analysis are not pseudoeffects of the correlated measure. We found a significant effect of the signed RPEs

on the single-trial amplitude (mean slope of regression .082, SD=.114, $t(19)$=3.418, p=.023), whereas there was a trend for the unsigned RPEs (mean slope of regression=.019, SD=.090, $t(19)$=1.877, p=0.076. Similar findings were obtained when binning small RPEs and large RPEs for wins and losses separately by identifying the 25th and 75th percentile of each individual's range of RPEs, and testing the effect of RPE size and feedback valence on FRN amplitudes using a repeated measures ANOVA. We found a significant interaction effect of valence and RPE size $(F(2,18)=10.68$, p=.004, see Figure 7-3C), whereas no significant main effect of size and valence was observed (all ps >.10, all Fs ≤ 2.81). We infer that signed double-update RPEs uniquely explain variance in the FRN trial-by-

trial amplitudes. Thus, we argue that the FRN codes learning signals rather than reflecting mere surprise (Hauser *et al.*, 2014b).

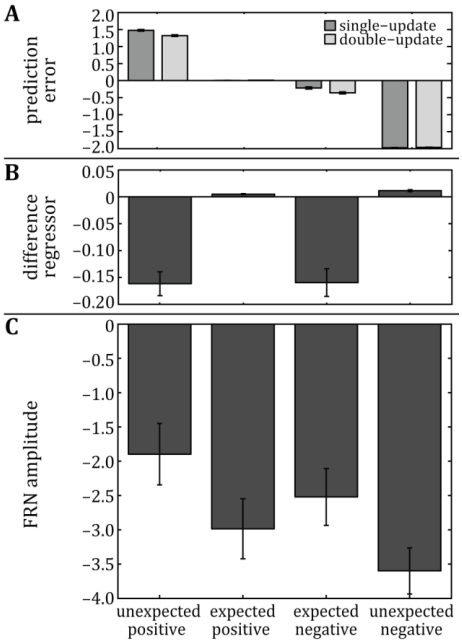


Figure 7-3. A) Mean Reward Prediction Errors, derived from the Single-Update and Double-Update Model, as a function of valence and expectedness. We plot mean RPEs of the 25th and 75th percentile of each individual's range of RPEs. B) Mean difference of chosen decision values estimated by the Double-Update and the Single-Update Model (equation 15), that is, estimates of abstract 'double-update' inference components in decision-making. Trial-by-trial differences were used as regressors to predict the electrophysiological signal. C) Mean FRN amplitudes, plotted as a function of valence and expectedness of the feedback in the trial. FRN amplitudes were influenced by the interaction of valence and expectedness and showed a positive association with reward prediction errors. FRN amplitudes were furthermore negatively correlated with the difference regressor, indicating that the FRN codes abstract 'double-update' inference components - the influence of 'what might have happened' - in reward-guided decision-making.

7.4 Discussion

While it is essential for an agent to learn from observed outcomes emerging as a consequence of actual choice hypothetical inference on 'what might have happened' is thought to additionally guide decision-making and improve behavioral adaptation. In this study, we could identify separate contributions of single-update versus double-update learning, the latter reflecting abstract inference on the task structure, by focusing on unique variances explained by the difference between their value estimates in the FRN. Thereby, we demonstrate that the FRN codes the influence of both, a model-free, single-update RPE and additional components related to abstract inference on the anti-correlated task structure.

Revisiting the role of the FRN in reinforcement learning. For the investigation of electrophysiological signatures of reinforcement learning, a candidate deflection is the FRN. An influential theory suggests that the FRN is a neural signature of model-free, single-update RPE processing (Holroyd & Coles, 2002). Studies using cross-trial averages, and recently also parametric analyses based on computational modeling, have partially confirmed these theoretical claims (for a review see Walsh & Anderson, 2012). Our findings are in line with and extend Holroyd and Coles' seminal theory: We argue that the FRN in fact mirrors model-free learning signals but additionally codes influences of inferred stimulus values deduced via abstract inference on the task structure.

Learning from experiential, observed outcomes versus inferred outcomes based on abstract inference, has been previously related to the distinction between model-free and model-based control of behavior (Bromberg-Martin *et al.*, 2010; Lucantonio *et al.*, 2012): model-free control is driven by rewards achieved in the past and therefore retrospective and reflexive. By contrast, model-based behavior as the deliberative, prospective mode of control relies on an internal representation of the environment and allows forward planning of future actions based on their potential outcomes. Consequently, model-based control is computationally more expensive but enables individuals to rapidly adapt their behavior in a dynamically changing environment (Daw *et al.*, 2005; Dolan & Dayan, 2013). In the present study, the double-update model modifies model-free learning signals by incorporation of the anti-correlated task structure which leads to more successful behavioral adaptation in a dynamic environment. It is therefore conceivable that abstract double-update inference is associated with the model-based system, as the model-free system is by definition blind towards the environmental structure. Our result that the FRN codes both, model-free single-update RPEs and additional components reflecting abstract inference on the anti-correlated task structure fits neatly to the notion of a common architecture for the human control systems over decision-making with ubiquitous higher-order model-based influences in neural reward processors (Doll *et al.*, 2012). In line with previous studies, this cuts against dual systems accounts of isolated model-free vs. model-based

control. However, an alternative explanation might include that the abstract double-update inference as observed in the present study does not arise from a full model-based system but rather temporal-difference learning about the relationship of the choice options (Shohamy & Wagner, 2008; Doll *et al.*, 2012; Wimmer *et al.*, 2012). While there is no unique formulation of model-free and model-based control (Dolan & Dayan, 2013) and tasks may differ in which aspects of the system they capture, another approach of dissociating model-based versus model-free decision-making is to use sequential decision tasks (Glascher *et al.*, 2010; Daw *et al.*, 2011). In sequential decision-making, the model-based learner acquires knowledge about the task-immanent transition structure and uses this to evaluate decision options. This task differs from the experiment used here by capturing a more complex learning environment and thereby offers the possibility to investigate one important feature of model-based control, namely, inferring action values by a learnt cognitive sequential model of the consequences of one's actions (Doll *et al.*, 2012). Our findings motivate further electrophysiological investigations of different aspects of behavioral control, e.g., in more complex environments via the application of sequential learning tasks.

Potential generators of the FRN and neural correlates of behavioral control. Compared to methods such as fMRI, EEG is limited in tracking signals from deep subcortical structures that have been postulated to play key roles in reinforcement learning, such as the striatum. However, it has been claimed plausible that local field potentials are modulated by afferent midbrain prediction error signals (Talmi *et al.*, 2012). Possible generator regions of the FRN are a matter of debate. Notably, the FRN is measured over the medial frontal cortex (MFC), a region that has been implicated in the coding of model-free as well as model-based RPE signals (Daw *et al.*, 2011). Interestingly, studies using EEG source localization discuss the origin of the FRN in the striatum (Foti *et al.*, 2011a), a brain structure that has also been shown to be involved in the processing of both model-free and model-based learning signals (Daw *et al.*, 2011). Recently, a combined EEG-fMRI study likewise adopted a single-trial approach in order to track coupling of feedback signals in hemodynamic and electrophysiological responses (Becker *et al.*, 2014). Their data

imply contributions of multiple frontal midline generators to the FRN signal. Notably, hemodynamic activity in the medial prefrontal cortex (mPFC) and the striatum was also coupled to the magnitude of electrophysiological FRN responses.

FRN and signed versus unsigned RPEs. Whether the FRN codes signed or unsigned RPEs is a matter of ongoing debate (Alexander & Brown, 2011; Talmi *et al.*, 2013; Cavanagh & Frank, 2014; Hauser *et al.*, 2014b; Ullsperger *et al.*, 2014). A previous modeling-based study has found a correlation of the FRN with unsigned RPEs. The authors conclude that the FRN rather reflects salience coding than learning (Hauser *et al.*, 2014b). Our findings contribute to this discussion by showing that the FRN is explained by signed RPEs and also reflects abstract higher-order components. This points towards a role for the FRN in learning beyond coding of expectedness or salience only. However, albeit on a trend level only, we also found a correlation between the FRN and unsigned PEs. Our findings are in line with a recent interpretation which discusses contributions of signed versus unsigned RPE to the FRN (Ullsperger *et al.*, 2014): the authors argue that the presence of both signed and unsigned RPEs in the FRN is plausible because unsigned RPEs may, beyond signed RPEs, play a particular role in learning and behavioral adaptation. The authors suggest that surprise signals can be used to modify a weighting factor (such as learning rate or volatility of the environment) of signed RPEs. While our findings corroborate this unifying notion, we moreover suggest a closely related interpretation of the association of unsigned PEs with the FRN. The absolute value of the model-free RPE signal has been claimed to function as information on the reliability of the model-free system (Roesch *et al.*, 2012; Lee *et al.*, 2014). Such reliability signals are thought to be used by an arbitration mechanism which allocates the degree of control exerted by one of the systems at a given point in time. Based on the findings by Hauser et al. (2014b) and also the trendwise correlation observed in our data, we suggest that components of the FRN additionally code the reliability of the model-free system and may thereby also reflect an electrophysiological signature of this arbitration process. Specific modeling strategies are warranted to address this question (Li *et al.*, 2011; Roesch *et al.*, 2012; Lee *et al.*, 2014).

Limitations. It has been argued that limitations inherent to the ERP methodology render conclusions about the role of the FRN in reinforcement learning opaque (Cavanagh *et al.*, 2010; Cohen *et al.*, 2011b). Our findings suggest that by adopting a modelling-derived parametric approach, FRN accounts can contribute to a more profound understanding of electrophysiological correlates of human decision-making processes. Building on that, we believe that for future electrophysiological studies in the framework of behavioral control, it seems promising to additionally take dynamic changes in systems-level oscillatory synchronization into account (Cavanagh *et al.*, 2010; Cohen *et al.*, 2011b; Cavanagh & Frank, 2013).

In conclusion, our findings provide an electrophysiological correlate of incorporating abstract inference into the decision-making process. Reduced neural tracking of prediction errors (Tanabe *et al.*, 2013; Parvaz *et al.*, 2015), disturbed mechanisms of inference (Lucantonio *et al.*, 2012; Lucantonio *et al.*, 2014; Huys *et al.*, 2015b) and altered behavioral control, e.g. an imbalance between model-based and model-free control are suggested to have psychopathological implications (Deserno *et al.*, 2013; Huys *et al.*, 2014). For instance, patients suffering from disorders characterized by failure in behavioral adaptation, e.g., addiction or obsessive compulsive disorder, have been reported to show a bias towards model-free as compared to model-based learning (Sebold *et al.*, 2014; Voon *et al.*, 2015). As EEG, in comparison to fMRI, is advantageous with regard to feasibility, we offer new means of studying these processes in patient populations characterized by aberrant reinforcement learning mechanisms.

8 Study 4: Lateral prefrontal model-based signatures are reduced in healthy individuals with high trait impulsivity[5]

8.1 Introduction

Impulsivity can be defined as a tendency for premature choices without foresight but despite adverse consequences (Dalley *et al.*, 2011; Robbins *et al.*, 2012). Impulsivity, a multifaceted construct, has been established as a vulnerability factor for addiction (Verdejo-Garcia *et al.*, 2008). Recent studies support the view of self-report trait impulsivity as an endophenotype for addiction disorders (Ersche *et al.*, 2010; Ersche *et al.*, 2013): non-addicted, albeit cognitively impaired and at-risk, first-degree family members showed intermediate levels of trait impulsivity when compared to addicted siblings and unrelated controls (Ersche *et al.*, 2010; Ersche *et al.*, 2013). This endophenotype research characterized unaffected siblings by intermediate brain alterations, most prominently by means of structural measures of fronto-striatal circuits (Ersche *et al.*, 2012a). Interestingly, fronto-striatal structural measures were shown to correlate with the expression of the dominant mode of behavioral control (de Wit *et al.*, 2012b; Voon *et al.*, 2015). An important proposal linked the personality trait impulsivity to an overreliance on habitual behavioral control (Everitt *et al.*, 2008; Hogarth *et al.*, 2013). Empirical evidence for this hypothesis mainly stems from animal models of drug addiction showing that high-impulsive rats are predisposed for escalation of repeated drug self-administration and early relapse after abstinence (Dalley *et al.*, 2007).

Behavioral control is postulated to be parsed between competing habitual and goal-directed systems (Dickinson, 1985; Dolan & Dayan, 2013). This dual system theory was formalized through the use of computational models (Daw *et al.*, 2005): habitual control can be described by 'model-free' temporal-difference algorithms, which retrospectively update expectations by reward prediction errors. Dominance of model-free control is accompanied by

[5] This chapter corresponds to the following article: Deserno, L., Wilbertz, T., Reiter, A.M.F., Horstmann, A., Neumann, J., Villringer, A., Heinze, H.J. & Schlagenhauf, F. (in press). Lateral prefrontal model-based signals are reduced in healthy individuals with high trait impulsivity. *Translational Psychiatry*.

reduced immediate sensitivity to outcome devaluation because new outcome experiences are required to gradually adapt outcome expectations (Dayan, 2009a). In marked contrast, goal-directed control relies on the prospective consideration of possible actions and their potential future outcomes (Balleine & Dickinson, 1998). This can be described by 'model-based' algorithms, which capture a task as a map in a forward-planning manner and therefore model-based control enables flexible behavioral adaptation in dynamic environments (Doll *et al.*, 2012). Using sequential decision-making and computational modeling, it was demonstrated that healthy individuals use a mixture of both control strategies meanwhile prefrontal cortex and ventral striatum code signatures of both model-free and also model-based control (Daw *et al.*, 2011; Deserno *et al.*, 2015b). Strikingly, when using the same task, a balance of behavioral control shifted towards model-free was reported across several psychiatric conditions characterized by high levels of trait impulsivity, including addiction (Sebold *et al.*, 2014; Voon *et al.*, 2015).

Adopting such a *Computational Psychiatry* approach (Montague *et al.*, 2012; Stephan & Mathys, 2014; Wang & Krystal, 2014), it has yet not been studied whether a shift towards model-free control also extends to the vulnerability factor impulsivity. One study could show that high-impulsive smokers showed reduced goal-directed control in a devaluation paradigm when compared to low-impulsive smokers (Hogarth *et al.*, 2012b). However, the latter study could not rule out potential effects of smoking addiction and did not include functional or structural brain measures. To fill this gap, we utilized sequential decision-making, as in previous studies (Daw *et al.*, 2011; Sebold *et al.*, 2014; Deserno *et al.*, 2015b; Voon *et al.*, 2015), in healthy low- and high-impulsive individuals taken from a larger sample. Finally, 50 participants underwent task-based fMRI to examine neural correlates of model-free and model-based control based on computational modeling of the observed behavior. First, we explored whether high-impulsive individuals show reduced model-based control similar to patients (Sebold *et al.*, 2014; Voon *et al.*, 2015). Dimensional approaches to psychiatry suggest that impairments in behavioral control, as observed in drug addiction, could lie at the end of a continuum including healthy high-

impulsive individuals (Robbins *et al.*, 2012). Therefore, it appears conceivable that healthy individuals with levels of impulsivity comparable to patients show intermediate alterations in behavioral control. Second, on the neural level, we tested whether high-impulsive individuals show elevated model-free prediction errors or reduced model-based signatures. Such effects were expected in ventral striatum or prefrontal cortex as these regions were previously indicated in coding model-free prediction errors and additional model-based signatures (Daw *et al.*, 2011; Deserno *et al.*, 2015b). Structural MRI was analyzed by means of gray matter density to assess its co-variation with the behavioral and functional imaging effects.

8.2 Materials and methods

Participants and instruments. A total of 452 participants completed the Barratt Impulsiveness Scale-11, a self-report measurement of trait impulsivity with high retest reliability in clinical and non-clinical populations (Patton et al., 1995). Among these, 52 right-handed individuals were selected from the upper and lower ends. Sample size for this study was determined in accordance to previous between-group studies with the same task (Sebold et al., 2014; Voon et al., 2015). According to the literature (Stanford et al., 2009), mean total BIS-scores of each group met criteria for high and low impulsiveness (Table 8-1). Both groups were matched for age and gender and screened for axis-I psychiatric disorders using the SCID-IV interview (First et al., 2001). Based on this screening, one participant was excluded because of a recent episode of major depression and another participant fell asleep during task-based fMRI. The final sample consisted of 50 participants (24 high-impulsive and 26 low-impulsive participants (Sobell, 1992). Intelligence was examined based on a German vocabulary test (Schmidt & Metzler, 1992) as well as working memory using the backward digit span test and processing speed using the digit symbol substitution test (Wechsler, 1955). For detailed group description see Table 8-1. The local ethics committee (University of Leipzig) approved the study. All participants gave written informed consent and received monetary compensation on an hourly basis in addition to their monetary gain during the task.

Table 8-1. Sample Characteristics. Group means with standard deviations and range in brackets are reported; for group comparisons two-sample t-test were used. Drinking was assessed with the time line follow back interview, verbal intelligence with a German vocabulary test and working memory with a backward digit span test.

Healthy participants	with high trait impulsivity (n=24)	with low trait impulsivity (n=26)	Sig.
Age (years) (24/26)	27.29 ± 3.67 (22-33)	27.58 ± 3.74 (20-33)	.78
Gender (24/26)	12 female 12 male	13 female 13 male	
BIS total (24/26)	74.76 ± 5.07 (68-90)	50.31 ± 3.78 (41-58)	<.001
Smoking (24/26)	1 smoker	1 smoker	
Drinking (g) (24/25)	23.48 ± 21.28 (0-74)	17.48 ± 14.84 (0-63)	.26
Verbal Intelligence (24/25)	109.88 ± 7.60 (97-129)	113.08 ± 5.93 (104-122)	.11
Working memory (24/25)	8.08 ± 2.10 (4-12)	8.12 ± 1.97 (5-12)	.82
Processing speed	88.26 ± 11.29 (4-12)	82.71 ± 11.45 (4-12)	.09

Sequential decision-making task. A two-step sequential decision task was implemented as in previous studies (Daw et al., 2011; Wunderlich et al., 2012b; Voon et al., 2015). Participants had to make two sequential choices between pairs of stimuli to receive a monetary reward after the second choice. Within each trial, participants had to decide between two gray boxes at the first stage or two colored boxes at the second stage (Figure 8-1A). Crucially, each first-stage choice was associated with a different pair of colored boxes at the second-stage via a fixed transition probability of 70%, which did not change during the experiment (Figure 8-1B). Thus, choice of each first-stage stimulus was commonly (70%) associated with a certain second-stage pair of stimuli and this is labeled a 'common state'. In reverse, choice of each first-stage stimulus rarely resulted (30%) in the other second-stage pair of stimuli and this is labeled a 'rare state'. Model-free control neglects this transition probability and staying with the same first-stage action that lead to a reward after a second-stage choice is most likely (a main effect of reward). In contrast, model-based control takes into account the transition probabilities. Thus, staying at the first-

stage decreases after having received a reward in a rare state but increases after having received no reward in a rare state (reward x state interaction).

All stimuli were randomly assigned to the left and right position on the screen. At the first stage, the chosen gray stimulus was surrounded with a red frame, moved to the top of the screen after completion of a 2-s decision phase and remained there for 1.5s. Subsequently, participants entered the second stage (a common or rare state depending on the type of transition) and decided between two colored boxes. After a second-stage choice, feedback (reward or no reward) was delivered according to slowly and independently changing Gaussian random walks. These random walks were identical to Daw et al. (2011) as it was shown that less distinct random walks for reward delivery reduce the degree of model-based behavior (Eppinger et al., 2013). Slowly changing reward probabilities at the second stage challenge the subject with ongoing learning and thus maximize the dissociation of the two control strategies at the first stage. Thus, non-stationary reward probabilities at the second-stage induce ongoing model-based evaluation while stationary reward probabilities would favor a dominance of model-free control at some point in time. The task consisted of a total of 201 trials with two choice stages within each trial. Trials were separated by an exponentially distributed inter-trial interval (ITI) with a mean of 2s. Prior to the experiment and similar to Daw et al. (2011), participants were explicitly informed that the transition structure from the first to the second stage would remain constant throughout the task. Information was provided about the independence of reward probabilities and their change over time. Before MRI scanning, participants performed a 55-trial version of the task with different stimuli and reward probabilities and were instructed to maximize reward in the main experiment, which they received as monetary payout after completion of the task.

First-stage stay-switch behavior was analyzed as a function of reward (reward/no reward) and state (common/rare) in the previous trial. Each individual's first-stage stay probabilities were subjected to repeated-measures analysis of variance (ANOVA, using anovan in MATLAB) with reward and state as within-subject factors and impulsivity (high/low) as between-subject

factors. A main effect of reward shows an influence of model-free control while the interaction of reward and state reveals influences of model-based control. Previously, healthy individuals showed a mixture of both control strategies (Daw et al., 2011; Wunderlich et al., 2012b; Smittenaar et al., 2013; Sebold et al., 2014; Deserno et al., 2015b; Voon et al., 2015) expressed by a significant main effect of reward and a significant interaction of reward and state. In the following, we describe a more fine-grained dissociation of the two control strategies via computational modeling, which also provides individual trial-by-trial signatures for the analysis of neural measurements. All behavioral analyses were performed using MATLAB 2010b.

Computational model. As in previous studies (Daw et al., 2011; Wunderlich et al., 2012b; Deserno et al., 2015b; Voon et al., 2015), we adopted a computational modeling approach to disentangle influences of model-free and model-based control on participant's choice behavior. To this end three types of models were applied. (1) A model-free algorithm capturing only a main effect of reward in first-stage stay-switch behavior. This algorithm was the temporal-difference model SARSA(λ) which learns decision values retrospectively after prediction errors occur (Sutton & Barto, 1998). (2) A model-based algorithm, which only gives an interaction of rewards and state but no main effect of reward. To this end, first-stage values were computed prospectively by multiplying maximum values at the second stage with explicitly instructed transition probabilities (Daw et al., 2011). (3) A combination of both algorithms, a so-called hybrid model which can reproduce a main effect of reward and an interaction of reward and state (Daw et al., 2011). Values from all three models were transformed into choice probabilities using a softmax rule with three parameters accounting for stochasticity separately at the first and second stage ($\beta_{1\&2}$) and a repetition parameter (ρ) accounting for perseverance of first-stage choices.

Leaving out parameters of the softmax, the model-free algorithm SARSA(λ) has 3 parameters: first- and second-stage learning rates (α_1/α_2), which describe how quickly values change with respect to first-stage and second-stage prediction errors; stage-skipping update λ (another learning rate), which connects the two stages via an influence of reward prediction errors at the second-stage on first-stage values. Importantly, λ describes how quickly first-stage

values change with respect to second-stage reward prediction errors and thus accounts for the main effect of reward in first-stage stay behavior but not for an interaction of reward and state. Thus, a high value of λ signifies a stronger influence of reward prediction errors at the second stage on first-stage values. The model-based algorithm shares one parameter with the model-free algorithm (α_2) because both algorithms converge at the second stage. In line with previous work (Daw et al., 2011; Deserno et al., 2015b), we also show that including the parameter λ improves the fit to the data (see S-Table 3-1). To give an interaction of reward and state no further parameter is required as the interaction results from multiplying maximum values of second-stage stimuli with explicitly instructed transition probabilities (Daw et al., 2011). The hybrid algorithm has a total of 4 parameters: three parameters from SARSA(λ) and a fourth parameter (ω) that weights the influence of model-free and model-based values and is therefore of most interest because it represents a relative balance of the two control strategies. Please see supplementary information for equations and model fitting.

Model comparison. The aim of model comparison is to identify one best-fitting algorithm. In other words, a control strategy that is most likely in groups of high- and low-impulsive individuals. To compare the three models for their relative goodness of fit, we subjected the model evidence (approximated via sampling from the empirical prior distribution) to a random-effects Bayesian model selection procedure (Stephan et al., 2009). The resulting exceedance probabilities show which model is most likely in a population (Stephan et al., 2009). In the supplementary information, we show that other measurements of relative model fit proved consistent with this approach (S-Table 3-1) and show that best-fitting parameters reproduce the observed behavior (S-Figure 3-1).

Group comparison of model parameters. The predictions of the two control strategies differ at the first-stage of the task. In accordance with raw data analysis, parameters that explain variance in first-stage decision values are of main interest here. In the hybrid model, the winning model in both groups, a weighting parameter (ω) determines to which extent overall first-stage decision values are influenced by model-free and model-based values. Two further parameters, originally

from the model-free algorithm, also directly influence the update of first-stage values: first-stage learning rate (α1), which determines how quickly first-stage values change with respect to prediction errors at the onset of the second stage; and a stage-skipping update (λ), which determines to what extent first-stage values change with respect to reward prediction errors and accounts for the main effect of reward. Finally, there is also a second-stage learning rate (α2), which determines how quickly second-stage values change with respect to reward prediction errors but do not directly influence first-stage values; we subjected all four parameters of the hybrid model (ω, α1, α2, λ) to a one-way multivariate analysis of variance (MANOVA, using manova1 in MATLAB) with the between-subject factor impulsivity (high/low).

Magnetic Resonance Imaging. Functional imaging was performed using a 3 Tesla Siemens Trio scanner to acquire gradient echo T2*-weighted echo-planar images with blood oxygenation level dependent contrast. Covering the whole brain, 36 slices were acquired in oblique orientation at 20° to AC-PC line in ascending order with 2.5-mm thickness, 3x3mm^2 in-plane voxel resolution, 0.5-mm gap between slices, TR=2s, TE=22ms and a flip angle α=90°. Prior to functional scanning, a field map was collected to account for individual homogeneity differences of the magnetic field. T1-weighted structural images were also acquired (TR=1300ms, TE=3.46ms, flip=10°, matrix=240×256, voxel size: 1×1×1mm, slices=170).

Analysis of fMRI data. Two participants had to be excluded due to artifacts in ventral sections of the brain. Thus, functional imaging results are reported for a sample of 48 participants (23 high-impulsive and 25 low-impulsive participants). FMRI data were analyzed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). For preprocessing, images were corrected for delay of slice time acquisition. Voxel-displacement maps were estimated based on field maps. All images were realigned to correct for motion and were also corrected for distortion and the interaction of distortion and motion. The images were spatially normalized into the Montreal Neurological Institute (MNI) space using the normalization parameters generated during the segmentation of each subject's anatomical T1 scan (Ashburner & Friston, 2005); spatial smoothing was applied with an isotropic Gaussian kernel of 6mm full width at half maximum.

Prior to statistical analysis, data were high-pass filtered with a cutoff of 128s. An event-related analysis was applied to the images on two levels using the general linear model approach (GLM) as implemented in SPM8. As in the original paper by Daw et al. (2011), the analysis focused on the two time points within each trial when prediction errors arise: at onsets of the second stage and at onsets of reward delivery. Prediction errors at second-stage onsets compare values of first- and second-stage stimuli and therefore vary with respect to the weighting parameter (ω), which gives the balance of the two control strategies. At the first level, both time points were entered into the model as one regressor, which was parametrically modulated 1) by model-free prediction errors and 2) by the difference of model-based and model-free prediction errors, which reflects the difference between model-based and model-free values (the partial derivative of the value function with respect to ω). Note, that this difference regressor equals zero at reward delivery, because both algorithms converge at this time point. To avoid any confound of the neural results due to activity differences between these two time points per se, the difference regressor was mean-centered within each subject and the time-point of reward delivery was additionally included as a separate regressor. As in Daw et al. (2011), the design also included first-stage onsets with two parametric modulators, the softmax probability for choosing one of the two first-stage probabilities as well as its partial derivative with respect to ω, but these onsets were not in the focus of the present analysis. Individual (random-effects) model parameters were used to generate modeling-derived regressors. Invalid trials (no choice within response window) were modeled separately. All regressors were convolved with the canonical hemodynamic response function as provided by SPM8 and its temporal derivative. The six movement parameters from the realignment were included in the model as regressors of no interest as well as the first derivative of translational movement with respect to time. An additional regressor was included censoring scan-to-scan movement >1 mm.

At the second level, contrast images of model-free prediction errors and the difference of model-based and model-free prediction errors were taken to a second-level random effects model. For correction of multiple comparisons, family-wise error (FWE) $p<.05$ at the cluster

level was applied to statistical maps displayed at p<.001 uncorrected with a cluster extent k=20. Previous research revealed an important role of prefrontal cortex and ventral striatum in coding signatures of both systems (Glascher *et al.*, 2010; Daw *et al.*, 2011; Lee *et al.*, 2014; Deserno *et al.*, 2015b). Thus, mean parameter estimates for clusters of ventral striatum and prefrontal cortex were extracted and then tested between groups using three repeated-measurers ANOVAs with control mode (model-free/model-based) as within-subject factor and impulsivity (high/low) as between-subject factor. Subsequently, a one-way MANOVA with the between-subject factor impulsivity was used to assess regional specificity by comparing the difference between both effects (model-free prediction errors and the difference of model-based and model-free prediction errors) in all three regions of interest.

Voxel-based morphometry. For segmentation of each subject's anatomical T1 the unified segmentation approach was applied as implemented in SPM8 (Ashburner & Friston, 2005). Subsequently, each individual's modulated image of gray matter density was smoothed with an isotropic Gaussian kernel of 6mm full width at half maximum. The smoothed images were then subjected to a random-effects model containing total intracranial volume as a covariate.

Using fMRI clusters named above, gray matter density was extracted for medial and lateral PFC as well as ventral striatum and included as covariates in between-group comparisons of behavioral and functional imaging data. We also tested for between-group effects. Independent of impulsivity, we examined a co-variation of the parameter ω with gray matter density as reported previously for medial prefrontal and orbitofrontal cortex (Voon et al., 2015). Given these results by (Voon et al., 2015) but also studies that implicate lateral prefrontal cortex in model-based control (Smittenaar et al., 2013), we constructed a bilateral search volume (taken from the AAL Atlas, Tzourio-Mazoyer et al., 2002) of medial prefrontal and orbitofrontal cortex (superior medial frontal gyrus, medial orbitofrontal gyrus and anterior cingulate cortex) and lateral prefrontal cortex (middle frontal gyrus and inferior frontal gyrus).

8.3 Results

Sample characteristics. As BIS was selection criterion, groups differed significantly (Table 8-1). Notably, mean BIS of high-impulsive individuals (74.76 ± 4.96) lay in a similar range as for drug users and their siblings (Ersche et al., 2010). As shown in Table 8-1, groups were matched for age and gender and did not differ regarding measures of drinking and smoking or neurocognitive measures.

Behavioral raw data. First-stage choice behavior of all participants showed a significant main effect of reward and an interaction of reward and state (reward $F(1,48)=75.30$, $p<.001$, reward x state $F(1,48)=64.30$, $p<.001$, Figure 8-1C) indicating that across all participants aspects of model-free and model-based control were present. These effects were also present when looking at both groups separately.

Individuals with high trait impulsivity did not show a reduction of model-based control tested by a three-way interaction (reward x state x impulsivity $F(1,48)=.73$, $p=.40$, Figure 8-1C) but there was a trend towards a significant reward x impulsivity interaction ($F(1,48)=3.50$, $p=.07$, Figure 8-1C). Close inspection of Figure 8-1C suggest that this reward x impulsivity interaction results from slightly lower stay probabilities in high-impulsive compared to low-impulsive individuals after unrewarded (particularly unrewarded-rare trials) but not rewarded trials. Thus, the main effect of reward appeared slightly stronger in high-impulsive individuals. To confirm this, a one-tailed between-group t-test (high > low) was performed on the main effect of reward (the difference between staying after rewards and staying after non-rewards). Indeed, this difference between staying after rewards and staying after non-rewards was significantly higher in high-impulsive individuals ($t(48)=1.94$, $p=.04$ Figure 8-1D) indicating a subtle accentuation of model-free control in high-impulsive individuals. Although the repeated-measures ANOVA did not reveal any interaction of impulsivity with state or reward and state, following a reviewer's suggestion, we further unpacked the reward x impulsivity interaction for rare and common trials separately. This one-tailed post-hoc test revealed that the observed

effect was mainly driven by the difference between rewarded and unrewarded trials in rare transitions ($t(48)=1.6$, $p=.06$) but not in common transitions ($t(48)=.26$, $p=.40$).

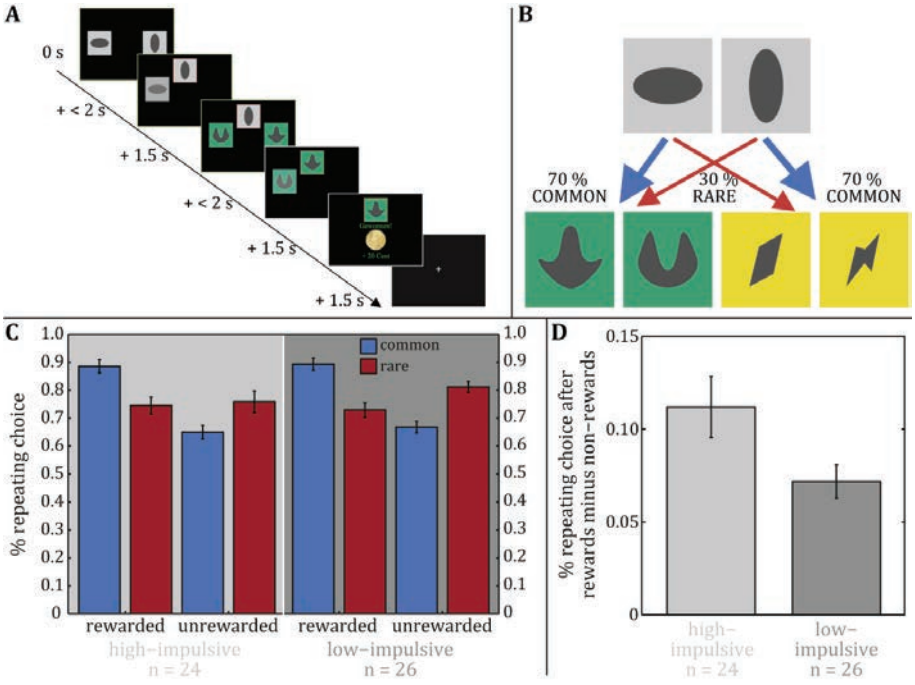


Figure 8-1. Task and behavioral raw data. A) Exemplary trial sequence of the task. B. State-transition probabilities. C) Stay-switch behavior at the first-stage was analyzed as function of reward and state in the previous trial. These stay probabilities were subjected to repeated-measures ANOVA with reward and state as within-subject factors and group as between-subject factors. This revealed a significant main effect of reward ($F(1,48)=75.30$, $p<.001$) and reward x state interaction ($F(1,48)=64.30$, $p<.001$); no significant main effect of state ($F(1,48)=1.32$, $p=.26$) and no significant state x group ($F(1,48)=.07$, $p=.80$) or reward x state x group ($F(1,48)=.73$, $p=.40$) interactions. There was a trend towards a significant reward x group interaction ($F(1,48)=3.50$, $p=.07$). D) In a one-tailed post-hoc t-test, the difference between staying after rewards and staying after non-rewards was significantly increased in the high- compared to the low-impulsive group ($t(48)=1.94$, $p=.04$). Error bars represent standard error.

Computational modeling. Model selection revealed the hybrid model as best-fitting in both groups (high-impulsive exceedance probability=.9974, low-impulsive exceedance probability=.9997, S-Table 3-1). This underlines that a mixture of both control modes, provided by this hybrid model, is the most likely control mechanisms in low- and high-impulsive groups.

All four parameters of the hybrid model (ω, α1, α2, λ, for their distribution see S-Table 3-2) weresubjected to a MANOVA with the between-subject factor impulsivity. This revealed a significant effect of impulsivity $F(4,45)=2.72$, $p=.041$). Post-hoc univariate tests (Figure 8-2) showed no difference for the balance of control ω (high-impulsive .6076 ± .1114, low-impulsive .5943 ± .1080, $F(1,48)=.19$, $p=.64$, Figure 8-2A), for first-stage learning rates α1 (high-impulsive .5272 ± .2077, low-impulsive .4330 ± .1928, $F(1,48)=2.8$, $p=.10$, Figure 8-2B), nor for second-stage learning rates α2 (high-impulsive .5803 ± .1706, low-impulsive .6006 ± .1328, $F(1,48)=.22$, $p=.64$, Figure 8-2C), but significantly higher stage-skipping update λ (high-impulsive .6854 ± .0756, low-impulsive .6202 ± .0965, $F(1,48)=7.00$, $p=.01$, Figure 8-2D). In addition, to demonstrate that the effect of impulsivity on λ was not due to fitting ω simultaneously, we also tested whether λ was significantly different between groups when comparing parameters of the model-free algorithm with ω=0. This was indeed the case (high-impulsive .69±.07, low-impulsive .64±.07, $t(48)=2.96$, $p=.005$). The parameter λ signifies a stronger influence of reward prediction errors at the second stage on first-stage decision values and accounts for the main effect of reward observed in first-stage stay behavior. In line with raw data analysis, this speaks for a subtle, albeit significant, elevation of model-free control in high-impulsive individuals. This result remained significant when including neurocognitive measures, amount of alcohol intake or gray matter density as covariates. Explorative comparison of parameters of the softmax observation model (β1, β2, ρ) and the negative log-likelihood showed no significant differences ($t(48)<=1.61$, $p>=.11$). See S-Table 3-2, for distribution of all parameters and the negative log-likelihood.

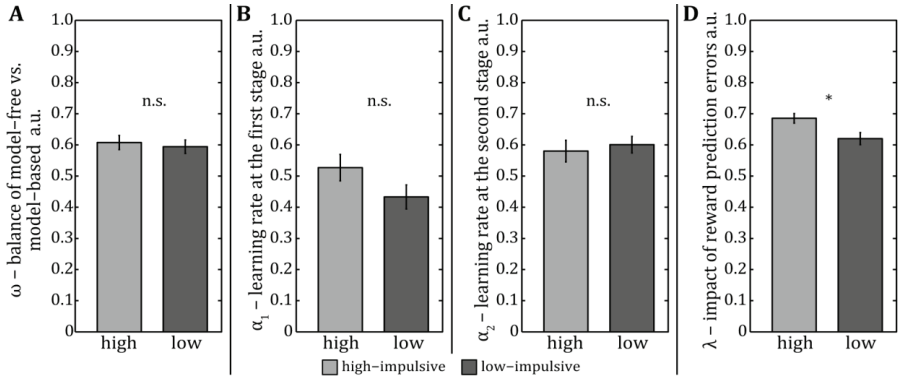


Figure 8-2. Hybrid model parameters. Four parameters of the hybrid model (ω, α1, α2, λ, S-Table 3-3) were subjected to a multivariate analysis with the between-subject factor impulsivity. This revealed a significant effect of impulsivity ($F(4,45)=2.72$, $p=.04$). Post-hoc univariate tests showed no difference for A) the balance of model-free and model-based control (ω), B) first-stage learning rates (α1), nor C) for second-stage learning rates (α2), but D) significantly higher stage-skipping update (λ). Error bars represent standard errors.

Functional MRI. As a replication of previous work (Daw et al., 2011; Deserno et al., 2015b), the conjunction of model-free prediction errors and the difference of model-based and model-free prediction errors across both groups reached significance (whole-brain p-FWE<.05 at the cluster level) in right and left ventral striatum, medial prefrontal cortex and right ventro-lateral prefrontal/orbitofrontal cortex (S-Table 3-3, Figure 8-3). Thus, for between group-comparison, parameter estimates of the clusters for bilateral ventral striatum, medial and right ventro-lateral prefrontal cortex were tested between groups using three repeated-measures ANOVA with control (model-free/model-based) as within-subject factor and impulsivity (high/low) as between-subject factor. As depicted in Figure 8-3, no main effect of impulsivity ($F(1,46)<=.28$, $p>=.60$) nor an impulsivity x control interaction was observed ($F(1,46)<=1.79$, $p>=.19$) in ventral striatum and medial prefrontal cortex. In right lateral PFC, we observed no main effect of impulsivity ($F(1,46)<.01$, $p=.99$, Figure 8-3B) but a significant impulsivity x learning interaction ($F(1,46)=4.80$, $p=.03$, Figure 8-3B).

To assess regional specificity, a MANOVA with the between-subject factor impulsivity was used to compare the difference between two effects of interest (model-free prediction errors and the difference of model-based and model-free prediction errors) in all three regions of interest, which indeed reached significance *(Δ_{Roy}=.28, F*(44)=4.09*, p*=.01*).* All between-group fMRI findings remained significant when adding neurocognitive measures, amount of alcohol intake or gray matter density as covariates.

Structural MRI. First, no differences were observed between low- and high-impulsive groups at a whole-brain level nor when looking at anatomical or fMRI-derived regions of interest. Second, a significant positive correlation between dorsolateral prefrontal gray matter density and parameter ω (a higher ω indicates more model-based choices) was observed (MNI x=-42, y=22, z=50, t=5.04, p-FWE=.05 for bilateral medial and lateral PFC, r=.59, R^2=.35, 95% confidence interval [.37, .75], Figure 8-4).

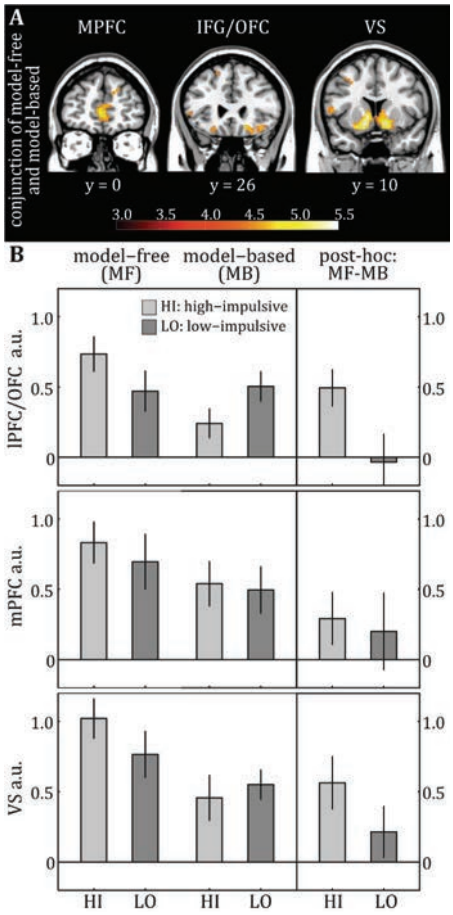


Figure 8-3. fMRI results across the entire sample. A) Across both groups a significant (whole-brain p-FWE<.05 at the cluster level) conjunction of model-free reward prediction errors and the difference of model-based and model-free prediction errors was observed in right and left ventral striatum, medial prefrontal cortex, right ventro-lateral prefrontal/orbitofrontal cortex, right and left parietal cortex and posterior cingulate cortex. For display purposes, maps are thresholded at p<.001 uncorrected and a cluster extent of k=20. B) mean parameter estimates of the cluster for bilateral ventral striatum, medial and right ventro-lateral prefrontal cortex tested between groups using three repeated-measures analysis. No main effect of impulsivity (F(1,46)<=.28, p>=.60) nor an impulsivity x control interaction was observed (F(1,46)<=1.79, p>=.19) in ventral striatum and medial prefrontal cortex (B, middle and lower panel). In right lateral PFC (B, upper panel), we observed no main effect of impulsivity (F(1,46)<.01, p=.99) but a significant impulsivity x learning interaction (F(1,46)=4.80, p=.03).

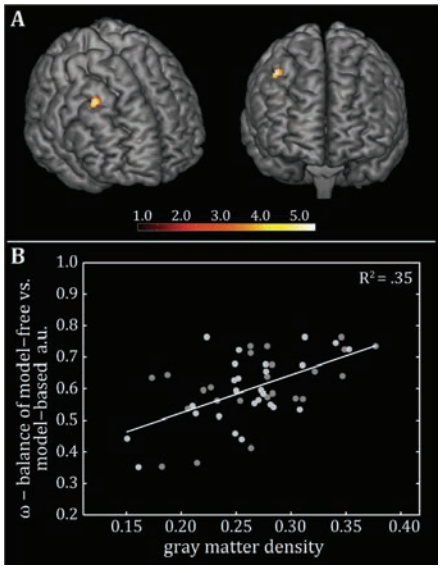


Figure 8-4 Gray matter density and the balance of behavioral control. A positive correlation between gray matter density in dorsolateral prefrontal cortex (MIN x=-42, y=22, z=50, t=5.04, p-FWE=.05 for bilateral medial and lateral PFC) and the balance of model-free and model-based control (ω) was observed. For display purposes, maps are thresholded at p<.001 uncorrected and a cluster extent of k=20.

8.4 Discussion

The present study shows high trait impulsivity in healthy individuals to be accompanied by behavioral and neural signatures in favor of a model-free system of behavioral control. While we did not observe a shift in a balance of behavioral control towards model-free control in high impulsive individuals, two main findings support this notion: first, in line with behavioral raw data analysis, computational modeling revealed a subtle but significant accentuation of model-free control in high-impulsive individuals; second, lateral prefrontal model-based signals were reduced in high-impulsive individuals.

Trait impulsivity, behavioral control and addiction. High-impulsive individuals showed an accentuation of a model-free control system, namely the impact of reward prediction errors on first-stage decision values was elevated. In contrast to addicted and other psychiatric patient samples (Sebold et al., 2014; Voon et al., 2015), we did not find evidence for an impairment of model-based behavioral control in our sample of high impulsive individuals. Utilizing the same sequential decision task, it was recently demonstrated that patients with addictive disorders and other conditions from the impulsivity-compulsivity spectrum show a shift of behavioral control from model-based towards model-free control (Sebold et al., 2014; Voon et al., 2015). In both patient studies (Sebold et al., 2014; Voon et al., 2015), model-based control was reduced (reward x state x group interaction or lower parameter ω) but patients did not differ from controls regarding measures of the model-free system alone (reward x group interaction or higher parameter λ). So far, the origin of behavioral findings in patients remain unclear: they could result from an antecedent accentuation in a model-free system ultimately reducing model-based control although this is not supported by hitherto existing studies; they could be linked to an arbitration or integration problem between two systems or they could be tied to impairments of a model-based system alone. Studies suggest that inter-individual variability in cognitive capacities relate to a model-based system (Otto et al., 2013a; Otto et al., 2013b; Smittenaar et al., 2013; Schad et al., 2014). Interestingly, Sebold et al. showed in alcohol-dependent patients that reduced model-based control was at least abolished when correcting for cognitive capacities

(Sebold et al., 2014) and similar control analyses were not reported in Voon et al., 2015. Here, we show that the risk factor impulsivity results in an accentuation within a model-free control system alone although – unlike in addiction and other patients groups - an overall balance of control was not altered. Importantly, general cognition is very unlikely to account for the findings in the present study. Nonetheless, it remains an intriguing question why some healthy individuals perform this task in a model-free way. One line of reasoning includes that ongoing model-based evaluation during the main experiment challenges limited computational resources (Daw et al., 2005). Studies in healthy individuals support this view by showing that inter-individual differences in cognitive capacities, in particular working memory (Otto et al., 2013a; Otto et al., 2013b; Schad et al., 2014), relate to the balance of model-free and model-based control in this task. Further associations were shown for acute stress reactivity and chronic stress levels (Otto et al., 2013b; Radenbach et al., 2015) as well as striatal presynaptic dopamine levels (Deserno et al., 2015b). Another idea involves that such individuals could have a 'false' model or a 'false' belief about the state transition, e.g. a subjective illusion of control. Interestingly, a recent study reported that healthy adults with a subjective belief of control over reward delivery, which was objectively not given, showed increased ventral striatal and lateral prefrontal activation during reward anticipation (Lorenz et al., 2015). However, this idea cannot be adequately tested with task applied in the present and instead requires experimental designs specifically tailored to address this. Together, these factors most likely also play an important role in explaining the emergence of a dominance of model-free control in psychopathological groups performing this task (Sebold et al., 2014; Voon et al., 2015). With respect to addictive behaviors, it is conceivable that longitudinal interactions with acute drug abuse (Hogarth et al., 2012a), chronic drug consumption (Deserno et al., 2015a) or acute and chronic stress (Otto et al., 2013b; Radenbach et al., 2015) may finally prompt a pattern of reduced model-based control and leave model-free control as the only available mode of control in patients.

Reduced lateral prefrontal model-based signatures in high-impulsive individuals. High-impulsive individuals exhibited reduced model-based signatures in a sector of the lateral PFC. In previous

research, measures of impulsivity were linked to inferior parts of lateral PFC (Farr *et al.*, 2012; Wilbertz *et al.*, 2014), which is an important region exhibiting top-down control (Koechlin *et al.*, 2003) also during sequential decision-making (Lee *et al.*, 2014). Indeed, it was proposed that altered behavioral control in addiction and impulsivity is associated or even results from reduced prefrontal top-down control exerted over striatal regions (Dalley *et al.*, 2011). In the present study, reduced model-based signatures in lateral PFC of high-impulsive individuals could indicate such deficient top-down control. In line, Daw et al. suggested that a covariation of ventral striatal activation with model-free but most strikingly also model-based signatures could result from a top-down, prefrontal to striatal, information flow between the two control systems (Daw *et al.*, 2011). However, it remains an important question what precisely determines the degree of control exerted over striatal regions given that ventral striatal model-based signatures remained unaffected in high-impulsive individuals. One potential explanation for unaffected ventral striatal signals, and also intact model-based behavioral control, is that medial PFC model-based signatures did not differ between high- and low-impulsive groups. Given a likely role of medial PFC in integrating decision values from both systems (Wunderlich *et al.*, 2012a; Lee *et al.*, 2014), intact model-based coding in mPFC may preserve neural top-down control and thus behavioral model-based control. A failure of this mPFC function may ultimately result in an overall shift of behavioral control as observed behaviorally in patients (Sebold *et al.*, 2014; Voon *et al.*, 2015).

Dopamine was also suggested to play an important role in modulating top-down control in fronto-striatal circuits (Braver & Cohen, 1999; Seamans & Yang, 2004; Cools, 2011). While blunted (ventral) striatal dopamine function was reported in addicted patients both pre- and postsynaptically (Volkow *et al.*, 1996; Heinz *et al.*, 2004; Martinez *et al.*, 2005; Martinez *et al.*, 2007; Martinez *et al.*, 2012), animal research has shown that PET measures of ventral striatal dopamine D2 receptor availability are lower in high-impulsive, stimulant-naive rats and predict escalated levels of stimulant self-administration (Dalley *et al.*, 2007). Interestingly, in human PET studies higher levels of impulsivity were shown to be mediated by lower levels of

presynaptic dopamine function (Buckholtz *et al.*, 2010; Schluter *et al.*, 2013). Using the same task and analytic strategy as in the present study, pharmacological elevation of presynaptic dopamine induced a bias towards model-based choices (Wunderlich *et al.*, 2012b). This positive association between model-based control and dopamine was confirmed in a human PET-fMRI study with respect to presynaptic dopamine levels in ventral striatum (Deserno *et al.*, 2015b). Interestingly, in the latter study ventral striatal presynaptic dopamine levels were also shown to be positively correlated with model-based signatures in lateral PFC (Deserno *et al.*, 2015b) at nearby coordinates where model-based signatures were found to be reduced in high-impulsive individuals in the present study. Although low dopamine levels appear to be associated with reduced model-based control, impulsivity and vulnerability to addiction, the exact interplay of these variables still remains to be elucidated in future translational and longitudinal studies.

One may further speculate that a lateral PFC dysfunction characterizes the impulsive spectrum. Indeed, the observed reduction of model-based signals in lateral PFC nicely matches endophenotype studies that revealed lateral PFC (in particular inferior frontal gyrus) as a vulnerability nexus in siblings of stimulant-dependent patients with regard to white matter integrity and gray matter density (Ersche *et al.*, 2012a). Reduced structural PFC integrity was not observed in our sample of high-impulsive individuals, which may be due to differences in sample characteristics. In particular, high-impulsive siblings of addicted patients also show cognitive impairments (Ersche *et al.*, 2012a). To isolate effects of impulsivity, we explicitly choose to study high-impulsive healthy individuals who did not show differences in cognitive measures when compared to low-impulsive individuals. Notably, all behavioral and fMRI results associated with high impulsivity were independent of individual variability in these cognitive measures or gray matter density. Irrespectively of impulsivity, dorsolateral prefrontal gray matter density was positively related to a balance of model-free and model-based control. This confirms previous findings linking prefrontal gray matter density to a balance of control, albeit in a different prefrontal area (Voon *et al.*, 2015). Thus and taken together so far, it is likely that

multiple 'hits' on the functional and structural level multiplex to a vulnerability pattern for addiction (Heinz *et al.*, 2011).

Limitations. The presented behavioral and neural results warrant replication and their predictive relevance remains an important target for future longitudinal studies. Thus, future studies should follow up healthy participants from extreme ends of personality traits, at-risk samples and patients to examine whether alterations in behavioral control predict future development of drug intake as suggested by animal models (Dalley *et al.*, 2007). Regarding decision-making tasks that aim to assess model-free habitual and model-based goal-directed behavior, construct validity remains an important issue. Work from our group has studied construct validity by testing the applied version of sequential decision-making and a selective devaluation task in the same participants (Friedel *et al.*, 2014). Indeed, although in a limited sample size, a positive correlation between the main outcome measures of both tasks was found (Friedel *et al.*, 2014). A similar observation has recently been confirmed in a larger sample and a different design incorporating devaluation into the sequential decision task (Gillan *et al.*, 2015). Although more indirectly, the by now repeatedly reported association between model-based control and general cognitive capacities (Otto *et al.*, 2013a; Otto *et al.*, 2013b; Schad *et al.*, 2014; Otto *et al.*, 2015) also supports construct validity of the applied task in terms of the computational costs, and thus higher cognitive demands, of model-based control.

Conclusion. We present first evidence for the idea that high impulsivity in healthy individuals is accompanied by behavioral and neural signatures in favor of model-free behavioral control. The behavioral results in healthy high-impulsive individuals were qualitatively different to findings in patients with the same task. Effects of smoking, alcohol intake, general cognition or structural brain measures did not account for the findings. Adopting a Computational Psychiatry approach, we show that these techniques represent feasible and mechanistically informative tools that may enrich future longitudinal studies (Montague et al., 2012; Brodersen et al., 2014a).

9 Study 5: Risk factors for addiction and their association with model-based behavioral control[6]

9.1 Introduction

Drug addiction tends to run in families and relatives of drug-dependent individuals have an eightfold increased risk of developing addictive disorders compared with the general population (Merikangas *et al.*, 1998). Endophenotype accounts of addiction postulate that unaffected relatives share alterations in behavioral or cognitive processes similar or intermediate to those observed in addicted individuals (Robbins *et al.*, 2012).

Inspired by a rich body of work in cognitive neuroscience, recent developments in addiction research highlight a shift from goal-directed toward habitual instrumental control systems as biasing addicted individuals to repeatedly choose certain maladaptive behaviors even in the face of negative consequences (Everitt & Robbins, 2005; Dayan, 2009a; Sebold *et al.*, 2014; Voon *et al.*, 2015). This view on addiction builds upon the prominent notion that instrumental control in healthy decision-making arises from contributions of both a deliberative, goal-directed and a reflexive, habitual system (Balleine & Dickinson, 1998; Dolan & Dayan, 2013). Computational formulations have amended this theory (Daw *et al.*, 2005): on the one hand goal-directed behavior uses a mental model of the environment; future actions and potential outcomes are planned in a forward manner and these costly computations enable flexible behavioral adaptation. On the other hand, habitual behavior is retrospective and rigid, but computationally efficient. It relies on 'stamped-in' past rewards and neglects environmental structure.

A shift from goal-directed toward habitual behavior has not only been suggested for addiction itself, but also for recognized risk factors for addiction like acute and chronic stress (Otto *et al.*, 2013b; Radenbach *et al.*, 2015) or impulsivity (Everitt *et al.*, 2008; Hogarth *et al.*, 2012b; Deserno *et al.*, in press). Studies in healthy at-risk populations are of particular importance as

[6] This chapter corresponds to the following article: Reiter, A.M.F.*, Deserno, L.*, Wilbertz, T., Heinze, H.J., & Schlagenhauf, F. (under review). Risk factors for addiction and their association with model-based behavioral control.

*authors contributed equally

they help to elucidate whether a shift toward model-free instrumental control precedes the development of addiction or is a consequence of addictive behavior. Further, they help to rule out potential confounders like neurotoxic effects on brain structure and globally impaired cognitive functioning. In particular, inter-individual differences in cognitive functioning were shown to be associated with the degree of model-based control in healthy individuals (Otto *et al.*, 2013a; Otto *et al.*, 2013b; Schad *et al.*, 2014) but also with impairments in patients (Sebold *et al.*, 2014).

Adopting a dimensional computational psychiatry approach, we asked whether healthy individuals with a positive family history of alcohol-dependence show a bias toward model-free control as it had been observed in addicted individuals (Sebold *et al.*, 2014). Building on previous evidence pointing toward an important role of impulsivity and cognitive capacity in instrumental control within populations at risk for or suffering from addiction (Ersche *et al.*, 2012b; Sebold *et al.*, 2014), the study was also designed to assess these factors as additional moderators of behavioral control.

9.2 Materials and methods

Participants. 20 healthy participants with positive family history of alcohol dependence were recruited based on the CAST-6 (Children of Alcoholics Screening Test (Hodgins et al., 1993). Only individuals with a score >= 5 were included (usually a score of score of >= 2 is considered as a positive family history, Hodgins et al., 1993). In the lab, participants were additionally interviewed on parental alcohol consumption, confirming their fathers' fulfillment of DSM-IV criteria of addiction. To exclude any influence of potential prenatal alcohol exposure, only individuals with a father suffering from alcohol dependence were included. 17 healthy participants without positive family history of alcohol use disorders (CAST-6 score of zero and no indication of any substance abuse for 1st – 3rd degree relatives in a personal interview) were included as a control group. Both groups did not differ in age or gender distribution and were screened for axis-1 psychiatric disorders using the SCID-IV interview (First, 1997). None of the participants fulfilled criteria of an axis-1 disorder at the time of the study.

To further characterize the sample, all participants underwent neuropsychological assessment containing 4 tests: the Digit Symbol Substitution Test (DSST, Wechsler, 1955) and the Reitan Trailmaking Test (TMT, Reitan, 1955), part A for processing speed, TMT part B for complex attention/executive function and Backward Digit Span Test (DS, Wechsler, 1955) for working memory. All test scores were z-transformed and z-transformed scores of all four tests were averaged for a composite measurement of cognitive capacities (compare Schlagenhauf *et al.*, 2013). Crystalline intelligence was examined based on a German vocabulary test (Schmidt & Metzler, 1992). In addition, participants completed the BIS-11 (Patton *et al.*, 1995; Stanford *et al.*, 2009), a well-established measurement to assess trait impulsivity. Participants also indicated alcohol consumption in the preceding 4 weeks using the Time Line Follow Back questionnaire (Sobell, 1992). For a detailed group description, please see Table 9-1. The study was approved by the local ethics committee and written informed consent was obtained from all volunteers. Participants were reimbursed on an hourly basis.

Table 9-1. Sample characteristics of the original sample. Group means with standard deviations and range in brackets are reported; for group comparisons two-tailed two-sample t-test or Chi-Square Tests were used. DSST: Digit Symbol Substitution Test, TMT: Trail Making Test, BIS: Barrat Impulsiveness Scale, CAST 6: Children of Alcoholics Screening Test, TLFB: Time Line Follow Back Questionnaire.

	positive family history of alcohol-dependence (n=20)	negative family history of alcohol-dependence (n=17)	Sig.
Age (years)	28.65 ± 5.76 (19-42)	29.24 ± 5.47 (21-41)	.76
Gender	10 female /10 male	8 female / 9 male	.86
DSST (19/16)	83.89 ± 10.55 (70-105)	86.75 ± 16.76 (57-120)	.54
TMT A (19/16)	26.62 ± 8.55 (12-45)	20.46 ± 6.13 (9-31)	.02
TMT B (19/16)	54.33 ± 26.92 (30-95)	51.06 ± 19.73 (16-88)	.61
Digit span (19/16)	7.95 ± 2.48 (4-13)	8.06 ± 2.77 (4-14)	.90
Z-fluid IQ (19/16)	0.14±0.66 (-1.33-0.91)	0.16±0.89 (-1.61-1.83)	.28
Verbal IQ (19/16)	109.79 ± 9.31 (92-129)	112.38 ± 9.14 (97-133)	.42
BIS total (18/16)	60 ± 7.76 (49-73)	59.63 ± 7.51 (45-74)	.89
CAST-6	5.70±.47 (5-6)	0	<.001
TLFB (18/17)	19.39 ± 17.97 (1-70)	20.32±24.81 (0-98)	.90

Sequential decision-making task. A two-step choice task was implemented as in previous studies (e.g., Daw et al., 2011; Deserno et al., 2015b). The task consisted of 201 trials, each trial involved two choice stages. At each stage, subjects were required to give a forced choice (maximum decision time 2s) between two stimuli presented; stimuli were two gray boxes at the first stage and two pairs of differently colored boxes at the second stage (Figure 9-1A). Position of the screen where stimuli were presented (left vs. right) was randomized over trials. After a choice the respective stimulus was framed in red, moved to the top of the screen and remained there for 1.5s. Rewards were delivered only after the second-stage choice. Reward probabilities of second-stage stimuli were identical to (Daw et al., 2011). First and second stage choices were connected via a fixed transition probability: each first-stage choice was associated with one pair of the second-stage stimuli via a fixed probability of 70% (Figure 9-1B). Each trial was ended by an exponentially distributed inter-trial interval (ITI) with a mean of 2s.

During an instruction session prior to the experiment, participants were explicitly informed that the transition structure would not change throughout the task. Participants were also told about the independence of reward probabilities and their dynamic change over the course of the experiment. Participants were instructed to win as much money as possible and informed that the balance of their account would be paid out in addition to the reimbursement for study participation. After detailed instruction including teach-back, participants trained on a shortened version of the task (50 trials) with different reward probabilities and stimuli.

Behavioral data analysis. Data were analyzed using MATLAB R2012 and Statistics Toolbox Release 2012b (The MathWorks, Inc., Natick, Massachusetts, United States), IBM SPSS Statistics for Windows, Version 22 (IBM Corp., Armonk, NY) and R (R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/). Stay-switch behavior on the first step was analyzed as a function of reward (reward or no reward) and state (common or rare) on the previous trial. Individual stay probabilities were subjected to a repeated-measures ANOVA with reward and state as within-subject factors and group as between-subject factors.

Computational modeling. The aim of model-free and model-based algorithms is to learn values for each of the stimuli, which appear in the task as three pairs (sA, sB, sC). sA refers to the first-stage stimuli where values derived from model-free and model-based algorithms differ. sB and sC refer to the two pairs of second-stage stimuli. a refers to the chosen stimuli. The index i denotes the two stages of the task (i=1 for SA at the first stage and i=2 for SB or SC at the second stage) and the index t denotes the trial.

First, the model-free algorithm was SARSA(λ)(Sutton & Barto, 1998):

$$Q_{MF}(s_{i,t+1}, a_{i,t+1}) = Q_{MF}(s_{i,t}, a_{i,t}) + \alpha_i \delta_{i,t} \quad (16)$$

$$\delta_{i,t} = r_{i,t} + Q_{MF}(s_{i+1,t}, a_{i+1,t}) - Q_{MF}(s_{i,t}, a_{i,t}) \quad (17)$$

Notably, $r_{1,t}$ = 0, because no reward is delivered after a first-stage choice. Further, we allow for an additional stage-skipping update of first-stage values by introducing the parameter λ, which connects the two stages and allows the reward prediction error at the second stage to influence first-stage values:

$$Q_{MF}(s_{1,t}, a_{1,t}) = Q_{MF}(s_{1,t}, a_{1,t}) + \alpha_1 \lambda \delta_{2,t} \quad (18)$$

λ additionally accounts for the main effect of reward as observed in the analysis of first-stage stay-switch behavior but does not reflect the interaction of reward and state. Instead, the influence of learning values for the transition matrix accounts for the interaction of reward x state.

Second, the model-based algorithm learns values in a forward-planning way and computes first-stage values by multiplying maximum values at the second stage (model-free algorithm) with transition probabilities:

$$Q_{MB}(s_A, a_j) = Q_{MB}(S_B | S_A, a_j) \; max \; Q_{MF}(s_B, a) + Q_{MB}(S_C | S_A, a_j) \; max \; Q_{MF}(s_C, a) \quad (19)$$

Note that this approach simplifies transition learning because transition probabilities are not learned explicitly. This approach is in line with the task instructions. Daw and colleagues report

a simulation that verified that this approach outperforms incremental learning of the transition matrix (Daw *et al.*, 2011).

Third, the hybrid algorithm connects model-free and model-based learning via the weighting factor ω:

$$(20) \qquad Q(s_A, a_j) = \omega Q_{MB}(s_A, a_j) + (1-\omega) Q_{MF}(s_A, a_j)$$

Importantly, ω reflects the relative influence of model-free and model-based values and is therefore the parameter of most interest, representing the balance of the two decision-making systems.

Finally, we transform values into action probabilities using a softmax for Q:

$$(21) \qquad p(a_{i,t} = a | s_{i,t}) = \frac{exp(\beta_i [Q(s_{i,t}, a) + \rho * rep(a)])}{\sum_{a'} exp(\beta_i [Q(s_{i,t}, a') + \rho * rep(a')])}$$

Here, β_i controls the stochasticity of the choices and stochasticity is assumed to be different between the two stages. The additional parameter ρ captures first-stage choice perseveration and rep is an indicator function that equals 1 if the previous first-stage choice was the same.

In summary, the algorithm totals 7 parameters. It can be reduced to its special cases ω = 1 (4 parameters) and ω = 0 (5 parameters).

Model fitting. We fit bounded parameters by transforming them to a logistic (α, λ, ω) or exponential (β) distribution to render normally distributed parameter estimates. To infer the maximum-a-posteriori estimate *MAP* of parameters θ, we use a Gaussian prior with mean μ and variance σ:

$$(22) \qquad MAP_i = \text{argmax} \log p(Y | \theta) \; p(\theta | \mu, \sigma)$$

where Y represents the data in terms of actions A_i per subject *i*. We set priors empirically to the maximum likelihood estimates *ML* of μ and σ given the data by all subjects:

$$(23) \qquad ML_i = \text{argmax} \log p(Y | \theta)$$

and achieve this by using Expectation-Maximization. For an in-depth description please compare Huys *et al.*, 2011; Huys *et al.*, 2012. All *7* parameters of the best-fitting model (Table 9-2) were subjected to a multivariate ANOVA with group (family history: positive/negative) as between-subject factor.

Model selection. To compare models for their relative goodness of fit, we compute the model evidence by integrating out free parameters. This integral was approximated by sampling from the empirical prior distribution (Huys *et al.*, 2011; Huys *et al.*, 2012). The integrated likelihood was subjected to the spm_BMS function contained in SPM8 (http://www.fil.ion.ucl.ac.uk/spm/) to compute expected posterior probabilities and their exceedance probabilities (Stephan *et al.*, 2009).

Regression models and moderator analyses. To test the potential influence of impulsivity and cognitive capacities on the balance of model-based and model-free behavioral control, we built a linear regression model with ω as dependent variable and family history (positive vs. negative), impulsivity (BIS-11 total score), as well as the sum score across cognitive capacities as predictor variables. In all models, we additionally included age as nuisance variable as it is known to impact model-based behavior (Eppinger et al., 2013). Further, the negative log-likelihood of the hybrid model was included as independent variable to control for unspecific effects of individual variability in model fit. To test potentially interacting effects of the risk factors on ω, we applied moderator analyses (Hayes & Matthes, 2009).

9.3 Results

Behavioral raw data. As in previous studies with the same task (e.g. Daw et al., 2011; Deserno et al., 2015b), analysis of stay-switch behavior at the first-stage as a function of reward and state in the previous trial revealed a main effect of reward ($F(1,35) = 23.657$, $p < .001$) and a reward x state interaction effect on first-stage decisions $F(1,35) = 43.826$, $p < .001$, Figure 9-1C). In individuals with a positive family history of alcohol dependence neither evidence for a reduction of model-based choices (reward x state x family history interaction $F(1,35) = .570$, $p = .461$, Figure 9-1C) nor for a shift toward model-free control (reward x family history interaction $F(1,35) =$

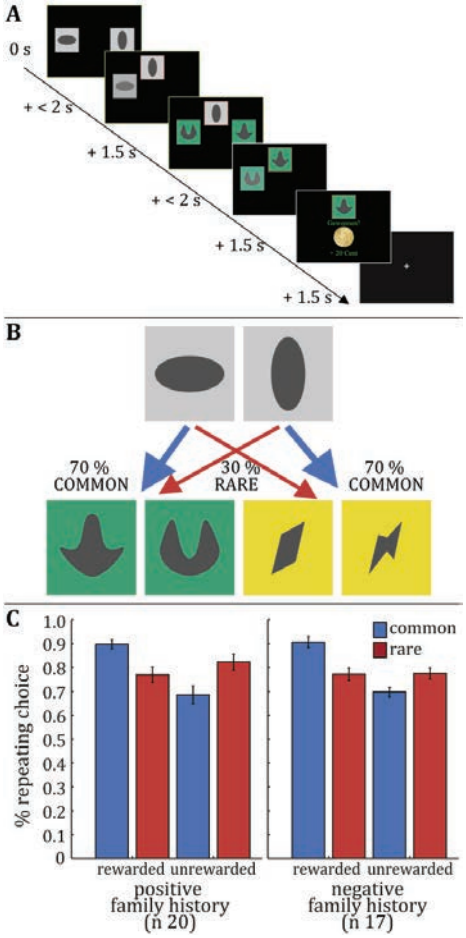


Figure 9-1. Task and Raw Data Results. A) Exemplary trial sequence. At each stage, subjects made a choice (maximum decision time 2s) between two stimuli presented: two gray boxes at the first stage and two pairs of differently colored boxes at the second stage. After this choice the respective stimulus was framed in red, moved to the top of the screen and remained there for 1.5s. before the subject entered the second stage, where another choice had to be made. Reward was delivered after the second-stage choice. B) First and second stage choices were linked via a fixed transition probability: each first-stage choice led one pair of the second-stage stimuli with probability of 70% C) Stay-switch behavior at the first-stage of the task was analyzed as function of reward and state in the previous trial. These stay probabilities were subjected to repeated-measures ANOVAs with reward and state as within-subject factors and group as between-subject factors. We observed significant main effect of reward (F=23.66, p<.001) and reward x state interaction (F=43.83, p<.001); no significant main effect of state (F=.95, p=.34) and no significant reward x group (F=.38, p=.54), state x group (F=1.85, p=.18) or reward x state x group (F=.57, p=.46) interactions.

Table 9-2 Distribution of best-fitting parameters (hybrid model). Min: minimum, 1st / 3rd Qu: first and third quartile, Max: maximum

	Min	1st Qu.	Median	Mean	3rd Qu.	Max
β_1	2.56	5.09	6.47	7.02	8.23	14.11
β_2	1.38	2.64	3.15	3.63	4.19	7.96
α_1	.07	.36	.50	.51	.67	.83
α_2	.04	.41	.53	.50	.66	.91
λ	.23	.49	.65	.62	.78	.93
ω	.34	.55	.66	.63	.73	.83
ρ	.02	.09	.15	.16	.21	.32
-LL	-280.60	-217.70	-188.80	-183.90	-156.80	-88.43

.379, p=.542, Figure 9-1C), nor a main effect of group on stay/switch behavior (F(1,35)=.029, p=.864) was observed.

Computational modeling. We compared three computational models: a model-based algorithm (ω=1), a model-free algorithm (ω=0) and a hybrid model with ω as a free parameter. Confirming previous studies with the same task and modeling analysis, model selection across all participants revealed that the hybrid model explained the observed choice behavior best ($XP_{model\text{-}based}$=.029, $XP_{model\text{-}free}$=.006, $XP_{hybrid\ model}$=.965). With respect to family history of alcohol dependence, we tested for between group differences by subjecting all 7 parameters of the hybrid model to a multivariate ANOVA with family history (positive/negative) as between-subject factor. There was no significant effect of group (F(7,29)=.760, p=.280).

Next, we aimed to probe whether the effect of cognitive capacities on model-based control is moderated by the two risk factors family history and impulsivity, respectively. See Figure 9-2 for the distribution of impulsivity scores in this sample and the subsequently described replication sample. In the respective moderator analyses, the interaction between positive family and cognitive capacities did not show a significant effect (R^2-change due to interaction=.002, F=.056, p=.814), whereas the interaction effect between impulsivity (BIS score) and cognitive capacities on ω was significant (R^2-change due to interaction=.134, F=5.394, p=.028, Figure 9-3A). We also explored the included neurocognitive subdomains by subsequently entering the four different test scores (TMT A, TMT B, DS, DSST) as independent variable in separate moderator analyses (dependent variable ω, moderator variable impulsivity). This revealed a positive effect

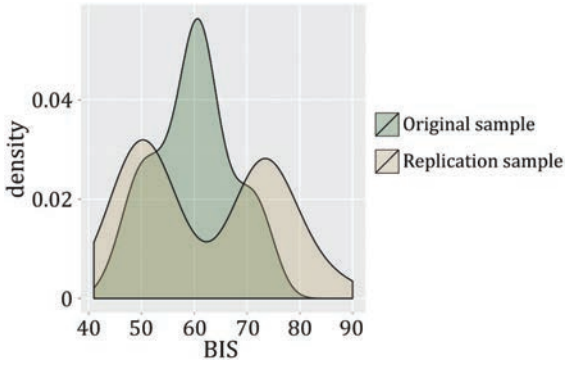


Figure 9-2. Density function of BIS-11 values in the original sample and the replication sample. The different distributions are due to differences in recruitment strategy: in the replication sample, participants were specifically chosen based on particularly low vs. high values on the BIS-11.

on the association of impulsivity and model-based behavior, whereas the other cognitive subdomains failed to contribute significantly to an interaction effect (TMT A p=.119, DS p=.0763, DSST p=.100). We explored this interaction effect in a post-hoc fashion by using the median of the BIS score to split our sample in a subgroup with higher vs. lower trait impulsivity scores. The regression analysis was then repeated for those groups separately. We observed that the interaction effect between impulsivity and executive control was driven by a significant effect of executive control on ω in the lower impulsive subgroup (*beta*=.591, t=2.574, p=.024). Contrary, there was no association of executive control and ω in the higher impulsive subgroup (*beta*>-.110, t<.341, p>.738). See Figure 9-4A for an illustration.

Cognitive capacities, impulsivity and model-based choices: replication analysis in an independent sample.

Using the same task and computational modeling analysis, we previously investigated the association between high vs. low trait impulsivity (defined according to BIS-11) and the balance between model-based and model-free decision-making (Deserno et al., in press). For a detailed description of recruitment strategy, sample characteristics and results, please compare Deserno et al., in press. In short, no evidence for an influence of high vs. low trait impulsivity on the parameter ω was found. Given the above reported findings, we now reanalyzed these data with respect to an interaction effect of impulsivity and cognitive capacity on ω, an analysis that had

not been performed in the original investigation. All datasets included in the previous investigation (n=50, 24 high-impulsive and 26 low-impulsive participants) were reanalyzed for replication. It is important to note that in the original study, participants were selected from the upper and lower ends of the BIS-11 range in a larger sample (n=453) according to their particularly high vs. low values in the BIS-11. According to the previous literature (Stanford et al., 2009), mean total BIS-scores of each group met the criteria for high or low impulsiveness, respectively. This difference in study design results in a different distribution of total BIS scores in the replication sample from Deserno et al., in press as compared to the original sample of participants with and without positive family history of alcohol-dependence. Specifically, the respective high vs. low impulsive subgroups of both samples were significantly different from each other (comparing lower impulsive groups of the original and replication sample using an independent samples t-test: $mean_{original\_sample}$=53.625, *SD*=4.674, $mean_{replication\_sample}$=50.308, *SD*=3.782, *t*(26.900)=2.397, *p*=.024; comparing higher impulsive groups of the original and replication sample: $mean_{original\_sample}$=65.333, *SD*=4.703, $mean_{replication\_sample}$=74.792, *SD*=5.065, *t*(38.000)=6.214, p<.001). See Figure 9-2 for a plot of the distribution of BIS values in the two samples.

We repeated the identical analyses as described above in the replication sample: a moderator analysis with ω as dependent variable, independent variable sum score cognitive capacities, moderator variable impulsivity (high vs. low), as well as age and negative log-likelihood of the hybrid model as nuisance variables was conducted. Replicating the findings in the original sample, we again found a significant interaction effect of impulsivity and cognitive capacities on ω (R^2 increase due to interaction=.081, *F*=4.669, *p*=.036, Figure 9-3B). Next, we again tested for the role of the included cognitive subdomains and thus entered the four different test scores (TMT A, TMT B, DS, DSST) as independent variables in separate moderator analyses (dependent variable ω, moderator variable impulsivity). This revealed a significant effect of the interaction "cognitive speed (DSST) by impulsivity" (R^2 increase due to interaction=.112, *F*=6.906, *p*=.012), and of the interaction "attention (TMT A) by impulsivity" (R^2 increase due to interaction=.082,

F=64.594, *p*=.038) on ω. Executive control (TMT B, R^2 increase due to interaction=.048, *F*=2.685, *p*=.109) and working memory (Digit Span, R^2 increase due to interaction=<.001, *F*=.043, *p*=.836) did not significantly interact with impulsivity in their effect on ω. Post-hoc regression analyses for both groups (high vs. low impulsive individuals) separately revealed that this effect was driven by a significant relation of cognitive speed (DSST) on ω in the high-impulsive group (*beta*=.503, *t*=2.683, *p*=.014); this was absent in the low-impulsive group (*beta*=-.126, *t*=.629, *p*=.536, Figure 9-4B). Post-hoc analyses for TMT-A did not indicate a significant effect of TMT-A in any of the groups (*beta* <.309, *t*<1.649, *p*>.11).

9.4 Discussion

In the present study, we did not observe evidence for altered model-free and model-based instrumental control in adult participants with an alcohol-dependent father. Independent of family history, our findings however suggest that an interaction of impulsivity and cognitive capacities influences the degree of model-based decision-making. The latter effect could be replicated in an independent sample of high and low impulsive individuals (Deserno *et al.*, in press).

Family history of addiction and model-based control. The present work does not provide evidence in favor of a shift from model-based to model-free control in healthy participants with family history of alcohol-dependence. At first glance, this seems in contrast to findings with the same sequential decision task in addicted and other psychiatric patient samples characterized by loss over behavioral control (Sebold *et al.*, 2014; Voon *et al.*, 2014): in these two studies, patients suffering from addictive and other compulsive disorders showed reduced model-based control. It is interesting that, specifically for alcohol-addiction, after a closer look, a more complex picture arises: Voon and colleagues found no reduction of model-based control in alcohol-dependent subjects per se but a correlation of model-based control with duration of abstinence (Voon *et al.*, 2015). In the study by Sebold and colleagues, reduced model-based control was found in alcohol-dependent patients overall but effects were attenuated when adjusting for general cognitive functioning (Sebold *et al.*, 2014). Based on the presented null finding in

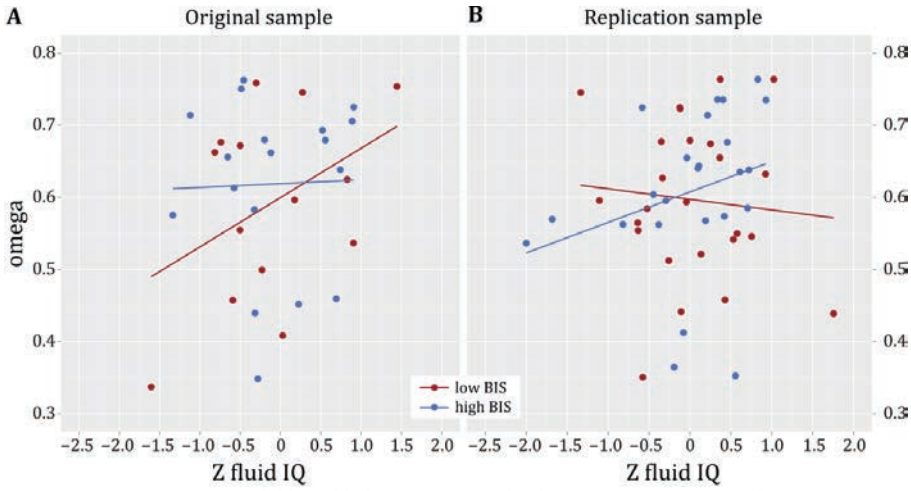


Figure 9-3. Association of model-based behavior (as given by the parameter omega) with cognitive capacity (Z-score of fluid intelligence) in the lower, but not in the higher impulsive group of the original sample. In the replication sample, a positive association of omega with cognitive capacity was found. In the original sample, high and low impulsive groups were defined based on a median split. In the replication sample, groups were defined by sampling from the upper and lower ends of the BIS-11 range in a larger sample (n=453) according to their particularly high vs. low values in the BIS-11 (Deserno *et al.*, in press), see study 4.

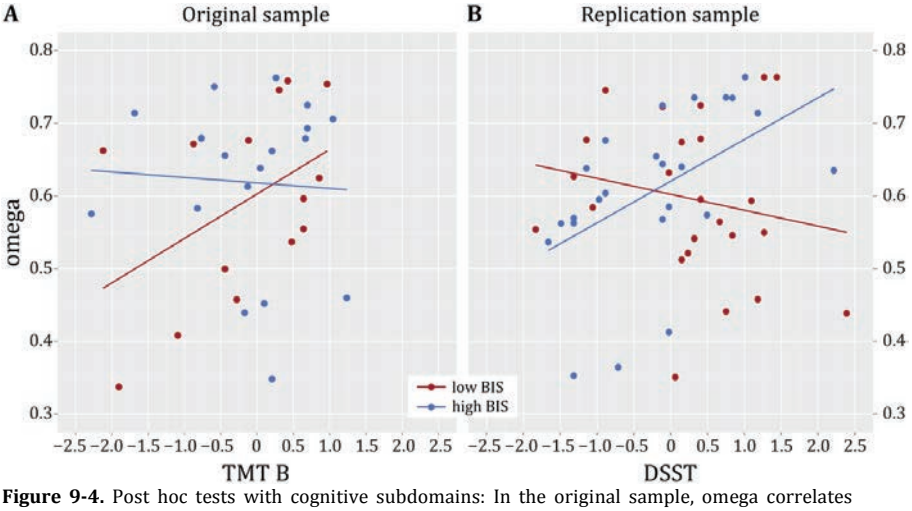


Figure 9-4. Post hoc tests with cognitive subdomains: In the original sample, omega correlates positively with TMT B in the low impulsivity group. In the replication sample, omega correlates positively with DSST scores in the high impulsivity group. In the original sample, high and low impulsive groups were defined based on a median split. In the replication sample, groups were defined by sampling from the upper and lower ends of the BIS-11 range in a larger sample (n=453) according to their particularly high vs. low values in the BIS-11 (Deserno et al., in press), see study 4. We plot z-transformed scores of the cognitive test scores.

relatives, one might speculate that a bias toward model-free control arises as a consequence of chronic alcohol-consumption rather than preceding it as a vulnerability factor.An additional explanation takes into account our cross-sectional design and the inclusion criteria of unaffected adult participants without any indication of alcohol abuse or other addictive behavior and within an age-range that exceeds the typical onset of addictive disorders. Given this sample selection strategy, participants included in this study might be those who were particularly resilient *not* to develop an addictive disorder - and thus show no alteration in behavioral control. In line, Volkow and colleagues have found putatively protective traits in terms of dopaminergic neurotransmission in a similar sample of unaffected adult relatives of addicted patients (Volkow *et al.*, 2006) and follow the same line of reasoning. To tackle this important question appropriately, longitudinal designs are required to map instrumental control across the developmental process from risk to addiction in adolescence to abstinence and potential relapse in adulthood.

Further, it is to be noted that our study comprised a rather small sample-size albeit in a similar range as the previous between-group patient studies (Sebold *et al.*, 2014). Thus, the null finding of an absent association between family-history of addiction requires replication in a larger population, ideally including 1st degree relatives not only of alcohol-dependent subjects, but also of other substance addictions and other psychiatric states characterized by loss of behavioral control like obsessive-compulsive disorder or binge eating.

Addiction, cognitive capacities and instrumental control. Studies suggest that inter-individual variability in cognitive capacities relates to a model-based system (Otto *et al.*, 2013a; Otto *et al.*, 2013b; Schad *et al.*, 2014) which was indeed shown to moderate group differences in studies involving patients characterized by cognitive impairment (Sebold *et al.*, 2014). These and our findings suggest that, when observing differences on instrumental control between groups which differ systematically in cognitive factors, one ought to tread carefully when interpreting these; differences might be an epiphenomenon of a more general impairment rather than a specific characteristic for alcohol-dependence. This is also in line with a study using instructed

devaluation tasks in alcohol-dependent patients, which revealed a global impairment in learning per se (Sjoerds *et al.*, 2013). Here, we replicate the previously reported correlation between cognitive function and model-based control per se and find evidence for interaction effects of between cognitive function and impulsivity, a recognized risk factor for addiction, on model-based control.

Impulsivity, Cognitive Capacities and Instrumental Control. Interestingly, cognitive dysfunction itself (more specifically, executive functioning), as well as the combination of deficits in cognitive function and impulsivity have been suggested as endophenotypes for drug dependence (Ersche *et al.*, 2012b). Thus, our finding of an interaction of cognitive functioning and impulsivity on model-based behavior in two independent samples amend previous studies reporting on an influence of impulsivity on reduced goal-directed control in a devaluation task (Hogarth *et al.*, 2012b) or on accentuated model-free control together with intact model-based control (Deserno *et al.*, in press). Our finding and its replication in an independent sample suggest that the interaction of cognitive capacities and impulsivity plays an important role. Interestingly, in the sample at hand, for which impulsivity measures were not the selection criterion, a positive correlation of cognitive capacity and model-based behavioral control was found in the relatively lower impulsive group. To interpret this finding, one might speculate that relatively, but not extremely low impulsiveness in addition to relatively high cognitive capacities provides an optimal ground for model-based control in this task. In the replication sample, which was specifically recruited to consist of a low vs. a high impulsive group from the extreme ends of the impulsivity measures, impulsivity scores at the higher end matched those of addicted patients. Interestingly, in this sample the correlation of cognitive speed with model-based behavior was driven by the high-impulsive group, suggesting it as a potential compensatory factor in high-impulsive individuals, which were initially assumed to be impaired in goal-directed, model-based behavioral control (Hogarth *et al.*, 2012b; Deserno *et al.*, in press).

Conclusion. In sum, we did not find evidence for an influence of the risk factors positive family history or impulsivity on model-based control per se. Our findings could speak in favor of a

multiple hits account with different risk conditions playing together to impair or protect model-based behavioral control. Longitudinal designs might help to disentangle these rather complicated interaction effects on model-based control and eventually, on the potential development of addiction.

10 Study 6: The interaction of acute and chronic stress impairs model-based behavior[7]

10.1 Introduction

Making effective decisions is particularly relevant in stressful situations and may depend on individual responsiveness during acute stress as well as on the long-term stress load. Dual-system theories of decision-making postulate a goal-directed system and a habitual system to compete for behavioral control (Balleine & Dickinson, 1998; Balleine & O'Doherty, 2010). Recently, computational modeling accounts of reinforcement learning have amended these theories (Daw *et al.*, 2005): here, goal-directed, model-based behavior is seen as a flexible, albeit computationally complex strategy, which builds an internal mental model of the environment. Thereby, future actions and their potential outcomes are planned in a forward manner. In contrast, habitual, model-free control is seen as a retrospective and therefore more rigid strategy driven by past rewards, which neglects environmental structure for the advantage of computational efficiency. Crucially, human decision-making involves both control systems with considerable inter-individual variability (Daw *et al.*, 2011). However, it remains an intriguing question how control over actions is allocated between the two systems depending on the particular situation and on inter-individual trait differences (Dolan & Dayan, 2013).

Among situational factors that influence this allocation of control, stress is a key candidate for biasing the balance of the two systems towards habitual decision-making (Schwabe & Wolf, 2009; 2011; 2013). At the neurobiological level, cortisol, the endproduct of the hypothalamus-pituitary-adrenal (HPA-) axis, might affect prefrontal executive capacities, which may thus limit the degree of control exerted by the more sophisticated, model-based system. On the behavioral level, stress has been shown to influence decision-making, e.g. in terms of dysfunctional strategy use, automatic responding, goal implementation, response conflicts, risk

[7] This chapter corresponds to the following article: Radenbach, C.*, Reiter, A.M.F*., Engert, V., Sjoerds, Z., Villringer, A., Heinze, H.J., Deserno, L. & Schlagenhauf, F. (2015). The interaction of acute and chronic stress impairs model-based behavioral control. *Psychoneuroendocrinology*, *53*, 268-280.

*authors contributed equally

taking, feedback processing per se and reward vs. punishment sensitivity (Petzold *et al.*, 2010; Plessow *et al.*, 2011; Plessow *et al.*, 2012; Starcke & Brand, 2012). In a recent study, Otto *et al.* (2013b) compared acutely stressed and non-stressed participants and did not observe between-group differences in the balance of behavioral control. However, inter-individual differences in physiological stress response, as measured by cortisol increase, were negatively correlated with the degree of model-based control across both groups. Importantly, this points to the direction that inter-individual differences in stress reactivity, rather than a stress-eliciting condition per se, might impact decision-making.

Beyond acute stress, animal studies suggest that chronic stress shifts decision-making towards more habitual strategies: Dias-Ferreira *et al.* (2009) observed that chronically stressed rats became insensitive to outcome devaluation, a key characteristic of habitual behavior. In humans, the effect of chronic stress and the interplay between previous stress experience and acute stress on model-based decision-making has not yet been investigated.

Here, we utilized a within-subjects design to assess the influence of a potent acute psychosocial stressor on the balance between model-based and model-free control as assessed via sequential decision-making (Daw *et al.*, 2011). By means of computational modeling, we first asked if acute psychosocial stress diminishes the degree of model-based control within individuals. Second, we tested if inter-individual variations in physiological and subjective stress reactivity predict the balance of behavioral control per se. Finally, we examined the interaction of chronic and acute stress levels in human decision-making.

10.2 Materials and Methods

Participants. Thirty-nine healthy male subjects recruited by Internet advertisements completed the study (mean age: 25.2, *SD*=2.7, range: 21–30 years). All participants except for one had obtained university entrance qualification, one held the general certificate of secondary education. The average years of education (including school, university etc.) was 16.32 (*SD*=3.21), the average duration of unemployment counted 0.19 years (*SD*=.44). Exclusion criteria comprised presence or history of any neurological or psychiatric disorder and smoking,

as nicotine impacts the neuroendocrine stress response (Mendelson *et al.*, 2005). Exclusion criteria were assessed prior to study participation during a semi-structured telephone-screening. The study was approved by the ethics committee of the University of Leipzig. Written informed consent was obtained from all participants prior to the study.

Procedure. In two separate sessions (interval between test sessions: M = 7.03 days, SD = 0.28), participants performed a Markov two-step sequential decision task (Daw *et al.*, 2011). One of the sessions involved the standardized protocol Trier Social Stress Test (TSST, Kirschbaum *et al.*, 1993) to induce psychosocial stress before task performance (stress condition). In the control session, individuals were asked to read a neutral text before they executed the decision task (control condition). The order of the two sessions (stress vs. control) was counterbalanced across all participants (Figure 10-1). On the first test day, participants were introduced to the study procedures, provided with instructions, and underwent training of the decision task. Before the stress or control protocol was applied, participants rested for 16 min in order to adapt to the testing situation. Importantly, both experimental sessions were scheduled at exactly the same time of day and always between 12:00 pm and 6:00 pm to control for circadian effects on task performance and cortisol levels (Kudielka *et al.*, 2004). In a third test session, trait questionnaires and working-memory were assessed (interval between the second and the third session: M=15.69 days, SD=17.50)

Stress protocol. The TSST is a well-established experimental protocol to reliably induce acute psychosocial stress in the laboratory and to prompt an increase in saliva cortisol levels as described in detail elsewhere (Kirschbaum *et al.*, 1993; Kudielka *et al.*, 2007). After a 10 min anticipation period participants were asked to assume the role of a job applicant and to present themselves in front of an evaluation committee while they thought they were being audio- and video-recorded (5 min). The job talk was followed by a challenging 5-minute mental arithmetic task under evaluation by the committee. In the control condition, participants were undisturbed and requested to read a neutral, non-arousing, non-fictional text on the Mesozoic era for 20 minutes.

Assessment of cortisol response. To assess the physiological stress response of the hypothalamic pituitary adrenal axis, salivary cortisol was acquired five times throughout the experiment (Figure 10-1). Samples were collected using a Salivette device (SalivetteCortisol®, Sarstedt, Nuembrecht, Germany) at the following time points: after 16 min of rest (baseline, t1), directly after the termination of the TSST or control condition, respectively (t2), ~20 min (t3) and ~35 (t4) min after the onset of the stressor or control condition (during short breaks in the experimental task) and after completion of the task (~62 min after the onset, t5). Saliva samples were frozen at -20°C and analyzed using a time-resolved fluorescence immunoassay with a cortisol – biotin conjugate as a tracer. The intra-assay coefficient of variation was between 4.0% and 6.7%, and the corresponding inter-assay coefficients of variation were between 7.1% - 9.0%. Different cut-offs for relevant cortisol surges have been discussed in the literature; here, using a strict cut-off, a physiologically relevant cortisol surge was defined as an increase of at least 2.5 nmol/l above the individual baseline (Van Cauter & Refetoff, 1985; Schommer *et al.*, 2003).

Cortisol values were log-transformed to approximate normal distribution and then subjected to a repeated-measures ANOVA with the within-subjects factors time (t1–t5) and acute stress (stress vs. control). In case of violation of sphericity as indicated by Mauchly's test we report p-values based on Greenhouse Geisser estimates of sphericity (p_{gg}). To compute peak cortisol increase, log-transformed cortisol levels in the stress condition were first normalized for those in the control condition by subtraction of the corresponding time points (stress - control). Subsequently, individual cortisol increases were computed by subtracting normalized baseline levels (t1) from normalized peak levels (t3) (Starcke *et al.*, 2011).

Heart Rate. As a marker of sympathetic stress response, heart rate data were collected using a POLAR RS800sd heart rate monitor (POLAR, Buettelborn, Germany). Due to technical failure data were available in n = 29 subjects.

Individual heart rate increases were computed for stress and control condition separately by subtracting the mean of a five minutes interval in the middle of the resting period from the mean of a five minutes interval in the middle of the job interview or control condition,

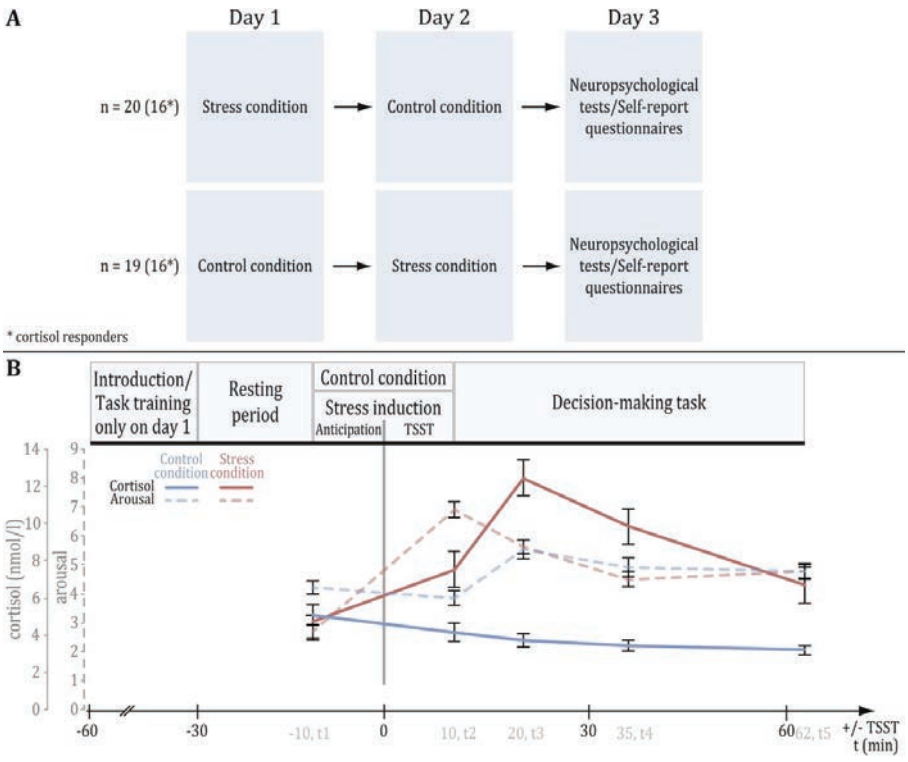


Figure 10-1 Within-Subjects study design and time line of stress intervention. A) All participants underwent the stress and the control condition. Order of conditions was counterbalanced across all participants. Trait measurements and working-memory capacity were assessed during a third test session. B) Course of one experimental session with salivary cortisol responses (in nmol/l, bold lines) and arousal response (Affect Grid Rating, dashed lines) to the stress (red lines) and the control condition (blue lines). Time scaling is relative to the onset of the Trier Social Stress Test (TSST). Note that the decision-making task was performed during the peak cortisol period in the stress condition.

respectively. A paired t-test was performed comparing increases of the stress and control condition *Subjective stress response.* At the predefined points of time, participants completed the following questionnaires (five times throughout the experimental session, respectively, Figure 10-1): the Affect Grid (Russell *et al.*, 1989), the state scale of State-Trait Anxiety Inventory (STAI;

Laux *et al.*, 1981) and the Aktuelle Stimmungsskala (ASTS), which is a German version of the Profile of Mood States scale (McNair *et al.*, 1971; Dalbert, 1992).

Arousal and *valence* (measured by the Affect Grid), *negative mood* (measured by the ASTS), and *state anxiety* (measured by the STAI) were assessed to verify the subjective effects of the stress induction by comparing the scores with the control condition. Peak increases in these scales were computed similarly to cortisol peak increase. For all correlations with arousal we use Spearman's correlation coefficient.

Sequential decision task. A two-stage decision task (Daw *et al.*, 2011; Wunderlich *et al.*, 2012b; Eppinger *et al.*, 2013; Otto *et al.*, 2013b; Smittenaar *et al.*, 2013) was used to assess the degree of model-based and model-free behavioral control.

The task was programmed in MATLAB (The MathWorks, Inc., Natick, Massachusetts, United States) with Psychophysics Toolbox extensions. It consisted of 201 trials with two stages each with a total length of approximately 35 minutes (Figure 10-2A). At the first stage, participants chose between two grey boxes randomly displayed on the left and right side of the screen each with different Chinese characters by pressing either a left or right button. The chosen stimulus was framed in red and moved to the top of the screen after the 2s decision time and remained there for 1.5s. At the second stage, one of two differently colored pairs of boxes, again with distinctive Chinese symbols, appeared on the screen and participants had to choose again between one of two boxes. Similar to the first step, the chosen stimulus was surrounded with a red frame and moved to the top of the screen. This second-stage choice could either be rewarded with 20 euro-cents or not.

Each of the first choices was predominantly associated with one of the two second-stage stimulus pairs (70% → common) and consequently less with the other (30% → rare; Figure 10-2B). These fixed transition frequencies remained constant during the task. Reward probabilities at the second stage changed slowly according to Gaussian random walks in order to induce ongoing learning (Figure 10-2C). In line with Daw *et al.* (2011), participants were explicitly introduced to the task-structure (including the stable transition frequencies and the

independent slow changes of the reward probabilities) and informed that the amount of money they would get after the testing session was completed would depend on the reward they received in the task. The instruction included a training period of 55 trials with different stimuli and reward probabilities and a post-training teach-back. In the main experiment, the task was paused after 41 and 121 trials for the collection of saliva samples (t3 and t4, see Figure 10-1).

In order to account for potential retest effects, two versions of the task differing in Gaussian random walks and Chinese characters were implemented. These two versions were counterbalanced between experimental days and participants (compare Wunderlich *et al.* (2012b) and Smittenaar *et al.* (2013) for other within-subjects designs of this task).

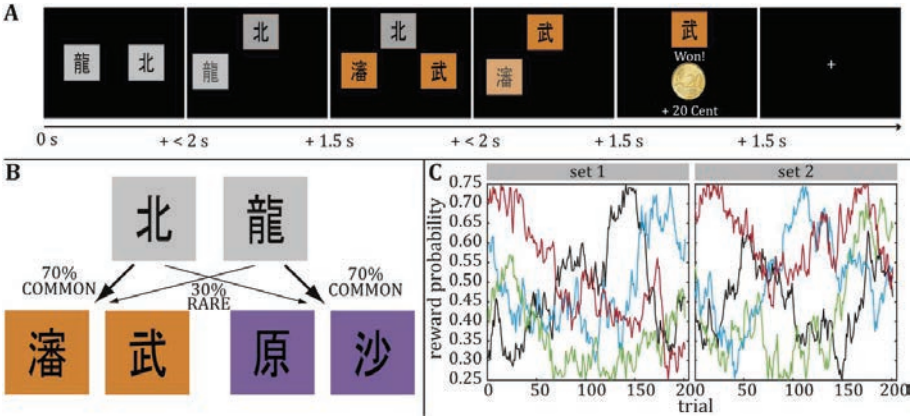


Figure 10-2. Sequential decision-making task. A) Trail sequence. Subjects are instructed to find the box with money inside and have to open first one of two grey boxes before the selected box moves up and a second pair of boxes appears. After one of those second step boxes is selected monetary reward is received depending on reward probability. B) State transition structure in the sequential decision-making task. Each first-step choice (grey box) is predominantly associated with one of the second-step states (orange and purple boxes) and leads there 70% of the time. These second-stage choices are probabilistically reinforced with money. C) Reward probabilities of the second step options change slowly over the course of the experiment according to Gaussian random walks. Random walks differed between the two experimental sessions to rule out retest effects.

Analysis of first-stage stay-switch behavior. Stay-switch behavior at the first stage was analyzed as a function of reward (reward vs. non reward) and state (common vs. rare) in the previous trial. These individual stay probabilities were subjected to a repeated-measures ANOVA with reward, state, and acute stress (stress vs. control) as within-subjects factors.

Computational Modeling. Trial-by-trial computational modeling of the observed behavioral responses is a powerful analysis technique that has recently been suggested to enrich the mechanistic understanding of stress effects on learning (Schwabe & Wolf, 2013). Here, the aim of model-free (MF) and model-based (MB) algorithms is to learn values for each of the stimuli, which appear in the task as three pairs (sA, sB, sC). sA refers to the first-stage stimuli and sB and sC to the two pairs of second-stage stimuli. In the following, a refers to the chosen stimuli and the indices i and t denote the stage (i=1 for SA at the first stage and i=2 for SB or SC at the second stage) and the trial, respectively.

First, the model-free algorithm was SARSA(λ):

$$Q_{MF}(s_{i,t+1}, a_{i,t+1}) = Q_{MF}(s_{i,t}, a_{i,t}) + \alpha_i \delta_{i,t} \quad (24)$$

$$\delta_{i,t} = r_{i,t} + Q_{MF}(s_{i+1,t}, a_{i+1,t}) - Q_{MF}(s_{i,t}, a_{i,t}) \quad (25)$$

Notably, $r_{1,t} = 0$ because no reward is delivered after a first-stage choice and $Q_{MF}(s_{3,t}, a_{3,t}) = 0$ because the task has only two states. We allowed different learning rates α_i for each stage i. Further, we allowed for an additional stage-skipping update of first-stage values by introducing another parameter λ, which connects the two stages and allows the reward prediction error at the second stage to influence first-stage values:

$$Q_{MF}(s_{1,t+1}, a_{1,t+1}) = Q_{MF}(s_{1,t}, a_{1,t}) + \alpha_1 \lambda \delta_{2,t} \quad (26)$$

The parameter λ additionally accounts for the main effect of reward as observed in the analysis of first-stage stay-switch behavior but not an interaction of reward and state.

Second, the model-based algorithm learns values in a forward-planning way and computes first-stage values by simply multiplying the better option at the second stage with the transition probabilities P:

$$(27) \qquad Q_{MB}(s_A, a_j) = Q_{MB}(S_B|S_A, a_j)\ max\ Q_{MF}(s_B, a) + Q_{MB}(S_C|S_A, a_j)\ max\ Q_{MF}(s_c, a)$$

Third, the hybrid algorithm connectsQ_{MF} and Q_{MB}:

$$(28) \qquad Q(s_A, a_j) = \omega Q_{MB}(s_A, a_j) + (1-\omega) Q_{MF}(s_A, a_j)$$

$$(29) \qquad Q = Q_{MB} = Q_{MF}$$

Importantly, ω gives a weighting of the relative influence of model-free and model-based values and is therefore the model's parameter of most interest.

Finally, we transformed values into action probabilities using a softmax for Q:

$$(30) \qquad p(a_{i,t} = a|s_{i,t}) = \frac{exp(\beta_i[Q(s_{i,t},a)+\rho * rep(a)])}{\sum_{a'} exp(\beta_i[Q(s_{i,t},a')+\rho * rep(a')])}$$

Here, β controls the stochasticity of the choices and we assume this to be different between the two stages. The additional parameter ρ captures first-stage choice perseveration and rep is an indicator function that equals 1 if the previous first-stage choice was the same. *I*n summary, the algorithm has a total of 7 parameters and can be reduced to its special cases ω = 1 (4 parameters) and ω = 0 (5 parameters). We fit bounded parameters by transforming them to a logistic (α, λ, ω) or exponential (β) distribution to render normally distributed parameter estimates. To infer the maximum-a-posteriori estimate of each parameter for each subject, we set the prior distribution to the maximum-likelihood given the data of all participants and then used Expectation-Maximization. For an in-depth description please compare Huys *et al.* (2011) and Huys *et al.* (2012). To compare models for their relative goodness of fit, we report the Bayesian Information Criterion (BIC) based on the log-likelihood (Table 10-1). Second, we computed the model evidence by integrating out the free parameters. This integral was approximated by sampling from the empirical prior distribution and we therefore added the subscript 'int' to the BIC (Table 10-1; Huys et al., 2011; Huys et al., 2012.). Third, we subject the integrated likelihood to the spm_BMS function, a random effects model selection procedure, contained in SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/) to compute so-called exceedance probabilities (Stephan *et al.*, 2009).

Testing the hypothesis that acute psychosocial stress shifts the balance of the two decision-making systems towards model-free control, the weighting parameter ω is of most interest here because it gives a measure of this balance. To analyze stress effects on model parameters, all parameters (7 in total) were entered into a repeated-measures MANOVA with the within-subjects factor condition. To test for an effect of physiological and subjective stress responses, assessed by cortisol increase and rating scales respectively, ω from the stress and control condition were entered into a repeated-measures ANOVA with the within-subjects factor condition and the following stress response measures as covariates: cortisol increase, arousal increase, valence decrease, anxiety increase, and increase of negative mood.

Influence of individual trait characteristics on stress-induced changes of the balance of the two systems. One aim of the study was to elucidate the impact of chronic stress levels and their interaction with acute stress on decision-making processes. During a third, independent test session, we therefore assessed questionnaires of chronic stress levels as perceived stress explicitly asked for the last month before the second test session (PSS-10, Cohen *et al.*, 1983) and stressful life events both within the last two years and for events within the whole life (Life Stress Scale, Holmes & Rahe, 1967). To characterize interaction effects of acute and chronic stress on the change in the balance between model-based vs. model-free control, while accounting for possible influences of working-memory capacity, scores of Life Stress (24 months and whole life, respectively), PSS-10, and the Digit Span Number-Backwards-Test (Wechsler, 1945; Von Aster *et al.*, 2006) were entered as independent variables into a multiple regression analysis. The dependent variable was Δ ω (ω_stress *minus* ω_control) which reflects the changes in the relative degree of model-based behavior under stress compared to the control condition. Individual's working-memory capacity was assessed during the third test session with the Digit Span Number-Backwards-Test. It was entered due to evidence that inter-individual differences in basic neurocognitive functioning, in particular working-memory, may play an important role in the degree of model-based control and might mediate the influence of acute stress on model-based vs. model-free decision-making strategies (Otto *et al.*, 2013a; Otto *et al.*, 2013b).

10.3 Results

10.3.1 Physiological and subjective stress response

Cortisol and Heart Rate Response. Seven participants did not display a cortisol increase of at least 2.5 nmol/l and thus were considered as non-responders (Van Cauter & Refetoff, 1985; Schommer *et al.*, 2003). The responder rate (~82%) was in line with other studies using the TSST (Kudielka *et al.*, 2007; Petzold *et al.*, 2010). Cortisol response in n = 32 responders was analyzed using a repeated-measures ANOVA with the within-subjects factors time (t1–t5) and acute stress (stress vs. control). A significant main effect of acute stress (acute stress, $F(1, 31) = 92.51$, $p < .001$, $\eta^2 = 0.749$), a significant main effect of time (time, $F(2.25, 69.66) = 39.48$, $p_{gg} < .001$, $\eta^2 = 0.56$) and a significant time and acute stress interaction (time x acute stress, $F(2.07, 64.07) = 107.41$, $p_{gg} < .001$, $\eta^2 = .776$) was found. A comparison between the peak cortisol response in the stress condition with the corresponding response in the control condition (both t3) showed a significant difference (t=14.36, p <.001, Cohen's d =.001). Baseline cortisol showed no significant difference between the conditions ($t(31) = 0.35$, p=.732, Cohen's d = .062; Figure 10-1B). The average peak was at t3. The cortisol level of the last sample which was determined after task performance was still significantly higher as compared to baseline cortisol levels ($t(31) = -6.75$, $p < .001$), indicating that the whole task was performed in a state of elevated cortisol levels.

Comparing heart rate increase using a paired t-test, we found a significantly higher increase in the stress condition than in the control condition ($t(31) = -2.39$, p =.042, Cohen's d = .983), indicating a significant increase in sympathetic nervous activity during the stress intervention.

Subjective Ratings. Subjective arousal ratings over the course of the experiment were analyzed using a repeated-measures ANOVA with the factors time (t1-t5) and acute stress (stress vs. control). We found a significant main effect of acute stress (acute stress, $F(1, 31) = 7.75$, p =.009, $\eta^2 = 0.200$), a significant main effect of time ($F(4, 124) = 8.51$, p <.001, η^2 =.215) and a significant time and acute stress interaction ($F(4, 124) = 17.75$, p =.007, $\eta^2 = 0.364$). A paired sample t-test showed that participants scored significantly higher directly after the TSST than after silent

reading during the control condition ($t(31) = -7.4$, $p < .001$, Cohen's $d = .118$). The average peak of arousal was found at t2. Ratings of unpleasantness, anxiety, and negative mood revealed similar results. Thus, subjective experience was significantly affected by the stress condition (Figure 10-1B).

10.3.2 Effects of acute social stress on model-based vs. model-free behavioral control

Analysis of stay-switch-behavior. In line with previous studies (Daw *et al.*, 2011), a three factors repeated-measures ANOVA (acute stress x reward x state) revealed a main effect of reward ($F(1, 31)=29.09$, $p<.001$, $\eta^2 = .484$) and an interaction effect of reward x state ($F(1,31) = 65.44$, $p < .001$, $\eta^2 = 0.679$) on the first-stage stay probabilities in cortisol responders (n = 32). This confirms that first-stage decisions are influenced by both rewards and states from the previous trials (Figure 10-3A). With respect to the acute psychosocial stress intervention, we found a main effect of acute stress on first-stage choices ($F(1, 31) = 5.69$, $p = .023$, $\eta^2 = .155$). This was due to higher switching in the stress compared to the control condition regardless of the previous trial's

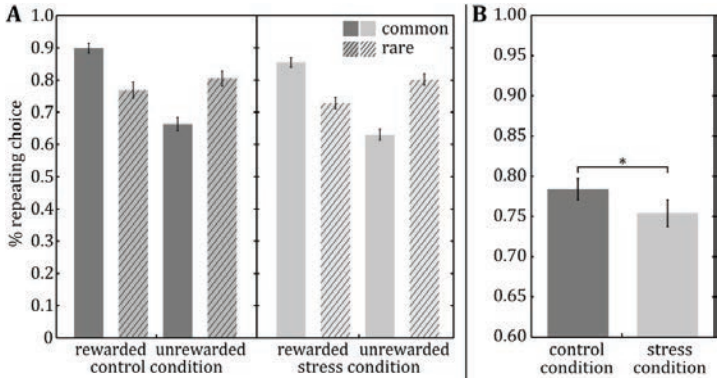


Figure 10-3. Stay probabilities of first-stage choices depending on second-stage states and rewards in the control condition as well as in the stress condition. A) In both conditions, participants' choices showed a main effect of reward and a reward x state interaction, indicating a mixture of model-based and model-free strategy to solve the task. B) Participants switched more in the stress compared to the control condition, irrespective of reward and state, indicated by overall reduced stay probabilities.

reward or state (Figure 10-3B). However, the hypothesized interaction effect of acute stress x reward x state was not significant ($F(1, 31)=.23$, $p=.634$, $\eta^2=.007$, all other interactions with acute stress *ps* >.2), indicating that the stress intervention did not influence the interaction of state and reward in the task. Thus, the analysis of stay-switch behavior does not indicate a direct impact of acute social stress on the balance between model-free and the model-based control (Figure 10-3A). Including order (stress on day 1 or day 2) as a between-subjects factor in the ANOVA did not change the observed results.

Computational Modeling. Across all participants and for both conditions, the hybrid model explained the observed data best (XP control condition = .99; XP stress condition = .99, see Table 10-1). This replicates previous studies in non-stressed participants (Daw *et al.*, 2011). Furthermore it indicates that the stress intervention did not change the learning mechanism, represented by a hybrid model engaging both systems gave the best account of the observed data in both conditions. There was no effect of acute stress on the weighting parameter ω, which represents the balance of the two decision systems ($t=.01$, $p=.99$, Cohen's $d = .011$). This is in line with the absence of such an effect on stay-switch raw data reported above. A comparison of all model parameters between stress and control condition using a repeated-measures MANOVA revealed a main effect of acute stress ($F(7, 25) = 2.92$, $p = .022$, $\eta^2 = .450$). Post-hoc paired t-tests showed significantly higher stochasticity of the participants'

choices at the first stage during stress compared with the control condition (β1: $t(31) = 2.6$, $p=.014$, Cohen's $d = .427$), which resembles the main effect of acute stress on stay-switch raw data. Further, the stage-skipping update λ, which connects reward prediction errors at the end of each trial to first-stage Q-values, was significantly lower in the stress compared to the control condition (λ: $t = 2.08$, $p = .046$, Cohen's $d = .424$). There were no differences for the remaining model parameters (beta2, alpha1, alpha2, rep: all *ps* > 0.2; Table 10-2). Including order (stress on day 1 or day 2) as a between-subjects factor in the MANOVA did not change the observed results.

Table 10-1 Model comparison. –LL: negative log-likelihood, BIC:Bayesian Information Criterion, BICint: Bayesian Information Criterion after integrating out the free parameters via sampling from the prior distribution, XP: Exceedance Probabilities after random-effects Bayesian model selection.

	-LL	BIC	BIC_{int}	XP
Control				
full hybrid model	7324	14715	15298	.9999
	Δ-LL hybrid	Δ BIC hybrid	Δ hybrid	BIC_{int}
$\lambda = 0$	-136	263	213	0
$\omega = 1$	-188	346	206	0
$\omega = 0$	-258	506	424	.0001
$\omega = 0, \lambda = 0$	-484	948	804	0
Stress				
full hybrid model	7643	15354	15908	.9997
	Δ-LL hybrid	Δ BIC hybrid	Δ hybrid	BIC_{int}
$\lambda = 0$	- 63	117	70	.0002
$\omega = 1$	-121	212	108	0.0001
$\omega = 0$	-226	441	393	0
$\omega = 0, \lambda = 0$	-345	671	554	.0001

Power analysis. For the purpose of a power analysis, we used effect sizes from two published within-subjects studies using a similar sequential decision-making task. We assume that an effect of the psychosocial stress intervention used here lies in a similar range as compared to the interventions used in these studies (a pharmacological challenge with L-DOPA [Wunderlich et al., 2012], Cohen's d = .67 and a TMS intervention [Smittenaar et al., 2013], Cohen's d = .49). Given a two-tailed alpha of .05 and sample size of 32 (cortisol non-responders excluded), power analysis revealed a power of .96 and .77, respectively. Thus, even when assuming a medium

Table 10-2. Distribution of best fitting parameters and the negative log-likelihood (-LL) of the hybrid model in n = 32 participants

	Control	Stress
$\beta 1$	7.89 (±3.07)	6.7 (±2.50)
$\beta 2$	3.80 (±1.26)	3.86 (±1.17)
$\alpha 1$	.52 (±.18)	.47 (±0.17)
$\alpha 2$	.49 (±.18)	.49 (±0.24)
λ	.55 (±.13)	.5 (±0.11)
ω	.66 (±.09)	.66 (±0.09)
ρ	.13 (±.04)	.13 (±0.04)
-LL	-182.13 (±39.29)	-188.83 (±43.48)

effect size (Cohen, 1988), the present study was well-powered to detect an effect of psychosocial stress on model-based behavior.

10.3.3 Association of subjective and physiological stress responses with model-based control

In a next step, we tested for an effect of physiological and subjective stress responses on the balance of model-free and model-based control as quantified by the parameter ω in all participants that completed the study (n = 39). Note that the effect of individual stress reactivity, including a potentially dampened cortisol response, was of central interest in this analysis. Thus, we explicitly included the cortisol non-responders into this analysis because we consider a non-significantly elevated cortisol level after stress induction as one important possible manifestation of individual cortisol reactivity. The model parameter ω during the stress and control conditions (= within-subjects factor "acute stress") was subjected to a repeated-measures ANOVA with cortisol increase, arousal increase, valence decrease, anxiety increase, and increase of negative mood as covariates (Table 10-3): this revealed a main effect of the two covariates cortisol increase and arousal increase, respectively (cortisol increase: $F(1, 33) = 6.81$,

p=.014, η^2 = .171, arousal increase: $F(1,33)$ = 9.82, p=.004, η^2 = .229) but no interactions with acute stress (all ps > 0.2). With respect to cortisol, post-hoc correlations showed that this effect was driven by a significant negative correlation of ω during acute stress with cortisol increase (i.e., the change from baseline to peak calculated from the normalized cortisol stress data; $r(38)$ = -.46, p=.004, Figure 10-4A) and a negative relationship between cortisol increase with ω during the control condition that did not reach significance ($r(38)$ = -.24, p =.149, Figure 10-4A). Regarding arousal, post-hoc correlations showed that this effect resulted from significant positive correlations between arousal increase with ω during the acute stress condition ($r(38)$= .41, p=.01, Figure 10-4B) and during the control condition ($r(38)$= .41, p=.01, Figure 10-4B). All reported correlations remained significant when excluding cortisol non-responders and when using a different normalization procedure (peak during stress induction minus resting phase measure on the same day i.e. without normalization on the control day data).

Table 10-3. Descriptive values of subjective stress-related measurements.

Stress measurement	Mean
Life stress score (whole life)	508.97 (±171.81)
Life stress score (24 months)	271.94 (±185.88)
Perceived stress scale (within the last month)	14.12 (±6.10)
Normalized arousal increase	2.62 (±2.15)
Normalized cortisol increase	9.43 (±4.13)

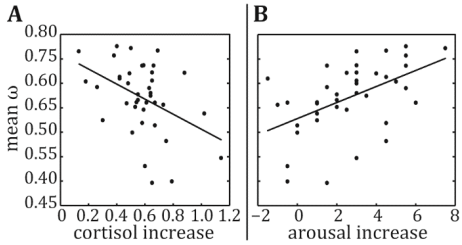


Figure 10-4. Correlation of stress reactivity with ω (averaged over the two sessions). A) Cortisol reactivity showed a significantly negative correlation with the model-parameter ω ($r(38)$ = -.46, p=.004). B) Arousal reactivity showed a significantly positive correlation with ω ($r(38)$= .41, p=.01)

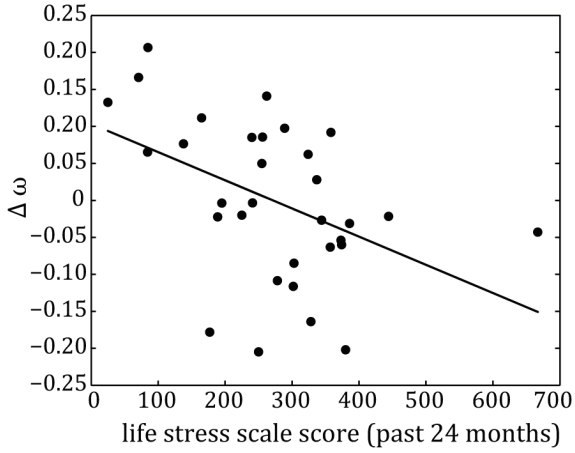


Figure 10-5. Association between chronic stress and shift in behavioral control due to acute social stress. A negative correlation between the Life Stress Scale Score and the difference score of ω (Δ ω = stress minus control condition) as determined by the computational model was found. The higher the score on the Life Stress Scale (24 months), the lower the values of the model parameter ω in the stress condition. Higher levels of chronic stress as measured by the Life Stress Scale correlate with a decrease in model-based behavior (*beta* = -.622, *t* = -2.93; *p*=.007).

10.3.4 Lifetime stress is associated with the change in model-based vs. model-free control during acute social stress

In order to investigate the influence of previous stressful life events on a change in the balance of the control systems induced by acute social stress, we calculated the difference of ω from the control and from the stress condition (Δ ω). We used a linear regression analysis for the cortisol responders (n = 32) with Δ ω as the dependent variable and Life Stress Scale (24 months and whole life), PSS-10, and the Digit Span (due to the finding of Otto et al., 2013b) as independent variables (table 10-3). Only chronic life stress in the past 24 months was significantly negatively associated with Δ ω (*beta* = -.622, *t* = -2.93; *p*=.007; Figure 10-5). This indicates that in individuals with high levels of chronic stress (24 months) the degree of model-based behavioral control is impaired after stress induction compared with the control condition. Since we

detected one outlier who scored more than three standard deviations above the mean of the chronic life stress scale for the past 24 months (z = 3.029), we repeated the linear regression analysis without this participant and found similar results (*beta* = -.648, t = -2.39; p=.036).

In an exploratory analysis, in order to test for possible neuroendocrine correlates of the chronic stress measure, a correlation of the Life Stress Score (24 months) with cortisol increase was tested, which did not indicate a significant association (r(31)= .009, p=.957).

10.4 Discussion

This study investigated the influence of stress on the balance of model-based and model-free behavioral control during a two-step decision task in a within-subjects design. First, after inducing acute psychosocial stress, we did not observe a shift toward model-free behavioral control across the entire sample. Second, we showed that variability in physiological and psychological stress reactivity is associated with inter-individual differences in the balance between model-free and model-based control. Third, we revealed that subjects with higher chronic stress as indicated by stressful life events displayed a shift towards reduced model-based control under acute social stress. Crucially, this finding demonstrates an interaction between acute and chronic stress and for the first time elucidates their joint consequences on behavioral control in humans.

Effects of acute psychosocial stress on sequential decision-making. First-stage decisions were affected by acute stress in terms of more frequent switching between options. In the computational modeling analysis, this effect was mirrored in the parameter β at the first stage, which was found to be significantly lower in the stress than in the control condition. Lower values of β indicate a higher degree of stochastic choices unrelated to the current choice value. Thus, our findings might be interpreted as pronounced although unsystematic exploration behavior triggered by acute stress. Moreover, the parameter λ, linking prediction errors at the end of each trial with first-stage choice values of the next trial, was reduced in the stress condition. As λ represents a parameter of the model-free system, hypothesized to be predominant during stress, this finding seems counterintuitive at first glance. However, in line

with the observations of enhanced stochasticity of first-stage choices under stress, it appears conceivable that in a state where decisions are marked by unspecific switching behavior at the first-stage, the use of prediction errors to update exactly these first-stage decision-values becomes attenuated.

Notably, apart from these general effects, our results do not suggest an acute stress-induced shift from model-based towards model-free strategies. In line with this, model-selection revealed no evidence for differences in model fit between the stress and control condition, supporting the conclusion that the task-solving strategy was not affected by acute stress. Further, the study was well-powered to detect such an effect. By using a different study design and stress induction protocol, these results replicate Otto *et al.* (2013b): The authors used a similar sequential decision-making task and did not find an acute effect of stress induction on behavioral control, as indicated by the reported non-significant 3-way-interaction of stress x reward x state in the decision-making task. In contrast to our findings, pioneering studies (Schwabe & Wolf, 2009; 2011) reported more habitual behavior after acute stress. However, these studies differ from our current study and Otto *et al.* (2013b) in terms of the paradigm used to examine behavioral control: there, a selective outcome devaluation protocol was applied (e.g. Valentin *et al.*, 2007). Importantly, similar to our findings, but without dissecting model-free versus model-based contributions to learning via computational modeling, Schwabe and Wolf (2009) did not observe differences in instrumental learning following acute stress. However, they provide evidence that acute stress promotes responding for the devalued outcome during a test in extinction as a measure of habitual behavior. These discrepant findings may reflect a stress-induced persistence of acquired behaviors rather than an effect of acute stress on different modes of instrumental acquisition. Accordingly, a recent validation study comparing both paradigms (selective devaluation and sequential decision making) suggests relatedness of both measurements in terms of goal-directed / model-based behavior, but points to the direction that both experiments might offer different insight into the habitual system (Friedel *et al.*, 2014).

*Effects of physiological and subjective stress reactivity on the balance of model-based vs. model-free control***.** Our findings suggest that physiological and subjective stress reactivity are associated with model-based control rather as trait factors in general, irrespectively of the acute exposure to psychosocial stress. Interestingly, both stress measures relate to behavioral control in opposite directions: stronger cortisol-reactivity was related to a lower degree of model-based control, whereas reactivity in terms of subjective arousal was associated with a higher frequency of model-based choices. These opposite effects are in line with the dissociation of subjective and physiological stress responses frequently described in the literature (Campbell & Ehlert, 2012).

Regarding physiological stress reactivity, we replicate the inverse relationship of cortisol increase and the degree of model-based behavior described by Otto *et al.* (2013b). However, we go beyond Otto et al.'s interpretation by arguing that the negative correlation across a stressed and non-stressed group reported in their between-group study likely captured a similar effect to the one observed in our within-subjects design: the importance of the persistent trait factor stress reactivity rather than an effect of acute neuromodulatory cortisol on model-based decision-making. Crucially, such an effect cannot be detected in a between-subjects design, whereas both factors are dissociable by the within-subjects design used here.

The relationship between subjective arousal reactivity and decision-making has, to our knowledge, not been directly addressed by previous studies. However, there is a substantial body of work on the interplay between arousal per se and decision-making: The somatic marker hypothesis (SMH) suggests bodily arousal feedback as a guiding influence on decision-making (Bechara *et al.*, 1996; Bechara *et al.*, 1997; Damasio, 1999; Critchley, 2005). A tendency to react to challenging environmental conditions by a higher arousal increase may therefore foster the integration of arousal feedback in the decision-making process. This might enable the individual to evaluate future outcomes of actions more precisely, boost mental mapping of environmental features and consequently promote the use of model-based strategies.

The interaction of chronic and acute stress attenuates model-based decision-making. It is widely accepted that stress influences cognition differently depending on the timing and duration of

exposure, e.g., with varying influences of acute and chronic stress (Lupien *et al.*, 2009). So far, the interaction of chronic and acute stress has not been addressed in the decision-making literature which has hitherto primarily focused on acute stress effects in between-subjects designs (Schwabe & Wolf, 2009; Otto *et al.*, 2013b). Evidence from animal studies indeed indicates that chronically stressed rats turn towards habit behavior (Dias-Ferreira *et al.*, 2009). The authors showed that chronic stress was associated with structural changes in fronto-striatal networks known to be critically involved in decision-making. One study investigating subacute stress exposure in humans suggests that this shift can be translated to human behavioral control (Soares *et al.*, 2012). Interaction effects of chronic and acute stress are widely discussed with respect to volume reduction in hippocampus and prefrontal regions and volume increase in amygdala as a result of chronic stress exposure, rendering the individual more susceptible for acute stress effects (Lupien *et al.*, 2009; Tse *et al.*, 2014). However, empirical evidence of interaction effects between acute and chronic stress on cognition is scant in humans as well as in rodents. Here, we extend these findings by providing first evidence for an interaction effect of chronic and acute stress on model-based control in human decision-making. It appears plausible that repeated exposure to stressful life events renders the individual more susceptible to detrimental changes in behavioral control brought about by an acute stressor. Further investigations are warranted to elucidate the underlying neural mechanisms of this suggested vulnerability-stress interaction.

Interestingly, stressful life events within the last 24 months, but neither whole-life stress nor perceived stress within the preceding month predicted acute stress-induced changes in model-based behavior. Given that reversibility of stress-induced effects on cognitive processes has been postulated (Luine *et al.*, 1994; Liston *et al.*, 2009; Soares *et al.*, 2012), it is conceivable that participants with more chronic stress over the whole course of their life might yet have had the chance to recover from the deteriorating effects of their stress experiences. In contrast, perceived stress within the last month might be too recent to significantly affect behavioral control in response to acute stress.

Limitations. It is to be noted that the sample selection criteria of this study might limit the generalizability of our results: only males within an age range of 21–30 years were included. Our findings emphasize the notion that inter-individual differences are crucial in the relationship between stress and model-based decision-making, and it is plausible that factors like age and gender are important covariates here (Eppinger *et al.*, 2013). We acknowledge that further studies are needed to replicate our findings in a broader range of the population.

Conclusion. We show that inter-individual differences in acute subjective and physiological stress response impact the degree of model-based behavioral control. Furthermore, a reduction in model-based control in response to acute stress was only observed in subjects with higher levels of chronic stress as indicated by a higher score in the Life Stress scale for the last 24 months.

By defining inter-individual differences in acute stress response and in chronic stress experience as crucial factors in behavioral control, our findings contribute to the intriguing question why some individuals shift towards attenuated model-based behavior whereas others do not (Dolan & Dayan, 2013; Schwabe & Wolf, 2013). This might be relevant for psychiatric conditions characterized by impaired model-based behavior like addiction, binge eating or obsessive compulsive disorder (Sebold *et al.*, 2014; Voon *et al.*, 2015) for which, notably, chronic as well as acute stress is postulated to be a pivotal factor in pathogenesis, maintenance, and relapse (Gluck *et al.*, 2004; McEwen, 2004; Gluck, 2006; Koob, 2008). Longitudinal designs are required to tackle the exact interplay of different dimensions of stress reactivity (subjective versus physiological), chronic versus acute stress and impaired model-based decision-making.

11 General Discussion

11.1 Summary

The aim of this thesis was to adopt a dimensional psychiatry approach in order to assess different aspects of goal-directed behavioral control 1) across different disorders characterized by loss of control and 2) in recognized risk factors. The studies were built on the computational formulation of a hypothesis which has traditionally been described for substance addiction: namely addiction as reduced goal-directed control with a potential overreliance on the habitual control system. In this thesis, I intended to investigate whether this hypothesis might be translated to other disorders characterized by loss of control over behavior as well as to risk factors of addiction.

The following research questions were asked:

1) Is substance addiction characterized by impaired mechanisms of flexible goal-directed behavioral adaptation? Can this deficit be explained by reduced abstract inference on un-chosen choice options ("what might have happened")? What are neural correlates of this deficit?
2) Do potential impairments of this kind transdiagnostically extend to binge eating disorder, a nosologically distinct diagnosis which shares clinical features with substance addiction? What are shared and differential behavioral mechanisms and neural correlates?
3) Does a shift from model-based to model-free behavioral control extend to recognized risk factors of addiction and might thus be seen as a vulnerability factor for addiction?

In study 1, combining counterfactual decision-making and computational modeling-informed fMRI, we showed that alcohol-dependent patients as compared to healthy controls are impaired in a key function of goal-directed decision-making: the ability to update unchosen choice values by using abstract inference on the task structure. Alcohol-dependent patients neglected the unchosen option particularly after punishment which neatly fits the clinical picture of maladaptive behavior despite negative consequences. This behavioral impairment

was accompanied by blunted abstract inference signals in mPFC. Within patients, these mPFC signals were negatively correlated with compulsive drinking habits. These findings contribute to research question 1: alcohol-dependent patients indeed show impaired mechanisms of goal-directed, flexible behavioral adaptation. While model-free learning was found to be intact in patients, one explanation for impaired behavioral adaptation in alcohol-dependent patients might be reduced abstract inference on unchosen choice options.

In study 2, the same design as in study 1 was applied to patients with binge eating disorder and matched healthy controls. In line with study 1, we could show that binge eating patients are impaired in adapting their behavior to changing environmental conditions. Our computational modeling analysis revealed that in contrast to alcohol-dependent patients, binge eating patients were not impaired in updating alternative choice options, but in balancing the exploration-exploitation trade-off appropriately. On the neural level, we detected reduced abstract inference signals in binge eating patients' medial prefrontal cortex, in a similar location as found in study 1 in alcohol-dependent patients. Furthermore, aI/vlPFC activation in explorative trials was reduced in binge eating patients compared to healthy controls. Thus, regarding research question 2, common and differential features of behavioral adaptation and its neural correlates were identified in binge eating disorder as compared to the results in alcohol addiction. This will be further discussed in the next section 11.2.

In study 3, combining the same counterfactual decision-making task as employed in study 1 and 2 with a modeling-informed parametric EEG analysis in healthy individuals, we assessed electrophysiological correlates of learning signals integrating alternative choice options. It was found that the FRN, which has previously been implicated in Reinforcement Learning, additionally codes these abstract inference components of learning signals. This opens up new possibilities to refine aspects of research questions 1 and 2 in future studies, in particular in the temporal domain.

In study 4, 5 and 6, a sequential decision-making task was used to assess whether a proposed shift from model-based to model-free control extends to risk factors of addiction. We

investigated individuals with high trait impulsivity as well as individuals with low trait impulsivity during fMRI (study 4) and adults with an alcohol-dependent father and matched healthy controls behaviorally (study 5). Importantly, for both risk factors, a reduction of model-based control per se was not found. This is in contrast to previous findings in patients (Sebold *et al.*, 2014; Voon *et al.*, 2015). For impulsivity, a subtle accentuation of model-free control, along with reduced model-based learning signals in the lateral prefrontal cortex (study 4) was observed. No such behavioral alteration could be detected in relatives (study 5). In the two independent samples from study 4 and 5, an interaction between impulsivity and cognitive capacities in predicting the degree of model-based control was found (study 5). In study 6, we probed whether acute psychosocial stress influences the balance between model-free and model-based control during sequential decision making in healthy individuals. We did not observe changes due to acute stress per se. However, associations with physiological and psychological stress reactivity were shown. Further, an interaction of acute and chronic stress was found: only in the presence of high chronic stress levels, acute stress reduced model-based behavior. With respect to research question 3, one general conclusion from the studies 4-6 is that findings in risk populations for addiction were qualitatively different to those in addiction and addiction-like disorders. This résumé is further discussed in section 11.3.

In the following sections, the findings will be discussed in depth and put in relation to a broader literature. As the results of the single studies were discussed in the empirical section (chapters 5-10), a focus of the following part is to conjoin the multiple results derived from the six studies (11.2 and 11.3). Thereby, I will identify limitations of the studies that might impact on the interpretation of results and suggest future approaches to overcome these. The subsequent section (11.4) offers a re-evaluation of theoretical and methodological aspects of the thesis based on the empirical findings. In the last part of the discussion (11.5), I will turn to future directions regarding questions that remained unanswered by the studies presented but are equally important for an across-the-board understanding of aberrant decision-making in addiction and addiction-like disorders.

11.2 Conjoint discussion of patient studies 1 - 2: transdiagnostic similarities and differences between alcohol addiction and binge eating disorder

Common and differential alterations in alcohol addiction and binge eating disorder. The translation of the addiction model to binge- and overeating has become increasingly popular over the last decade (Volkow & Wise, 2005; Smith & Robbins, 2013; Volkow *et al.*, 2013). So far, evidence of proposed parallels mostly stems from the animal literature. For example, rodent models of binge- or overeating showed interesting parallels to rodent models of substance addiction: sugar "bingeing" rats became insensitive to warning cues for punishments associated with further sugar intake (Epstein & Shaham, 2010; Johnson & Kenny, 2010). Other indications originate from a clinically observed overlap between binge eating and substance addiction (Ziauddeen *et al.*, 2012), namely a loss of control over the consumption of a "substance". This idea is appealing from a dimensional psychiatry point of view, which tries to overcome diagnostic boundaries and define behaviorally and biologically more plausible categorizations (compare section 3.6). However, there are also warning gestures against a premature adoption of this so-called food addiction model (Epstein & Shaham, 2010; Ziauddeen *et al.*, 2012; Salamone & Correa, 2013). Many concerns revolve around the critique that the cognitive phenotype of binge eating is insufficiently refined. This led to a call for detailed cognitive neuroscientific profiling of the disorder by strictly applying cognitive neuroscientific theories of addiction when investigating binge eating disorder (Ziauddeen *et al.*, 2012).

Studies 1 and 2 aimed to contribute to this matter by translating findings in substance addiction (study 1) to binge eating disorder (study 2). Adopting a Computational Psychiatry approach enables to test mechanisms underlying behavioral deficits. Interestingly, we observed in both samples that substance addicts and binge eating patients display a reduced amount of correct choices in a dynamically changing environment. This is in concordance with the shared clinical feature of failure of behavioral adaptation despite negative consequences. Again in both samples, mPFC coding of abstract inference components, integrating *what might have happened*, was found diminished. These parts of the data speak in favor of important parallels between the disorders. However, these two studies are also an illustrative example for the advantages of the

modeling-based approach; beyond behavioral raw data impairments (less amount of correct choices) in both groups, computational modeling allowed for a more fine-grained look at the behavioral differences: whereas alcohol-dependent patients showed reduced updating of alternative choices after punishment, binge eating patients did use abstract inference on the alternative choice. Instead, another parameter of the model was found altered, binge eating patients were characterized by accentuated and disadvantageous exploration behavior paralleled by reduced neural exploration signals in aI/vlPFC. The behavioral parameter capturing exploration, on the other hand, was not found significantly altered in alcohol addiction. Thus, one may conclude that behavioral adaptation is impaired in both patient groups but that the underlying mechanisms leading to this impairment differ. These data also implicate that one needs to tread carefully when adopting addiction models to binge eating or even obesity (which is even much more multifaceted in origin; Ziauddeen *et al.*, 2012; Dietrich *et al.*, 2014). It is interesting that, similar to the findings in alcohol-dependent patients, abstract inference signals in the mPFC were reduced in binge eating patients. This was related to behavioral performance, in terms of correct choices and exploration. However, in binge eating patients, we did not find behavioral evidence of not using abstract inference when updating alternative choices. This latter finding invites to explore more thoroughly how disadvantageous exploration behavior, which was indeed significantly correlated with the neural mPFC signature, relates to the capacity to use abstract inference on the task structure.

Abstract inference and exploration: a role for uncertainty? Conjecturing about the interplay between the capacity to use abstract inference and exploration behavior, it is noticeable that as shown via a simulation (study 1), double-updating of chosen and unchosen values is most advantageous after a change in environmental conditions, or in other words in situations of uncertainty. This is in line with key assumptions of a model-based system, which is most efficient when only little evidence is available, and is thus thought to exert most influence in uncertain situations (Daw *et al.*, 2005; Keramati *et al.*, 2011). Balancing the trade-off between exploratory and exploitative decisions also depends on uncertainty, with exploration taking

place proportional to experienced uncertainty (Frank *et al.*, 2009). Put differently, both the allocation of control to a model-based mode, as well as the degree of exploration behavior, depend on how uncertain the agent estimates its environment. The behavioral findings in studies 1 and 2, namely reduced double-updating, a heightened tendency towards exploration, as well as less expression of neural exploration and abstract inference signals might thus both be grounded in an inaccurate estimation of environmental uncertainty. Hierarchical Bayesian modeling techniques are warranted to test and compare the role of uncertainty estimation (e.g. Mathys *et al.*, 2011; Mathys *et al.*, 2014). In accord with this hypothesis, Schwartenbeck and colleagues (2015) have recently proposed a theoretical account on how different disorders from the compulsive spectrum might be characterized and differentiated by differing certainty estimates during decision-making (Schwartenbeck *et al.*, 2015). The authors suggest that addictive behavior is indeed caused by low certainty combined with low hazard rates (i.e., representation of task contingencies). This could similarly account for decision-making alterations observed in alcohol addiction and binge eating disorder. It could however be different in other disorders suffering from loss of behavioral control, such as Obsessive Compulsive Disorder (OCD). For example, in OCD, hazard rate estimation might not be altered. In sum, these are promising venues to further dissociate common and distinct aspects along the axes of behavioral control and psychopathological syndromes.

On the neural level, connectivity analyses might offer insight in how mPFC abstract inference and aI/vlPFC exploration signaling interact. One interesting hypothesis to be investigated is that stronger coding of abstract inference prediction errors helps to manage the exploration-exploitation tradeoff by impact of the mPFC on the aI/vlPFC. Interestingly, in previous studies aI activation has been frequently implicated in uncertainty coding (Grinband *et al.*, 2006; Singer *et al.*, 2009; Menon & Uddin, 2010). Thus, a further interesting question, which could be tested with connectivity analyses, is whether disturbed uncertainty signals influence the relevance of mPFC abstract inference coding.

To synthesize the similar finding of reduced goal-directed signals at close-by mPFC coordinates in alcohol-dependent patients and binge eating patients, one might reason that reduced mPFC function constitutes the final neural pathway leading to loss of control over one's actions. The mPFC might thus be speculated as an important neural point of convergence with respect to loss over behavioral control for various psychiatric conditions. Indeed, also previously, mPFC malfunction was implicated in the failure of different domains of self-regulation, across different diagnoses, e.g., behavioral adaptation in juvenile ADHD (Hauser *et al.*, 2014a), but also in emotion regulation in posttraumatic stress disorder (Etkin & Wager, 2007). It is plausible that alterations in (anterior) mPFC function lead to loss over behavioral control given theories on hierarchical functional organization of the prefrontal cortex which suggest most abstract representations being located in anterior parts of the prefrontal cortex. These superordinate regions are thought to operate on hierarchically abstract representations and to control subordinate (more posterior) regions (Badre, 2008). Certainly, malfunction of these *meta-controllers* could underlie a range of behavioral control problems across different disorders. To test this speculation more directly, one would, in future patient studies, ideally employ a computational model which explicitly embodies such a hierarchical structure (Mathys *et al.*, 2011; Mathys *et al.*, 2014; for application Iglesias *et al.*, 2013). In this model, prediction errors are weighted by certainty from the next higher level, which nicely relates back to the speculation on the role of uncertainty for abstract inference as well as exploration behavior in compulsive disorders (see previous paragraph and Schwartenbeck *et al.*, 2015).

Finally, one might argue that unspecific mPFC alterations (e.g. reduction in gray matter density) might underlie the observed overlap in mPFC malfunction found in both disorders. Notably however, the reduced mPFC signal was only found for the complex abstract inference signal, but did not differ between groups for the simpler, model-free learning signatures. This speaks in favor of a specific impairment in abstract inference coding in both disorders.

Matching Strategy. In a nutshell, studies 1 and 2 support the view that although the behavioral output is similar in alcohol addiction and binge eating disorder (i.e., impaired behavioral control

and goal-directed behavior in terms of correct choices in a changing environment), while underlying behavioral mechanisms and neural correlates appear to differ at least partially. This calls for a direct comparison of both groups. An important caveat when inferring on similarities and differences between two patient groups is that patient samples of study 1 and 2 do not only differ in diagnoses but additionally regarding a number of other factors: gender (predominantly men in the alcohol-dependent group, predominantly women in the binge eating group), age, educational status and neurocognitive functioning. This was accounted for by including separate, respectively as-good-as-possible matched control samples. Future studies might consider matching patients directly to each other, which, at the first glance, allows for a more direct comparison between the groups. However, as the patient populations differ naturally regarding these factors, and thus the differences are inherent to the diagnostic group, this entails the risk of studying artificial samples, which are no longer representative for the disease. It might also even out factors which are indeed tightly interwoven with the diagnostic entity.

A similar approach has been suggested regarding performance differences in cognitive-neuroscience studies of psychiatric patients: *matching for performance* was proposed to help to identify true biological differences beyond differences in cognitive abilities (Callicott *et al.*, 2003; Tan *et al.*, 2006). The search for truly *biological* differences might somehow seem circular: again, this results in non-representative samples, as cognitive impairments constitute a characteristic of many psychiatric conditions. Further, it evades the question of how brain signals have generated behavior. This question is certainly in the center of interest as the field still lacks any established biological marker of different behavioral phenotypes. Nevertheless, such performance-based subgroup analyses appear as informative control analyses and are more feasible when relatively large samples sizes are available, most likely from multicenter studies or via data-sharing tools (i.e. comparing patient groups with and without behavioral impairment regarding their neural correlates). Importantly, this debate once again advocates the computational approach: due to the inference of individual parameters, which are then transferred to the statistical analysis of neural data, one adjusts the test of brain-behavior

relationships for inter-individual differences in the latent process studied. However, for potential future advances in psychiatric diagnostics a different strategy could be more successful: namely to address biological heterogeneity within diagnostic labels and to test whether such biologically defined subgroups map on behavioral and symptom dimensions of clinical relevance (Stephan *et al.*, 2015).

11.3 Conjoint discussion of studies 4-6: reduced model-based control in risk factors of addiction – only crossing roads lead to Rome?

Across the examined risk factors of addiction, namely impulsivity, positive family history and stress, we did not find evidence for an impairment of model-based behavioral control per se. This is in contrast to findings in addicted and other psychiatric patient samples (Sebold *et al.*, 2014; Voon *et al.*, 2015). Lacking evidence in adult offspring of alcohol-dependent patients might either indicate that impaired model-based control is not an endophenotype of addiction; however it is important to consider the possibility that power of the study was too low to detect the effect. Indeed the effect of reduced model-based behavior might be much more subtle in at-risk than in addicted populations themselves (for an illustration of the hypothesized expression of model-based behavior in risk factors, see Figure 3-2). The picture emerging for stress and impulsivity is more complex, and points towards an important role of moderating factors: high trait impulsivity per se did not have an impact on the behavioral level regarding model-based choices (while the model-free system was found to be accentuated), but reduced lateral prefrontal model-based signals were revealed (study 4). A behavioral effect of impulsivity on model-based behavior could however be detected in interaction with cognitive capacities – another putative risk factor for addiction (study 5). Results of study 6 point in a similar direction: a detrimental effect of acute stress on model-based behavior was also not found per se, replicating a previous study (Otto *et al.*, 2013b). Only participants with a high load of chronic stress showed slightly reduced model-based control. In resonance with this reasoning, an interaction effect of physiological stress reactivity and fluid cognitive capacities had been detected in a previous study (Otto *et al.*, 2013b).

Taken together, these findings speak in favor of a *multiple hits* account in the relationship between risk factors of addiction and model-based behavioral control: a reduction of model-based control requires the concerted action of multiple risk factors (e.g., acute and chronic stress, high trait impulsivity and low cognitive functioning) to impact behavioral control. It cannot be determined from the studies at hand how exactly these risk factors act together to deteriorate or protect the model-based system. It might be that vulnerability factors sum up or compensate for each other (Robbins & Everitt, 1999). Alternatively, this could involve interaction effects in the sense of: only if factor A is present, factor B can exert its detrimental effect on behavioral control. It is also not clear which risk factor indeed interacts with which other risk factor. To piece this together, large-scale data acquisition in a broad range of the population, assessing a set of well-defined risk factors and their impact on model-based decision-making within subjects (compare e.g. Schumann *et al.*, 2010) in combination with analysis methods like structural equation models are warranted. The latter offer a means to test causal assumptions on the interactive effects of various assessed risk factors and have proven fruitful in risk profiling alcohol misuse and eating disorders before (Williamson *et al.*, 1995; Hawkins *et al.*, 1997).

It catches the eye that for all investigated risk factors, alterations in fronto-striatal circuits have been described previously (Dalley *et al.*, 2011; Dias-Ferreira *et al.*, 2009; Ersche *et al.*, 2012a; Ersche *et al.*, 2013). These regions are also key for goal-directed or model-based control (compare section 3.1 and 3.4). Diminished behavioral control in impulsivity was suggested to be associated with reduced prefrontal top-down control over striatal regions (Dalley *et al.*, 2011), which tallies with the finding of study 4 that high impulsive individuals were characterized by reduced model-based signals in the lateral prefrontal cortex. Also for first degree relatives, fronto-striatal structural abnormalities have been reported (Ersche *et al.*, 2012a). Interestingly relatives share with their dependent family member hypertrophy in the putamen and structural decline in the posterior insula, but not the gray matter reductions in the prefrontal cortex (Ersche *et al.*, 2012a, Ersche *et al.*, 2013). In a rodent model for stress, chronic

stress and associated reduced goal-directed control were related to a structural reorganization of fronto-striatal circuits, more specifically to atrophy of the medial prefrontal cortex and hypertrophy of dorsolateral striatum (Dias-Ferreira *et al.*, 2009).

From the set of at-risk studies in this thesis, it cannot be derived whether it is shared or differential neural substrates that underlie the putatively interactive effect of multiple risk factors. Puzzling together the finding of reduced lateral prefrontal model-based signatures in impulsivity while model-based behavior was intact (study 4) with the findings in patients (study 1 and 2, albeit using a different task) where goal-directed signals were found reduced in the medial prefrontal cortex, one might speculate that it is the failure of medial prefrontal function which ultimately leads to a break-down of goal-directed behavioral control. This is in line with the view of the medial prefrontal cortex as an arbitration region allocating control to one or the other system (Lee *et al.*, 2014). Such an arbitrator is thought to parse the degree of control exerted by a model-free vs. a model-based system at one point in time. It is suggested that this switching between strategies is brought about according to how reliable the predictions of the respective system are. Dysfunction of such arbitration was indeed discussed to lead to addictive or compulsive behaviors (Lee *et al.*, 2014; Gruner *et al.*, 2015).

In this thesis, all studies regarding risk factors for addictive behavior were of cross-sectional design and purposefully included young to middle-old adults, specifically selected not to meet any indication of substance misuse. It has previously been argued that this sample selection strategy might result in participants with particularly protective traits (Volkow *et al.*, 2006). This even invites the speculation that intact model-based behavior is a resilience factor for the development of addiction.

In any case, the contribution of multiple hits, putatively reducing behavioral control, to a transition into substance abuse cannot be elucidated from these cross-sectional studies. A hunch that the degree of model-based control might indeed change over the course of addiction development is however provoked by the observation that it covaries with time of abstinence in alcohol addiction (Voon *et al.*, 2015). To answer this critical question, longitudinal designs are

warranted. Ideally, in a large-scale data approach one would overcome power issues when investigating a large amount of teenagers with and without familiar risk for addiction, phenotype them with respect to additional risk factors (e.g. impulsivity, cognition, and stress) and track both, behavioral control and addictive behavior, over the course of the development.

In conclusion, inspired by the endophenotype account (Ersche *et al.*, 2012a; Robbins *et al.*, 2012), different populations at risk for addiction were studied. We used the same sequential decision-making task which had been used in previous studies (Sebold *et al.*, 2014; Voon *et al.*, 2015) to identify that addicted patients showed less model-based control. According to the endophenotype account, one would expect similar or intermediate expression of the behavioral deficit observed in patients (Figure 3-2). Across the presented studies on risk factors of addiction, we did not find evidence for an impairment of model-based behavioral control per se. Thus, the risk factors approach applied here was informative in the sense that it revealed qualitatively different alterations, not supporting the hypothesis that a shift from model-based to model-free behavior can be seen as an endophenotype for addiction and addiction-like disorders. Other than experiments used in humans, animal models of vulnerability factors in behavioral control study the transition from goal-directed to habitual control in paradigms involving self-administration of drugs. Building on such animal studies, authors have argued that these transitions are influenced by Pavlovian and instrumental controllers (Everitt & Robbins, 2005; Everitt *et al.*, 2008). It is an exciting speculation warranting future investigation, if - rather than reduced model-based control – alterations in Pavlovian-instrumental transfer effects constitute an endophenotype for addiction (see also section 11.5.1).

11.4 Re-evaluation of methodological aspects

The next section critically revisits methodological considerations of section 2: namely the use of the experimental tasks, computational modeling as a tool for analyzing behavioral data, the application of a dimensional approach regarding populations included and the use of fMRI and EEG brain measures. Based on this re-evaluation, I will suggest extensions for future studies.

11.4.1 Reflection of tasks used to measure behavioral adaptation

Two different tasks, namely counterfactual decision-making (studies 1-3) and a sequential decision-making task (studies 4-6) were used. They have been suggested as the two main paradigms to measure model-free and model-based control (Doll *et al.*, 2012). Both these tasks have been repeatedly used in previous studies investigating different aspects of behavioral control, mostly in healthy individuals (Hampton *et al.*, 2006; Daw *et al.*, 2011; Wunderlich *et al.*, 2012b; Otto *et al.*, 2013b; Schlagenhauf *et al.*, 2013; Smittenaar *et al.*, 2013; Deserno *et al.*, 2015b).

Sequential decision-making. Sequential decision-making was used in patient studies before, revealing reduced model-based behavior in substance addicted and binge eating patients (Sebold *et al.*, 2014; Voon *et al.*, 2015). There is now ample evidence that performance in sequential decision-making depends on general cognitive capacity (Otto *et al.*, 2013a; Otto *et al.*, 2013b; Schad *et al.*, 2014; Otto *et al.*, 2015). It was shown that cognitive capacity interacts with risk factors for addiction to impair model-based behavioral control (compare study 5 and Otto et al., 2013b). It is important to note that in the study by Sebold and colleagues group differences in model-based control between patients and controls were abolished when adjusting for cognitive capacities (Sebold *et al.*, 2014). For patient and risk group studies, this poses the important question whether reductions in model-based control found in this task are a specific characteristic for the disease, or rather an epiphenomenon of a more general cognitive impairment (which for alcohol-dependence, is inherent in the disease). As applied in the studies presented in this thesis, it seems particularly important to include a battery of measures of general cognition, and to test for an influence of cognitive factors on the investigated effect, in order not to incorrectly assign variance to the behavioral control systems which is better explained by general cognition (Collins & Frank, 2012). Evaluating this by now frequently used task, it is of note that a very recent simulation study revealed that in this task, behavior which appears as model-based behavior can be produced by (sophisticated) model-free agents without involving any planning (Akam *et al.*, 2015). These very recent findings should be taken into

account in future studies.

In the variant of the sequential decision-making task used here, the model-based system uses knowledge on environmental structure explicitly instructed and internalized beforehand (e.g., during the training session). Thus, using an already learnt transition-matrix is only one part of the model-based system, as it does not involve the acquisition of the transition function. The latter would be required of the model-based system in realistic environments where a beforehand training or instruction will obviously not be available in many situations. Indeed, it was argued that the ability to form representations of the world might be one of the most important factors in recovery from addiction (Kurth-Nelson & Redish, 2012). In a study involving healthy individuals, Gläscher and colleagues indeed introduced a variant of sequential decision-making which tests the latent acquisition and subsequent use of the transition matrix more explicitly (Glascher *et al.*, 2010). However, given the neurocognitive effort the variant used in studies 4-6 already requires, it seems challenging to apply the task designed by Gläscher and colleagues to psychiatric patients. When planning neuroimaging studies involving patients, it is important to take this point into account (Price & Friston, 1999): if patients are not able to perform the task, one cannot conclusively interpret the associated neural signal as abnormal neural processing as it might either be cause or consequence of impaired behavioral performance.

Counterfactual decision-making. Another aspect of capturing the environmental structure can be investigated by the counterfactual decision-making task used in the patient studies (studies 1 and 2) and study 3. Indeed, when adjusting performance for general cognitive functioning as in study 1, all behavioral and neuroimaging results were confirmed. This, and the fact that most patients were explained by a simple learning model, indicates that the demand of "scanning patients with tasks they can perform" (Price & Friston, 1999, p. 102) was indeed met by using this task. While sequential decision-making captures one of the key characteristics of model-based learning – the forward use of the transition matrix – in a more straightforward manner, counterfactual learning is well-suited to tackle another feature of behavioral control which clearly goes beyond pure model-free stimulus-action decision-making: abstract or hypothetical

inference on the task structure. During the counterfactual decision-making task applied in studies 1-3, the hidden task structure was learnt on the fly using abstract inference (via updating chosen and unchosen values) while in the sequential decision-making task the transition model was known beforehand. Implementing abstract inference via fictive outcomes, and integrating it into one's actions is an important hallmark of adaptive behavior (e.g., Boorman et al., 2009; Boorman et al., 2011), as in everyday life an individual often has to come to a decision while lacking explicit experience in favor of or speaking against the options at hand. Conjunctive coding of multiple actions and their outcomes enhances the efficiency of reinforcement learning (Abe & Lee, 2011; Takahashi et al., 2013). Several studies have linked this type of inference regarding the task structure, including the simulation of hypothetical outcomes, to the model-based system (Lucantonio et al., 2012; Lucantonio et al. 2014; Hampton et al., 2006). Given evidence from memory research (e.g. Shohamy & Wagner, 2008), it is well possible that generalization effects (here, generalizing from one stimulus being good to the other being bad) may *not* arise from a full model-based system but rather represent a short-cut approximation to the environmental structure by representing the relationship between two options. Human beings indeed can learn and benefit from correlation structures between different outcomes (Wunderlich *et al.*, 2011). It is plausible that the ability to infer values from the value of another stimulus, thus, to generalize, and to implement statistical regularities of the environment in the decision-making process, is a prerequisite for building mental models of the environment and thus plays an important role in model-based behavior. Hence, counterfactual decision-making captures a related but certainly more basic mechanism that might contribute to the previous findings of disturbed model-based control during sequential decision-making in various patient groups (Sebold *et al.*, 2014; Voon *et al.*, 2015).

Tasks to measure behavior control in addiction – what have we learnt so far? Regarding the original question of goal-directed vs. habitual control in addiction, these tasks, combined with the computational-modeling analysis, opened up new ways to dissect behavioral control in human beings. These tasks enable to test the hypothesis of a shift in the dominant mode

dimensionally. With former tasks, e.g. instructed devaluation (de Wit *et al.*, 2009; de Wit *et al.*, 2012a; de Wit *et al.*, 2012b; Sjoerds *et al.*, 2013), differences in behavioral control are more opaque as a reduction in goal-directed control implicates a rise in habitual control and vice versa. Given the (mostly very recent) studies using sequential decision-making or counterfactual decision-making, including the findings of this thesis, one résumé could be that there is accumulating evidence for reduced goal-directed or model-based behavior in addiction and addiction-like disorders like binge eating. This might be due to an impairment to use abstract inference on unchosen choice options or to biases in the exploration-exploitation tradeoff (Sebold *et al.*, 2014; Morris *et al.*, 2015; Voon *et al.*, 2015, and study 1 and 2 of this thesis). Interestingly and contrary to the initial hypothesis, these studies provide no evidence for an enhancement of the model-free or habitual system in addiction. The original formulation of the hypothesis is based on observations in animals which transit into escalating habitual drug-intake over the course of addiction development and in particular relapse to drug-intake (Everitt & Robbins, 2005). Thus, habitual behavior in this case is tightly coupled and assessed via drug intake, which is different from the experimental paradigms described here using monetary incentives as reinforcers. In a theoretical account, Simon and Daw propose a modification of these theories, promoting a more integrated view of the habitual, model-free and the goal-directed, model-based system (Simon & Daw, 2012). The authors argue that laborious (compare equation 5) model-based computations partially draw on cached values derived from the model-free system. Such a combination of the two modes of behavioral control would indeed be able to plan sequential action trajectories, but these would be targeted at reaching states with high values as derived from model-free valuation (Simon & Daw, 2012). One crucial point about these model-free values is that they are thought to be *inflated* as an effect of drug abuse, e.g., by uncontrolled, excessive effects of the abused substance on dopaminergically modulated prediction error signals (Redish, 2004; Redish *et al.*, 2008; Simon & Daw, 2012). Behavior resulting from such combinations of model-free and model-based valuation would bias the agent towards actions that lead to the high model-free values. An alternative model formalizing

combinations of model-free and model-based valuation as described is Sutton's Dyna-Q architecture (Sutton, 1990). The application of such an integrated algorithm appears as a highly promising next step for research on the model-free vs. model-based dichotomy in addiction.

11.4.2 Computational Psychiatry as a tool for refining neurocognitive disease phenotypes

Refining clinical phenotypes. The computational approach as applied here proved to foster a fine-grained understanding of the clinical phenotypes investigated (Montague *et al.*, 2012). For example, in study 1, we have shown that simpler learning mechanisms and the neural representation of model-free prediction errors are upheld in individuals suffering from alcohol addiction. Instead, they are impaired in one specific feature of behavioral control, namely updating alternative options after punishment. This goes beyond very general explanations, e.g. when defining addiction globally as a "disease of learning" (Hyman, 2005, p. 1414) or "a pathology of motivation and choice" (Kalivas & Volkow, 2005, p. 1403). Thus, it at least partially answers the question what patients can and cannot do (Price & Friston, 1999). A more fine-grained view on psychopathology might prove advantageous for developing specifically tailored psychopharmacological and psychotherapeutic treatment strategies. Also in light of the common stigmatization of drug-addicted individuals (Corrigan *et al.*, 2009), a more differentiated view may be beneficial.

Defining shared and differential pathways of disorders. Relatedly, the computational approach promoted the understanding of shared and differential pathways of the two patient populations, both characterized by loss of control. Regarding the two patient groups included in this thesis (study 1 and 2), behavioral deficits were similar at first glance (reduced correct responses in the counterfactual decision-making task). However, when looking at the modeling parameters, this could be pinned down to different underlying processes (compare section 11.2). This is particularly important, as binge eating disorder is a newly defined disease in DSM-5, which is understudied to date and still warrants thorough phenotyping (Ziauddeen *et al.*, 2012). Similarly, for risk factors of addiction such as impulsivity, alterations in behavioral or cognitive processes similar or intermediate to those observed in patients have been hypothesized

(Robbins *et al.*, 2012). By using the computational approach, alterations observed in risk factors could be more specifically defined as not quantitatively but rather qualitatively different from addicted populations (i.e., study 4 revealed alterations in the model-free system in impulsivity in contrast to the model-based system in addiction; Sebold *et al.,* 2014, see also section 11.3).

Analyzing the dynamic nature of behavioral adaptation. Behavioral adaptation is by definition a dynamic process over time. In the studies at hand, computational approaches enabled to mirror this process on a trial-by-trial basis, on the behavioral and neural level (studies 1- 4). It is to be noted that models applied in these studies might not fully reflect the dynamics inherent in both experimental tasks: as for studies 1-3, the applied models assume that double-updating is stable over the course of the experiment, and in studies 4-6 the models implement a stable use of the transition function. Other computational models might mirror the dynamics of learning more explicitly (compare Behrens *et al.*, 2007; Mathys *et al.*, 2011; Mathys *et al.*, 2014).

Classification and clustering as promising future avenues. However, despite these advantages, the current work has not made full use of the opportunities the computational psychiatry approach offers. As we have seen, observed behavior is distilled to specific components expressed by parameters for each subject. These would provide an optimal basis for machine-learning-based classification or clustering methods which aim at identifying subgroups and clinically significant conditions (Brodersen *et al.*, 2014; Stephan *et al.*, 2015; Wiecki *et al.*, 2015). For example, by feeding parameters derived from a model of reaction time in choice tasks (*Drift Diffusion* Model, Ratcliff & Rouder, 1998), combined with EEG data, into a classifier, Wiecki and colleagues could classify Parkinson patients according to whether they were on vs. off deep brain stimulation (Wiecki *et al.*, 2015). This approach could also be adopted for the studies at hand (e.g. studies 1 and 2): one might use the behavioral parameters derived from the modeling analysis as well as the associated fMRI signals to classify all participants of study 1 and 2. It might depend on the included parameters whether this would indeed reveal three subgroups (as suggested by the DSM-5 based sample selection criteria: alcohol-addicted, binge eating and healthy participants), or end up in one patient (independent of diagnosis) vs. control group pattern. The latter pattern

might arise from a classification on the basis of blunted mPFC abstract inference signatures and further support a transdiagnostic view. It would also be interesting to use the parameters gained in study 1 and 2 in order to cluster subgroups, agnostic to the actual diagnostic status. In a second step, these computationally informed subgroups can be related to clinical measures and outcome. For instance, in study 1, we found that the neural abstract inference signature in mPFC was correlated with compulsive substance consumption in alcohol-dependent patients. Based on these findings, one could ask, if clusters characterized by reduced mPFC signatures (probably containing participants of both patient groups) specifically relate to the clinical characteristic of compulsive habits, independent of diagnosis and substance (food vs. alcohol). Based on these mechanistically informed clusters, differentiated treatment strategies could be developed targeting at a subgroup's specific deficit: for example different medication, psychotherapeutic or psychosocial intervention options might be beneficial for different subgroups. As long as these subgroups are not known, clinicians are forced to rely on a trial-and-error principle when applying different treatment options. This touches upon the important question of *differential indication* in the treatment of psychiatric disorders (Grawe *et al.*, 1990; Roth & Fonagy, 2013). In sum, this approach could open up an exciting avenue towards describing, assessing, investigating and treating psychiatric diseases on the basis of measurable, reliable computational phenotypes by overcoming symptom-based classification (Brodersen *et al.*, 2014b; Stephan *et al.*, 2015; Wiecki *et al.*, 2015).

11.4.3 Benefit of the dimensional approach

For the patient studies in this thesis, diagnostic labels were used to recruit samples, test hypotheses and interpret results. This is in contrast to the approach described in the last paragraph, which might help to overcome the drawbacks of the current diagnosis systems (compare section 3.6) by defining new computationally grounded and mechanistically informed phenotypes. It is a legitimate objection to ask whether a dimensional approach based on diagnostic categories is truly dimensional (Wiecki *et al.*, 2015). For future studies, it thus seems

promising to make use of computational clustering and classification techniques to overcome this bias of symptom-based classification.

One might also consider rethinking the definition of the risk factors applied here. For positive family history, it has proven a powerful design to include patients and their siblings in one study, such that genetic and environmental factors are mostly similar as compared to the control group (Ersche *et al.*, 2012a). The studies at hand included adult participants explicitly screened not to meet any abnormalities regarding substance use. It might be that these participants are particularly resilient not to develop an addictive disorder. In order to dissect risk- from resilience factors, it could be fruitful to study risk populations at an earlier stage of development, e.g. adolecents.

Regarding impulsivity, the trait has been defined as a multidimensional construct (Dalley *et al.*, 2011; Wilbertz *et al.*, 2014). Based on rodent studies, it was suggested that specifically one subdomain, namely *waiting impulsivity* (i.e., premature responding) is a predisposing factor for addictive behavior (Belin *et al.*, 2008). A task measuring this subdomain of impulsivity has recently been translated to test humans and has revealed higher waiting impulsivity in addicted individuals (Voon *et al.*, 2014). Thus, for further studies, it seems a promising option, rather than sampling individuals based on the self-report questionnaire, to use performance in this specific task as definition criterion for risk populations.

11.4.4 Using multiple methods to study neural signatures of behavioral control

Behavioral control as conceptualized in this thesis has more frequently been investigated using (modeling-based) fMRI than any other method to measure brain function. Also in the studies at hand, fMRI was used most frequently when measuring neural correlates of habitual and goal-directed behavior (study 1, 2, 4). This is motivated by the fact that deeper brain regions like the basal ganglia (which cannot be directly measured via EEG) are key for behavioral adaptation (Hikosaka & Isoda, 2010). However, those studies in this thesis which probed neural group differences using fMRI could show that these were present in medial and lateral prefrontal structures (compare study 1, 2 and 4). The latter regions are principally also accessible via EEG

measurements. In line, in study 3 we used a modeling-based, trial-by-trial analysis to demonstrate ERP correlates of implementing abstract inference in counterfactual decision-making. This might prove valuable in further studies with special populations like children or patients tackling similar questions. It might open up the opportunity of large-scale data acquisition as suggested at various points in this discussion (compare section 11.2 and 11.3), because EEG is by far more economic and feasible than fMRI. For example, on-site data acquisition directly measured at in- or out-patient departments are easy to imagine when using EEG. In this vein, the methodological approach should be broadened towards exploiting the temporal resolution more clearly by time-frequency analyses. Time-frequency components have been analyzed using a modeling-based approach before; e.g., frontal theta oscillations were shown to be key correlates of cognitive control and prediction error processing (Cavanagh *et al.*, 2010; Cavanagh & Frank, 2014). This might be an exciting target for future studies with similar research questions, asking whether frontal theta signals covary with a shift from model-free, habitual to model-based, goal-directed control.

11.5 Limitations, further considerations, future directions

11.5.1 The role of the Pavlovian valuation system

Regarding decision-making processes in healthy individuals and patients, the current thesis focused on the dichotomy of habitual and goal-directed instrumental control. In reality, to come to a decision, we undergo many sub-processes, which can be broken down into several basic computations where habitual or goal-directed learning and control are only pieces of the puzzle (Rangel *et al.*, 2008). For the purpose of this thesis, this view on decision-making was intentionally focused on one specific aspect, which results in oversimplification. To illustrate this, imagine an addicted individual faces a situation where he or she is running out of substance. Conceivably, the individual will seem quite flexible and engage in planning when it comes to foraging the craved substance (e.g. show sequences of drug-seeking behavior with delayed or in the absence of the rewarding outcome). To explain this, one might need to consider a third valuation system, the Pavlovian system (Dayan *et al.*, 2006; Rangel *et al.*, 2008). Based on

rodent studies, it has been suggested that drug seeking behavior is controlled by Pavlovian stimuli (Belin *et al.*, 2013). These conditioned reinforcers have themselves taken on rewarding features because they had been repeatedly paired with the drug (s*econd-order conditioning*). They can thus influence instrumental behavior and guide the addicted individual during drug-seeking (Belin *et al.*, 2013). This can result in a vicious circle of self-reinforcing loops of choice and reward (Hogarth *et al.*, 2007; Huys *et al.*, 2014).

Relatedly, when studying goal-directed behavior (e.g. in paradigms using money as reinforcers) in addicted versus healthy populations, one has to take into account that it might not only be the way to achieve goals that differs between groups. Also the goal values per se, the expected value of a prospective outcome (O'Doherty, 2014), might be altered. Indeed, the importance of subjective values is emphasized in prominent addiction theories (Everitt *et al.*, 2008): evidence from reward anticipation paradigms points towards reduced predictive value coding of monetary outcome in alcohol-dependent patients (Wrase *et al.*, 2007; Beck *et al.*, 2009), which is in line with the hypothesis of a *hijacked* reward system in addiction (Hyman, 2005).

Already in healthy individuals, refining the computational and neural architecture of Pavlovian systems has turned out to be difficult so far (Rangel et al., 2008; O'Doherty, 2014). The interaction of Pavlovian valuation systems with instrumental valuation systems is at the beginning to be understood (Talmi *et al.*, 2008). Little is known about how the Pavlovian system acts in concert with the habitual vs. the goal-directed control modes; exciting first results points toward the Pavlovian system assisting the goal-directed system by yielding approximations for the laborious computations required for goal-directed control (Huys *et al.*, 2012). There is good reason to assume that Pavlovian-to-instrumental transfers might be of pivotal importance for addiction: in a very recent study in alcohol-dependent patients, it was shown that BOLD signals reflecting the influence of Pavlovian cues on instrumental behavior predict alcohol intake and relapse in alcohol dependence (Garbusow *et al.*, 2015). It remains to be elucidated whether Pavlovian-instrumental transfers also play a role in addiction vulnerability.

11.5.2 Neuro-chemical considerations

To date, it is almost undisputable that the neurotransmitter dopamine (DA) plays an essential role for learning and behavioral adaptation. Most prominently, dopamine firing was shown to scale with reward prediction errors and to function as teaching signals (Montague *et al.*, 1996; Schultz *et al.*, 1997). In fact, the hemodynamic correlates of prediction errors measured via fMRI (compare studies 1, 2 and 4) have been theoretically linked to dopamine prediction error signaling (O'Doherty *et al.*, 2003; O'Doherty *et al.*, 2004; Knutson & Gibbs, 2007) although one should keep in mind that the BOLD signal is too unspecific as to reflect the particular event of dopamine release. Importantly, phasic dopamine signals are thought to represent model-free learning signals as the firing of dopamine neurons can be captured by model-free RL-algorithms (see equation 5). The role of DA in model-based learning is less understood, but a recent study has stressed the modulatory role of dopamine in the balance between model-free and model-based control (Wunderlich *et al.*, 2012b). In support, higher levels of presynaptic ventral striatal dopamine were related to a higher degree of model-based control, enhanced coding of model-based signatures in the lateral prefrontal cortex and reduced representation of model-free prediction errors in ventral striatum (Deserno *et al.*, 2015b).

The development of addiction and substance abuse is thought to involve the usurpation of DA learning signals (Redish, 2004; Dayan, 2009a; Volkow *et al.*, 2009; Huys *et al.*, 2014). Addiction has been shown to be characterized by reduced D2 receptor availability, which also relates to drug craving and relapse (Heinz *et al.*, 2004; Volkow *et al.*, 2004; Heinz *et al.*, 2005; Volkow *et al.*, 2009). Intriguingly from a transdiagnostic perspective, also compulsively eating, obese rats were shown to be characterized by downregulated D2 striatal receptors (Johnson & Kenny, 2010). It is further noticeable that alterations in the DA system have not only been described for the disorders discussed, but indeed for all risk factors investigated in this thesis: with respect to impulsivity, Dalley and colleagues showed a reduction of D2 receptor availability in trait-impulsive drug-naïve rats, which was associated with subsequent escalation of self-administered cocaine-intake after withdrawal (Dalley *et al.*, 2007). Likewise for human

participants, it was demonstrated that a higher degree of impulsivity was correlated with less midbrain D2 auto-receptor binding and greater amphetamine-induced DA release in the striatum (Buckholtz *et al.*, 2010). Regarding stress, rhesus monkeys which had undergone a high load of social stress during adolescence were characterized by lower levels of DA D2 receptor binding in the striatum and showed pronounced self-administration of cocaine as compared to non-stressed animals (Morgan *et al.*, 2002). Young adults with a multigenerational family history of substance use disorders were not significantly different from drug-naïve and drug-experienced control groups in terms of baseline D2 receptor availability (Casey *et al.*, 2014). Interestingly, after amphetamine intake, adults with a positive family history exhibited reduced dopamine responses especially in right ventral striatum, when compared to drug-naïve and drug-experienced controls (Casey *et al.*, 2014).

Based on these previous studies, the findings of the studies at hand could be interpreted in light of the *dynamic dopamine model of cognitive reinforcement learning* (Frank *et al.*, 2004). In this model, two main projection pathways – a direct and indirect pathway – are suggested to convey information from the striatum, via the thalamus to the cortex. The direct pathway signals the execution of a response from the striatum through the thalamus to the cortex, whereas signals in the indirect pathway suppress competing responses indirectly via the Globus Pallidum. D1 receptors are prominent in the direct pathway, while D2 receptors predominate the indirect pathway. After positive reinforcement, dopamine bursts activate the direct pathway, and deactivate the indirect pathway. This prompts learning and the execution (or repetition) of previously reinforced actions. Contrary, dopaminergic dips after punishment cause the opposite effect, such that punished actions are suppressed in the future (Frank *et al.*, 2004). Given the above described previous findings of reduced D2 receptor availability in impulsivity, after chronic stress, in addiction and in a rodent model of compulsive eating, but not in individuals with family history of addiction, one might speculate that reduced D2 receptors hamper the indirect pathway which is responsible for inhibiting disadvantageous choices. This impaired inhibitory process might potentially account for disturbances of the goal-directed system,

especially when punishing events occur, as well as perseverative behavior in alcohol addiction (study 1), and enhanced switching and exploration despite negative consequences in binge eating disorder (study 2). Likewise, it might play a role for the observed reduction in goal-directed or model-based behavior seen in the interaction of acute and chronic stress (study 6) and diminished neural model-based signals in impulsivity (study 5). This hypothesis could be tested by manipulating the pathways by dopaminergic agonists and antagonists in healthy individuals: one would expect that the pharmacological intervention causes similar alterations in behavior and neural signaling as observed in the studies at hand.

Besides dopamine, a recent study emphasizes the role of serotonin for behavioral adaptation after punishment in a reversal learning task (den Ouden *et al.*, 2013). The role of serotonin in addiction is less studied, however low serotonin levels have been linked to aggressive behavior (Heinz *et al.*, 2011) and impulsivity (Cools *et al.*, 2005). This indicates that some of the findings observed in this study might be subject to serotonergic influences. However, causal insight into molecular mechanisms promoting drug-specific plasticity is so far limited and requires translational research approaches (Luthi & Luscher, 2014).

11.6 Conclusions

The aim of this thesis was to elucidate mechanisms of behavioral control in alcohol addiction, binge eating disorder and risk factors of addiction. Using computational modeling, the presented studies could dissect specific behavioral control deficits in nosologically different psychiatric diseases. This constitutes a step towards a more mechanistic understanding of these disorders. On the neural level, the medial prefrontal cortex appears as an important convergence zone monitoring goal-directed behavior. The findings suggest that malfunction of mPFC is a final pathway involving nosologically different psychiatric conditions characterized by loss of control over one's behavior. Further, I could demonstrate that behavioral control alterations in risk factors of addiction are qualitatively different to patients. The data speak in favor of a *multiple hits* account of different risk factors interacting or adding up to impede goal-directed control. This calls for applying the computational approach in future longitudinal studies before the

onset of psychiatric diseases. Further, the application of hierarchical computational models in combination with modeling-informed fMRI and EEG as well as machine learning and clustering methods appear as a promising next step. This could indeed enable identification of biologically informed subgroups and guide new treatment and prevention developments. In sum, this thesis provides answers to the question of why some people repeat behaviors, despite often times devastating consequences.

12 Summary

Why do we repeat behaviors we know are bad for us?

Influential theories of behavioral control distinguish two main systems to guide behavior, a reflexive, habitual system on the one hand, and a reflective, goal-directed system on the other hand (Balleine & Dickinson, 1998; Daw *et al.*, 2005; Dolan & Dayan, 2013). The habitual system takes actions by drawing on values 'stamped in' by past reinforcement, even though these values might be disparate from the current value of the action's outcome. It is characterized by a high degree of automaticity and computational efficiency, which comes at the cost of reduced behavioral flexibility. In contrast, the goal-directed system plans actions in a forward manner, mentally tracing down possible consequences of actions. This is computationally costly but has the advantage of flexibility when adapting behavior as soon as environmental conditions have changed. In healthy individuals, decision-making is influenced by both systems of behavioral control (Daw *et al.*, 2011); individuals are thought to balance the trade-off between the faster, automated 'habit' system, and the slower but more flexible system to come to optimal decisions (Keramati *et al.*, 2011). Computational accounts have amended this dual-systems-theory (Dolan & Dayan, 2013), by formalizing habitual control in model-free and goal-directed control in model-based algorithms. In the last years, this computational formulation has substantially advanced cognitive neuroscientific research on behavioral control. Neuroimaging studies building on this approach have identified signatures of model-free and model-based control in medial prefrontal cortex as well as ventral striatum (Daw *et al.*, 2011).

Addiction is a psychiatric condition which is marked by adverse decisions: by definition, addicted individuals continue substance consumption in the face of negative consequences (American Psychiatric Association, 2013). To explain this disadvantageous kind of decision-making, theories of addiction build on the notion of the two decision-making systems described above. Addiction is conceptualized as a disrupted balance between goal-directed and habitual control (Everitt & Robbins, 2005; Redish *et al.*, 2008; Dayan, 2009a). An overreliance on the habitual system along with an impairment of goal-directed control was hypothesized to lead to

addicted individuals' repeatedly choosing one action which has been reinforced in the past but is harmful in the present.

The presented thesis employs this view on addiction and tests it in a population of individuals suffering from alcohol dependence. Furthermore, by adopting a dimensional approach (Robbins *et al.*, 2012), this thesis extends the hypothesis, (1) towards another psychiatric condition characterized by the loss of control over behavior, namely binge eating disorder; (2) towards risk factors of addictive disorders, namely impulsivity, positive family history, low cognitive functioning and stress. This is studied by leveraging computational modeling techniques as well as functional Magnetic Resonance Imaging (fMRI) and Electroencephalography (EEG) to test inter-individual differences in behavioral and neural signatures of behavioral control.

Study 1 addresses the research question whether individuals suffering from alcohol addiction are impaired in an important function of the goal-directed system, namely the ability to update unchosen choice values ("what might have happened"). During fMRI, a counterfactual decision-making task with reversals of reward contingencies was employed. Out of two options, subjects had to find the stimulus with the higher reward probability (of 80%) in order to gain monetary reward. Crucially, this task implemented a higher-order structure: whenever one of the stimuli was the good one, the other one would be the worse (Hampton *et al.*, 2006). Even though feedback for the unchosen option is not delivered, the agent can infer from feedback gained for the chosen option what would have happened if he had chosen the other stimulus. By the use of computational modeling, it can be dissected how much individuals use this information derived from abstract inference. Comparing alcohol-dependent patients with healthy controls, choice behavior in the patient group was best explained by a more simplistic Reinforcement Learning algorithm, which neglects alternative choice options. This was due to reduced updating of the alternative options after punishment. Decision-making in healthy volunteers was best explained by an algorithm which quantifies inter-individual differences in the extent to which individuals update alternative choice options. Modeling-based fMRI analysis

revealed that alcohol-dependent patients showed reduced representation of abstract inference signals in the medial prefrontal cortex – a key region implied in healthy and disrupted model-based control (Daw *et al.*, 2011). No difference was found regarding neural signals of model-free decision-making.

Study 2 adopted the same study design and experiment as in study 1 and translated it to patients suffering from binge eating disorder. As alcohol addiction, binge eating disorder is a psychiatric condition characterized by loss of control over one's behavior; during recurrent episodes of binge eating subjects experience the feeling that they cannot stop eating, despite negative consequences like a high risk for obesity. binge eating disorder is a newly recognized diagnosis first included in DSM-5 (American Psychiatric Association, 2013). Sometimes the disease is referred to as *food addiction* (Smith & Robbins, 2013; Volkow *et al.*, 2013), but neurocognitive research on the disorder has been relatively scarce rendering the clinical phenotype poorly defined (Ziauddeen *et al.*, 2012). Using counterfactual decision-making, we revealed that binge eating patients showed impaired behavioral adaptation in a dynamically changing environment. Computational Modeling showed that, other than the alcohol-dependent patients, individuals suffering from binge eating disorder were not impaired in the ability to update alternative choices. However, binge eating patients showed a disadvantageous bias towards explorative decisions. Modeling-based fMRI analysis revealed reduced abstract inference signals in medial prefrontal cortex, interestingly, at close-by coordinates as found for alcohol-dependent patients in study 1. This neural signal was indeed positively correlated with behavioral adaptation and negatively related to exploration behavior. No difference between patients and controls was found with respect to model-free learning signals. Moreover, we identified reduced anterior insula / ventrolateral prefrontal cortex signals as a neural correlate for the enhanced tendency to explore in binge eating disorder.

The aim of *study 3* was to identify neural abstract inference signals, which had been found reduced in patients in study 1 and 2, using EEG. Adopting a modeling-based EEG analysis, it was asked whether an event-related potential associated with prediction error processing, the

Feedback-Related Negativity (FRN), codes these abstract inference signals. Indeed, we could show that the FRN implements *what might have happened*. This finding complements theories about the role of the FRN in behavioral adaptation, which is controversially discussed in the literature. Moreover, this result opens up new ways to study research questions like in study 1 and 2 via EEG. As EEG is more feasible as compared to fMRI, this might facilitate studying large sample sizes in special populations, for example, psychiatric patients. Moreover, EEG is advantageous with respect to temporal resolution, which is relevant for studying learning processes.

Studies 4-6 probed behavioral control in risk factors for addiction using sequential decision-making (Daw *et al.*, 2011). Sequential decision-making enables to dissect goal-directed, model-based from habitual, model-free control strategies. It tests the use of a previously trained transition matrix while subjects are confronted with a sequential (i.e., involving multiple steps) decision-making problem. Using the transition matrix would correspond to a model-based strategy. Previous research had shown that subjects suffering from disorders of the impulsive-compulsive spectrum, including alcohol addiction and binge eating disorder, are impaired in model-based behavior (Sebold *et al.*, 2014; Voon *et al.*, 2015). Studies 4-6 asked whether a shift from model-based towards model-free control constitutes a vulnerability factor for addiction.

To this end, *study 4* examined healthy high vs. low impulsive individuals using a sequential decision-making task during fMRI. We did not find evidence for reduced model-based control, but observed a slight attenuation of the model-free system in high impulsive individuals. On the neural level, we revealed reduced model-based signatures in lateral prefrontal cortex.

Study 5 is a behavioral study, probing the balance of model-based and model-free control in healthy adult individuals with an alcohol-dependent father. No evidence for reduced model-based or enhanced model-free control could be detected. However, independent of family history of alcohol dependence, an interaction effect of the risk factors impulsivity and cognitive functioning on model-based control was revealed. In a re-analysis of the data from study 4, this interaction was replicated.

Study 6 asked whether individuals exposed to acute and chronic stressors shift from model-based towards model-free behavior during sequential decision-making. Whereas no effect of acute stress (as experimentally induced via public speaking, Kirschbaum *et al.*, 1993) per se could be revealed, physiological stress reactivity in terms of cortisol responding towards the acute stressor correlated negatively with the expression of model-based control. Subjective stress reactivity measured via ratings correlated positively with model-based behavior. Furthermore, we found evidence that the interaction of acute and chronic stress impairs model-based behavior: only in individuals with a high load of stressful life events, a reduction of model-based control in response to the acute stressor was observed.

The aim of this thesis was to elucidate shared and differential mechanisms of behavioral control in alcohol addiction, in binge eating disorder, a putatively addiction-like disorder, as well as in risk factors for addiction. Altogether, the presented findings indicate reduced goal-directed control in both, alcohol addiction and binge eating disorder, on the behavioral as well as the neural level. In both populations, model-free signals were found preserved. By the use of computational modeling techniques, partially differential pathways leading to this similar behavioral and neural alterations could be dissected: alcohol-dependent patients displayed a failure to integrate alternative choice options in their decisions after punishment, while patients suffering from binge eating were characterized by a disadvantageous bias towards explorative choices. The shared finding of reduced medial prefrontal cortex learning signals in both patient groups suggests this region as a transdiagnostic convergence point essential for monitoring behavioral control. Regarding risk factors of addiction, findings of study 4-6 were qualitatively different to observations in populations suffering from addiction or other disorders from the impulsive-compulsive spectrum (Sebold *et al.*, 2014; Voon *et al.*, 2015). The observed interaction effects between cognition and impulsivity, as well as acute and chronic stress, suggest a *multiple hits* account for vulnerability factors impeding model-based control not alone, but in interaction with each other.

Findings of the thesis call for longitudinal designs including large sample sizes and involving detailed phenotyping of potential risk factors. This could have the potential to replicate presented findings and to validate them with respect to their clinical relevance in prospective study designs. Further, the demonstration of abstract inference learning signals via EEG could prove fruitful for the purpose of gathering large sample sizes due to the feasibility of EEG. Stimulated by the findings at hand, tracking goal-directed behavioral control and its neural correlates from vulnerability over development and potential remission of addiction or addiction-like disease is an important target for future studies; this will eventually offer causal insight into the question of why people repeat certain behaviors, despite devastating consequences.

13 Zusammenfassung

Warum tun wir immer wieder Dinge, von denen wir wissen, dass sie schlecht für uns sind?

Theorien der Verhaltenskontrolle unterscheiden zwischen zwei Systemen, die Verhalten steuern: Es wird ein reflexives, habituelles System und ein reflektives, zielgerichtetes System beschrieben (Balleine & Dickinson, 1998; Daw *et al.*, 2005; Dolan & Dayan, 2013). Das habituelle System agiert retrospektiv, indem es auf Entscheidungswerte zurückgreift, die durch Belohnungen in der Vergangenheit erworben wurden. Dabei kann es sein, dass diese alten Entscheidungswerte nicht mehr dem Wert der aktuellen Konsequenzen einer Handlung entsprechen. Dementsprechend kann der Rückgriff auf diese Entscheidungswerte zu einer Handlung mit unerwünschten Folgen führen. Das habituelle System arbeitet in hohem Maße automatisch und ressourcensparend. Diese Effizienz geht jedoch mit geringerer Flexibilität einher, wenn Verhalten an Umweltbedingungen angepasst werden muss. Im Gegensatz zum habituellen System plant das zielgerichtete System vorausschauend, indem es mental die möglichen Konsequenzen verschiedener Handlungsoptionen simuliert. Dies beansprucht in hohem Maße kognitive Ressourcen. Es ermöglicht dem Individuum jedoch sich schnell und flexibel an veränderte Umweltbedingungen anzupassen. Entscheidungsverhalten ist bei gesunden Individuen durch beide Systeme geprägt (Daw *et al.*, 2011): Um zu optimalen Entscheidungen zu kommen, wird ein Gleichgewicht zwischen dem schnellen, automatischen habituellen und dem langsamen, aber flexibleren, zielgerichteten System gehalten. Diese 2-System-Theorie wurde in den letzten Jahren durch computationale Ansätze erweitert (Dolan & Dayan, 2013): Mit aus der Informatik und dem Maschinellen Lernen entlehnten Konzepten wurde habituelle Verhaltenskontrolle in sogenannten *modellfreien* Algorithmen und zielgerichtete Verhaltenskontrolle in *modellbasierten* Algorithmen beschrieben. Diese computationale Rekonzeptualisierung prägte die kognitiv-neurowissenschaftliche Forschung auf dem Gebiet der Verhaltenskontrolle in den letzten Jahren nachhaltig. In fMRT-Studien konnten modellfreie und modellbasierte neuronale Signale in medial- und lateral präfrontalen Kortizes und im ventralen Striatum identifiziert werden.

Sucht ist eine psychische Erkrankung, die durch nachteilige Entscheidungen gekennzeichnet ist: Süchtige Individuen konsumieren weiter, obwohl sie wissen, welche schwerwiegenden Konsequenzen für Gesundheit und Lebensführung der Substanzkonsum mit sich bringt (American Psychiatric Association, 2013). Moderne Suchttheorien greifen auf die oben beschriebene 2-System-Theorie zurück, um dieses nachteilige Entscheidungsverhalten zu erklären (Everitt & Robbins, 2005; Redish *et al.*, 2008; Dayan, 2009a): Es wird angenommen, dass Sucht auf ein gestörtes Gleichgewicht zwischen habituellen (modellfreien) und zielgerichteten (modellbasierten) Systemen zurückgeht: Da bei Sucht eine Dominanz des habituellen Systems vorliege, während der Einfluss des zielgerichteten Systems reduziert sei, wähle der Süchtige immer wieder Verhaltensweisen, die früher einmal belohnend waren, heute jedoch schädlich sind.

Die vorliegende Arbeit untersucht diese Hypothese in einer Population von alkoholabhängigen Individuen. Zudem wird diese Hypothese um eine dimensionale Perspektive (Robbins *et al.*, 2012) erweitert: 1) durch die Untersuchung einer weiteren Erkrankung, die durch den Verlust über die Verhaltenskontrolle gekennzeichnet ist, der Binge Eating Störung, und 2) durch Experimente, die verschiedenen Risikofaktoren für Sucht einschließen, nämlich Impulsivität, positiver Familienanamnese für Sucht, niedriges kognitives Funktionsniveau und Stress.

Es kommen dabei Techniken der computationalen Modellierung sowie funktionelle Magnetresonanztomographie (fMRT) und Elektroenzephalografie (EEG) zum Einsatz, um interindividuelle Unterschiede im Verhalten und den neuronalen Korrelaten von Verhaltenskontrolle zu untersuchen.

Studie 1 untersucht die Frage, ob alkoholabhängige Menschen in der Fähigkeit, aus alternativen Wahlmöglichkeiten zu lernen („was hätte sein können"), beeinträchtigt sind. Diese Fähigkeit fiktive Konsequenzen von nicht-gewählten Optionen in Entscheidungen einzubeziehen stellt eine wichtige Funktion des zielgerichteten Systems dar. Um dies zu untersuchen, wurde während der fMRT-Messung eine kontrafaktische Entscheidungsaufgabe, in der sich die Belohnungskontingenzen regelmäßig umkehren, genutzt. Um einen Geldgewinn zu erhalten,

sollten die ProbandInnen aus zwei Stimuli immer den mit der höheren Belohnungswahrscheinlichkeit (80%) auswählen. Dieser experimentellen Aufgabe ist eine höhere Struktur inhärent: Die höhere Belohnungswahrscheinlichkeit eines Stimulus ist dabei mit der niedrigeren Belohnungswahrscheinlichkeit bei einem anderen Stimulus verknüpft. (Hampton *et al.*, 2006). So könnte die Versuchsperson aus der Konsequenz der Wahl eines Stimulus (zum Beispiel Geldgewinn) auf die Belohnungswahrscheinlichkeit bei Wahl des anderen (nicht gewählten) Stimulus schließen. Mit der Hilfe von computationalen Modellen kann bestimmt werden, wie sehr die Versuchspersonen diese Informationen („was hätte sein können") nutzen. So konnte gezeigt werden, dass das Verhalten der PatientInnengruppe am besten durch ein einfacheres Lernmodell erklärt wurde, in dem die alternative Wahlmöglichkeit nicht berücksichtig wird. Dies war zurückzuführen auf ein reduziertes Beachten der alternativen Optionen nach Bestrafung. Die gesunden KontrollprobandInnen hingegen bezogen fiktive Information darüber was hätte sein können in ihre Entscheidungen mit ein. Die fMRT-Analyse konnte zeigen, dass alkoholabhängige PatientInnen die zielgerichteten Lernsignale, die Rückschluss auf die alternative Option beinhalten, verringert neuronal repräsentieren. Dieses Defizit wurde im medialen präfrontalen Kortex gefunden, der für modellbasierte Verhaltenskontrolle essentiell ist (Daw *et al.*, 2011).

In *Studie 2* wurde dasselbe Experiment wie in Studie 1 genutzt, um PatientInnen mit Binge Eating Störung zu untersuchen. Diese Störung ist, wie Substanzabhängigkeit, gekennzeichnet vom Verlust der Kontrolle über eigene Verhaltensweisen. Beispielsweise können die PatientInnen während der wiederkehrenden Essanfälle nicht mit dem Essen aufhören. Die Binge Eating Störung wurde erst vor kurzem als eigenständige psychische Störung anerkannt und in die neueste Auflage des Diagnostisch Statistischen Manuals aufgenommen (American Psychiatric Association, 2013). Die Binge Eating Störung wird manchmal als *Esssucht* bezeichnet (Smith & Robbins, 2013; Volkow *et al.*, 2013). Neuro-kognitive Forschung gibt es bislang jedoch wenig. Der klinische Phänotyp wird deshalb als nicht ausreichend definiert kritisiert (Ziauddeen *et al.*, 2012). Um hierzu einen Beitrag zu leisten, wurde in Studie 2 die Verhaltensanpassung bei

PatientInnen mit Binge Eating Störung mit dem oben beschriebenen kontrafaktischen Entscheidungsexperiment untersucht. Dies zeigte verschlechterte flexible Verhaltensanpassung bei Binge Eating PatientInnen. In einem weiteren Analyseschritt wurden Parameter aus der computationalen Modellierung zwischen PatientInnen und KontrollprobandInnen verglichen. Dabei konnte demonstriert werden, dass das behaviorale Defizit, anders als bei den alkoholabhängigen PatientInnen aus Studie 1, nicht darauf zurückzuführen war, dass Binge Eating PatientInnen alternative Wahlmöglichkeiten vernachlässigten. Binge Eating PatientInnen waren hingegen durch nachteilige explorative Verhaltenstendenzen gekennzeichnet. Die modellierungsbasierte fMRT-Auswertung zeigte reduzierte zielgerichtete Lernsignale im medialen präfrontalen Kortex, und das in einer ähnlichen Region wie in Studie 1 bei den alkoholabhängigen PatientInnen. Dieses neuronale Korrelat korrelierte positiv mit erfolgreicher Verhaltensanpassung und negativ mit Explorationsverhalten. Binge Eating PatientInnen zeigten ein reduziertes neuronales Korrelat von explorativen Entscheidungen in der anterioren Insel und dem ventrolateralen präfrontalen Kortex. Es konnte kein Unterschied zwischen PatientInnen und Kontrollgruppe hinsichtlich einfacher, modellfreier Lernsignale festgestellt werden.

Ziel von *Studie 3* war, die bei PatientInnen reduzierten zielgerichteten Lernsignale (Studie 1 und 2) mit der Methode der EEG nachzuweisen. In einer modellierungsbasierten EEG Analyse konnte nachgewiesen werden, dass ein ereigniskorreliertes Potenzial, die sogenannte Feedback-Related Negativity (FRN), Lernsignale mit der Information darüber *was hätte sein können* kodiert. Diese Ergebnisse erweitern Theorien über die Rolle der FRN für Verhaltensanpassung, die in der Literatur kontrovers diskutiert wird. Darüber hinaus helfen diese Ergebnisse, Forschungsfragen wie in Studie 1 und 2 in großen Fallzahlen von speziellen Populationen (z.B. psychiatrischen PatientInnen) zu untersuchen, da EEG im Vergleich zu fMRT leichter durchführbar und flexibler einsetzbar ist.

In *Studie 4-6* wurde Verhaltenskontrolle bei Personen mit Risikofaktoren für Suchtentwicklung untersucht. Dabei kam eine sequentielle Entscheidungsaufgabe zum Einsatz (Daw *et al.*, 2011;

Sebold *et al.*, 2014). Mit der Hilfe von sequentiellen Entscheidungsaufgaben können modellfreie von modellbasierten Verhaltensweisen unterschieden werden: es wird überprüft, ob ProbandInnen eine vorab gelernte Transitionsregel nutzen, wenn sie mit mehrschrittigen Entscheidungsproblemen konfrontiert sind. Das Anwenden dieser Regel entspricht dabei einer modellbasierten Strategie. In vorangehenden Studien konnte gezeigt werden, dass PatientInnen, die an einer psychischen Störung vom impulsiven-kompulsiven Spektrum (z.B. Alkoholabhängigkeit und Binge Eating Störung) leiden, reduziertes modellbasiertes Verhalten an den Tag legen (Sebold *et al.*, 2014; Voon *et al.*, 2015). In den Studien 4-6 wurde untersucht, ob reduziertes modellbasiertes Verhalten und erhöhtes modellfreies Verhalten auch bei Risikopopulationen von Sucht gefunden werden kann und so einen Vulnerabilitätsfaktor darstellt.

In *Studie 4* wurden deshalb gesunde, hoch- und niedrig impulsive Individuen mittels einer sequentiellen Entscheidungsaufgabe während der fMRT untersucht. Hierbei konnte kein Hinweis auf eine Reduktion des modellbasierten Systems bei hoch impulsiven Individuen gefunden werden. Es wurde jedoch eine leichte Erhöhung des modellfreien Systems festgestellt. In der fMRT-Analyse zeigten sich bei hoch impulsiven im Vergleich zu niedrig impulsiven Individuen reduzierte modellbasierte Signale im lateralen präfrontalen Kortex.

In der rein behavioralen *Studie 5* wurde mit einer sequentiellen Entscheidungsaufgabe modellbasiertes und modellfreies Verhalten bei gesunden, erwachsenen Individuen mit alkoholabhängigem Vater untersucht. Es ergaben sich keine Hinweise auf reduzierte modellbasierte oder erhöhte modellfreie Verhaltenskontrolle. Unabhängig von positiver Familienanamnese konnte jedoch ein Interaktionseffekt des Risikofaktors Impulsivität mit dem kognitivem Funktionsniveau auf modellbasierte Kontrolle beobachtet werden. In einer Reanalyse der Daten aus Studie 4 konnte dieser Interaktionseffekt repliziert werden.

Studie 6 beschäftigte sich mit der Fragestellung, ob akuter und chronischer Stress dazu führt, dass das Verhalten in einer sequentiellen Entscheidungsaufgabe weniger modellbasiert und stärker modellfrei geprägt ist. Akuter Stress wurde experimentell durch eine Vortragssituation

induziert (Kirschbaum *et al.*, 1993). Ein Effekt von akutem Stress per se auf Verhaltenskontrolle konnte nicht beobachtet werden. Physiologische Stressreaktivität (Kortisolreaktion auf Stressor) korrelierte negativ mit modellbasiertem Verhalten, während subjektive Stressreaktivität (Fragebogenratings) positiv mit modellbasiertem Verhalten assoziiert war. Ferner konnte gezeigt werden, dass die Interaktion von akutem und chronischem Stress modellbasierte Kontrolle beeinträchtigt: Lediglich bei Individuen, die ein hohes Maß an stressbehafteten Lebensereignissen berichteten, konnte eine Reduktion von modellbasiertem Verhalten nach Exposition an den akuten Stressor gefunden werden.

Gegenstand dieser Arbeit war es, geteilte und differentielle Mechanismen der Verhaltenskontrolle bei Alkoholabhängigkeit, Binge Eating Störung als vermeintlich suchtähnlicher Störung und bei Menschen mit Risikofaktoren für Sucht zu untersuchen. Zusammenfassend kann festgestellt werden, dass in beiden untersuchten PatientInnengruppen, sowohl auf der Verhaltensebene als auch neuronal, reduzierte Korrelate zielgerichteten Verhaltens nachgewiesen wurden. In beiden Gruppen wurde keine Veränderung hinsichtlich einfacherer, modellfreier Lernsignale festgestellt. Durch computationale Modellierung konnten teilweise unterschiedliche Mechanismen identifiziert werden, die zu diesen zunächst ähnlich erscheinenden behavioralen und neuronalen Veränderungen führen. Alkoholabhängige PatientInnen zeigten nach Bestrafung eine reduzierte Fähigkeit, alternative Optionen in ihre Entscheidungen einzubeziehen, während das Entscheidungsverhalten von Binge Eating PatientInnen durch eine unvorteilhafte Erhöhung von explorativem Verhalten gekennzeichnet war. Der über beide PatientInnengruppen gemeinsame Befund reduzierter zielgerichteter Lernsignale im medialen präfrontalen Kortex lässt den Rückschluss zu, dass es sich beim medialen präfrontalen Kortex um eine transdiagnostische Konvergenzzone handelt, die essentiell ist, um Verhaltenskontrollprozesse zu steuern, und deren Fehlfunktion den Verlust über die Kontrolle des eigenen Verhaltens zur Folge haben kann. Hinsichtlich der Risikofaktoren für Sucht kann zusammengefasst werden, dass die Befunde in den Risikogruppen qualitativ unterschiedlich zu den Ergebnissen in PatientInnenstudien sind (Sebold *et al.*, 2014; Voon *et al.*,

2015). In den vorliegenden Studien konnten mehrfach Interaktionseffekte von verschiedenen Risikofaktoren auf modellbasiertes Verhalten festgestellt werden. Dies deutet darauf hin, dass nicht einzelne Risikofaktoren alleine die Verhaltenskontrolle beeinträchtigen, sondern dies im Zusammenwirken mit dem Auftreten anderer Faktoren geschieht.

Aus den vorliegenden Ergebnissen lässt sich ableiten, dass Verhaltenskontrolle innerhalb von Längsschnittstudien in hohen Fallzahlen abgebildet werden sollte. Das Ziel wäre eine detaillierte Abbildung von vielen potenziellen Risikofaktoren und deren Zusammenwirken im Hinblick auf die Entwicklung von Verhaltenskontrolle bis hin zu Störungen der Verhaltenskontrolle. Hierfür könnte anhand der Ergebnisse aus Studie 3 die EEG eine geeignete und realisierbare Methode sein, um neuronale Maße auch in hohen Fallzahlen zu akquirieren. Um kausale Erklärungsansätze über die Ursachen für die Fortsetzung mittel- oder langfristig schädigenden Verhaltensweisen bei gleichzeitig vorhandenem Wissen über die Konsequenzen entwickeln zu können, sollte vertiefend die behaviorale und neuronale Entwicklung von Verhaltenskontrolle über den kompletten Prozess der Suchtentwicklung (Vulnerabilität, Entwicklung von Sucht- oder suchtähnlicher Erkrankung, Remission und potenzieller Rückfall) untersucht werden.

Supplementary Information

1. Supplementary Information for study 1 (SI-1)

Supplementary Methods and Materials

Participants. fMRI data were available of all 35 healthy participants and 34 patients, as 7 patients met MRI exclusion criteria and fMRI datasets from two patients had to be excluded due to raw data artifacts. Patients were recruited from an inpatient detoxification and rehabilitation program (*Soteria Clinic Leipzig*). All patients were free of any psychotropic medication for at least four plasma half-lives except for one patient taking doxepin due to sleeping problems. All subjects underwent the Structured Clinical Interview for DSM IV, Axis I disorders (First *et al.*, 2001; SCID-I) and patients additionally underwent a semi-structured interview on their individual addiction history. All patients were diagnosed as alcohol-dependent according to DSM-V and ICD-10. Alcohol-dependent patients did not meet criteria of any current comorbid psychiatric disorder. Included control participants did not report any current nor past psychiatric disorder (SCID-I).

Measures of Addiction Severity. Addiction severity was assessed using 1) Time-Line-Follow-Back Score (TLFB; Sobell, 1992), to assess alcohol units consumed in the month prior to treatment, 2) Obsessive-Compulsive-Drinking Scale (OCDS; Anton *et al.*, 1995), 3) Alcohol Craving Questionnaire (ACQ; Tiffany *et al.*, 2000), 4) Alcohol Use Disorder Identification Tests (AUDIT; Allen *et al.*, 1997).

Neurocognitive Measurements. Alcohol dependence is known to be linked with a number of cognitive deficits (Bates *et al.*, 2002; Goldstein *et al.*, 2004) which have recently been shown to be associated with impaired model-based decision-making (Sebold *et al.*, 2014).Therefore, participants completed a battery of neurocognitive tests on the following domains: working memory (Digit Span, Wechsler, 1955), cognitive speed (Digit-Symbol-Substitution Test, Wechsler, 1955), reasoning (Matrices Test, Amthauer, 1999), verbal IQ (German vocabulary test, Schmidt & Metzler, 1992), visual attention (Reitan Trail Making A, Reitan, 1955) and complex attention (Reitan Trail Making B, Reitan, 1955). Results and group comparisons are summarized in Table S1-1. We computed a factor analysis to extract composite measures of neurocognitive functioning. Based on an Eigenvector cutoff of >1, a factor analysis with an oblique rotation (direct oblimin) yielded a single factor solution, accounting for 59.61% of variance in the six tests obtained. The composite measure of neurocognitive functioning was subsequently used as a covariate in control analyses.

Task. In any of the 160 trials, participants decided between two cards each showing a different geometric stimulus with a maximum response time of 1.5 sec. The location (right vs. left side of the screen), where each of the stimuli appeared, was randomized over trials. After the participant had chosen one stimulus by button press (left vs. right button) the selected stimulus was highlighted and depicted for 1.5 sec minus reaction time. Feedback was shown for 0.5 sec (monetary win vs. monetary loss, indicated by a 10 Eurocent coin or a crossed 10 Eurocent coin, respectively). During the intertrial interval, a fixation cross was presented for a variable duration (jittered and exponentially distributed, range 1 sec - 12.5 sec). If no response occurred during the decision window, the message “too slow” was presented and no outcome was delivered.

In an instruction and training session outside the MRI scanner prior to the experiment, participants were informed that one of the two cards had a superior chance of winning money (probabilistic nature of the task). They were told that depending on their choice they could either win 0.10€ or lose 0.10€ per trial, that the aim was to win as much as possible and that the total amount of money gained would be paid out at the end of the experiment. Participants performed 20 training trials with a different set of cards and without any reversal of reward contingencies. Subsequently, participants were instructed that reward probabilities could change over the course of the main experiment and that they should track such changes to win as much money as possible. Importantly, no other information or details on reversals or the anti-correlated task structure were provided.

Computational Modeling: Tested models. First, the model-free SU-algorithm updates a decision value $Q_{a,t}$ for the chosen stimulus via the RPE $\delta_{Q_{a,t}}$ which is defined as the difference between the received reward R_t and the anticipated reward for the chosen stimulus $Q_{a,t}$:

(1) $\delta_{Q_{a,t}} = R_t - Q_{a,t}$

The RPE $\delta_{Q_{a,t}}$is used to iteratively update decision values of the chosen decision value trial-by-trial:

(2) $Q_{a,t+1} = Q_{a,t} + \alpha\delta_{Q_{a,t}}$

Here, α depicts the learning rate, which weights the influence of RPEs $\delta_{Q_{a,t}}$on the updated values. α has natural boundaries between 0 and 1. Importantly, this model neglects the anti-correlated task structure by only updating decision values for the chosen stimulus while the value of the alternative, unchosen stimulus $Q_{ua,t}$remains unchanged:

(3) $Q_{ua,t+1} = Q_{ua,t}$

Second, the DU-algorithm updates chosen and unchosen decision values in each trial. This takes into account the anti-correlated structure of the task. In our modeling approach, this is captured by additionally updating the unchosen decision values based on a different error signal, which compares the fictive outcome that might have happened with the value of the unchosen option. The RPE for the DU-model is:

(4) $\delta_{Q_{ua,t}} = -R_t - Q_{ua,t}$

The same learning rate α is used for updating unchosen values:

(5) $Q_{ua,t+1} = Q_{ua,t} + \alpha\delta_{Q_{ua,t}}$

Equation 5 gives the same weight to the update of unchosen decision values as to the chosen decision values. Third, and in contrast, we assume that the degree of updating the alternative choice option differs across individuals. To account for inter-individual variability regarding this process, we additionally constructed a 'hybrid' model to quantify each individual's degree of DU-learning. This is provided by the parameter κ, which weights the learning rate α for the unchosen RPE $\delta_{Q_{ua,t}}$.

(6) $Q_{ua,t+1} = Q_{ua,t} + \kappa\alpha\delta_{Q_{ua,t}}$

In the hybrid model, the RPE $\delta_{Q_{ua,t}}$ is weighted by the product of the learning rate for the chosen value and the weighting parameter κ, where $\kappa = 0$ reduces to the SU-Model and $\kappa = 1$ to the DU-Model. Note that this results in lower learning rates for DU-learning, which is in line with the key assumption that double-update learning is computationally costly.

To verify whether a model with a dynamic learning rate captures the observed behavior better than the DU algorithm with a fixed learning rate, we additionally tested a RL-model featuring gain adaptation by a variable learning rate. The Sutton-K1 model was introduced to combine dynamic learning rate methods with ideas from Kalman filtering and least square methods (Sutton, 1992) and was previously discussed and used as a non-hierarchical approximation for a dynamic learning rate (Chumbley *et al.*, 2012; Kepecs & Mainen, 2012; Landy *et al.*, 2012; Iglesias *et al.*, 2013). In this model, the values are equivalently updated via prediction errors as in equations (1) and (2). The dynamic learning rate is transformed with a logistic function to remain in boundaries between 0 and 1:

(7) $\alpha_t = \frac{1}{1+\exp(-\iota_t)}$

This is initialized with $\iota = 0$ corresponding to an initial learning rate of .5. The update of ι for the next trial depends on the change in reward prediction errors where

(8) $\iota_{t+1} = \iota(t) + \mu\delta_{Q_{a,t}} h_t$

and

(9) $h_{t+1} = (h_t + \alpha_t + \delta Q_{a,t}) * \max((1 - \alpha_t), 0)$

μ given in (8) is a free parameter which controls the individual degree of dynamic update of the learning rate. ι[8] is a sensitivity parameter of the learning rate, controlling the influence of the RPE of the last trial on a trial-by-trial basis as a function of μ.

Decision model. For all models, decisions are transformed into action probabilities by applying a softmax equation. The softmax equation includes the temperature β, which reflects the stochasticity of the choices; a' indicates all available choice options:

(10) $p(a_t) = \frac{exp(\beta Q(a))}{\sum_{a'} exp(\beta Q(a'))}$

Learning from rewards versus punishments. We also aimed to test the hypothesis whether a potential deficit of alcohol-dependent patients in DU-learning differs specifically as a function of learning from rewards versus learning from punishments in our task. In our models, we account for this by estimating separate learning rates and temperatures for reward and punishment α_{rew}, α_{pun} and β_{rew}, β_{pun}, respectively.

Model fitting. Fitting was performed in the same Bayesian framework as introduced in (Huys *et al.*, 2011; Huys *et al.*, 2012). To infer the maximum-a-posteriori estimate *MAP* of parameters θ for each individual *i*, we use a Gaussian prior with mean and variance μ and σ:

$$MAP_i = \text{argmax} \log p(Y \mid \theta)\ p(\theta \mid \mu, \sigma)$$

where Y represents the data in terms of actions A_i per subject i. We set priors empirically to the maximum likelihood estimates *ML* of μ and σ given the data by *all* subjects included:

$$ML_i = \text{argmax} \log p(Y \mid \theta)$$

[8]Note that this parameter is called β in the original publication which we here change to ι because β is used throughout the main manuscript to refer to the temperature in the decision model.

and achieve this by using Expectation-Maximization. Constrained parameters were transformed to a logistic (alphas, kappa) or exponential (beta) distribution to enforce constraints and to render normally distributed parameter estimates. All modeling analyses were performed using Matlab 2010b. This method was introduced and in-depth discussed in (Huys *et al.*, 2011; Huys *et al.*, 2012), was employed in several studies (Schlagenhauf *et al.*, 2013; Deserno *et al.*, 2015b; Huys *et al.*, 2015a) also in between-group designs (Chowdhury *et al.*, 2013; Deserno *et al.*, in press) including patient studies (Deserno *et al.*, 2014; Schlagenhauf *et al.*, 2014). It should be noted that the empirical prior mainly serves to mildly regularize parameters at the population level. As this was done based on the data of *all* participants, this renders between-group parameters valid.

Model comparison. For all models, we approximate the model evidence by integrating out the free parameters. This integral was approximated by sampling from the empirical prior distribution (Huys *et al.*, 2011; Huys *et al.*, 2012). Then, this integrated or marginal likelihood was submitted to a random-effects Bayesian model selection procedure (spm_BMS function contained in SPM8, Stephan *et al.*, 2009). The resulting exceedance probabilities *XP* show which model best accounts for behavior in each population. As this powerful technique is a relative comparison, we further show the validity of the inferred parameter by running simulations of the task based on the inferred parameters. Indeed, this reproduced the observed choice behavior well (S-Figure 1-2).

MRI data acquisition. Functional imaging was performed using a 3 Tesla Siemens Trio scanner to acquire gradient echo T2*-weighted echo-planar images with blood oxygenation level dependent contrast. Covering the whole brain, 40 slices were acquired in oblique orientation at 20° to the AC-PC line and in ascending order with 2.5-mm thickness, 3x3mm^2 in-plane voxel resolution, 0.5-mm gap between slices, TR=2.09s, TE=22ms, flip angle α=90°. Prior to functional scanning, a field distortion map was collected to account for individual homogeneity differences of the magnetic field. Additionally, T1-weighted anatomical images were acquired.

Preprocessing of fMRI data. For fMRI data analysis, we used SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). Images were corrected for delay of slice time acquisition. Voxel-displacement maps were estimated based on acquired field maps. In order to correct for motion, all images were realigned and additionally corrected for distortion and the interaction of distortion and motion. The images were spatially normalized to the Montreal Neurological Institute (MNI) space using normalization parameters generated during the segmentation of the individual T1-weighted structural image (Ashburner & Friston, 2005); thereafter, all images were spatially smoothed with an isotropic Gaussian kernel (6mm full width at half maximum).

Statistical analysis of functional MRI. Based on each individual's set of parameters identified during model fitting (random-effects parameters), we computed regressors for the statistical analysis of fMRI data. Using the general linear model approach (GLM) as implemented in SPM8, smoothed images were analyzed in an event-related manner. At the first level, onsets of feedback were entered into the model and convolved with the canonical hemodynamic response function and modulated parametrically by two trial-by-trial regressors from our modeling analysis: First, RPEs for chosen values were computed on basis of the SU-Model with $\kappa = 0$ (Model-free learner, RPE_{SU}).

Second, RPEs for chosen values were computed based on the DU-Model with $\kappa = 1$ (RPE_{DU}). Missing trials were modeled separately. To account for possible confounds due to movement, we included the six realignment parameters, the first temporal derivative of the translational realignment parameters and a further regressor censoring scan-to-scan movement higher than 1mm.

Voxel-based Morphometry. Each subject's anatomical T1-weighted image was segmented into different tissue classes using the unified segmentation approach implemented in SPM8 (Ashburner & Friston, 2005). Modulated images of gray matter density were smoothed using an isotropic Gaussian kernel (6mm full width at half maximum) and subjected to a random-effects model. The volume of gray matter, white matter and CSF tissue classes were summed in order to gain an individual estimate of total intracranial volume, which was entered as covariate in between-group comparisons. As there is strong evidence for pronounced cortical gray matter density loss in alcohol-dependent individuals (Beck *et al.*, 2012), we first tested for differences in gray matter density between patients and control group. The patient group was characterized by significantly reduced gray matter density (FWE-corrected for the whole brain $p<.05$) predominantly in a large cluster covering the cingulate cortex (see S-Table 1-7 and S-Figure 1-4). Second, in order to control for differences in gray matter density as a potential confound of our fMRI results, we extracted gray matter density from two regions of interest: (1) based on the fMRI analysis, the conjunction of both RPEs across the entire sample (thresholded at p-FWE <.05), (2) an anatomically predefined mask of combining frontal lobe and cingulate cortex (obtained from aal templates, WFUPickAtlas Toolbox).

Supplementary Results

Supplementary Raw Data Results. Following a reviewer's suggestion, we tested whether specifically patients acquire values but fail to reverse by analyzing correct responses as a function of stimulus (initially correct vs. post-reversal correct). An ANOVA with the between factor group and the within factor stimulus (including 15 trials respectively, to render realizations per stimulus equal) revealed a main effect of the factor stimulus ($F=24.11$, $p<.001$) and a main effect of group ($F=11.61$, $p=.001$), but, importantly, no interaction effect between the two factors ($F=1.52$, $p=.22$). The latter missing indication for a group x stimulus interaction points against a simple effect of perseveration regarding the value acquired for stimulus with an initially high reward probability as an account of the behavioral impairment observed in patients.

Model Simulation of Behavior. We simulated choices of the used RL-models by setting $\kappa = 0$ (SU-model), $\kappa = 0.5$ (hybrid model) and $\kappa = 1$(DU-model), in 1000 simulations, respectively and confirmed superiority of the hybrid model in terms of correct choices (S-Figure 1-1).

Model space. In the main manuscript, we report four models (SU, DU, hybrid, Sutton K1). The first three of these contained separate learning rates as well as separate temperature parameter in the softmax decision model for reward and punishment. Beforehand, we had systematically varied the influence of reward and punishment trials on learning rates in a 2x2 fashion. We verified by Bayesian Family Selection that the models containing both, separate learning rates for reward and punishment outperform models without this distinction (exp_r=.75, XP=.99). In a likewise manner, models with separate temperatures for reward and punishment outperform models without this distinction (exp_r=.93, XP=1).

Model Selection: Relative Model fit. Observed group differences in model evidences indicated that patients were best explained by the model-free Single-Update Model, whereas controls were best explained by the Hybrid Model, which includes individual variability regarding abstract inference on the alternative choice. We additionally confirmed that these results were robust against excluding n=7 participants who were not fitted better than chance (as indicated by a non-significant binominal test): Exceedance Probabilities for controls XP_{Hybrid}=.98, XP_{DU}=.002, XP_{SU}=.02, Exceedance Probabilities for patients XP_{Hybrid}=.32, XP_{DU}=.003, XP_{SU}=.68.

Further, we verified that these group differences were not driven by a small subgroup of patients. Looking at individual relative model fit, 25 of 35 healthy controls were better explained by the hybrid model than by the SU-Model, 25 of 43 patients were better explained by the SU than by the hybrid model. It is worth mentioning that while relative model fit indicated that patients were explained comparably better by the SU-model than by the hybrid model, in terms of absolute model fit, the hybrid model notwithstanding was able to explain the same majority of alcohol-dependent patients better than chance (n=37) as participants who were fit poorly were not fit by any model.

Model Selection: Absolute Model fit. To verify that the models explain actually observed behavior better than chance, we calculated a so-called predictive probabilities based on the negative loglikelihood (Daw, 2009), derived by exponentiating the average log likelihood per trial. Predictive Probabilities indicate how many of the observed data points can be accounted for by the inferred parameters. See S-Table 1-4 for results. Using a binominal test, we verified that individual predictive probabilities were significantly higher than chance. This resulted in n=7 participants which were not fitted better than chance by any of the models, meaning that 90% of the total sample and still 85% of the patient group was explained significantly better than chance by each of the models fit to the data. See S-Figure 1-3 for an illustration based on each individual's negative log likelihood. In line with this observation of appropriate model fit, adjusted McFadden's Pseudo R^2 for the hybrid model was .52, .60 for controls and .46 for patients.

Between-Group comparison of modeling-inferred parameters. We verified that the group difference on the double-update weighted learning rate was robust against excluding participants that were not fitted better than chance by any of the models. This was indeed the case (F=4.90, p=.03).

Analysis of simulated choice data. We were interested in whether the model replicates observed group differences on choice behavior (correct choices, win-staying, repetition of punished actions). Thus, we analyzed the choices simulated by the model (S-Figure 1-2) in the same manner as the original, empirical data. Hypotheses were directed, as we were interested in the replication of the empirically found effect. Thus, one-tailed tests were used. The model replicated the main effect of group on correct choices (t=4.56, p<.001), as well as the group effect on win-stay rates (t=3.05, p=.02) and on repetition behavior (t=1.76, p=.04). Inferred model parameters did not recover the group x phase interaction observed in the raw data.

Between-Group comparison of neural Single-Updating vs. Double-Updating signals. We verified that the result of significantly reduced coding of RPE_{DU} signatures in patients in the medial prefrontal cortex was robust against excluding participants that were not fitted better than chance by any of the models. Indeed, when excluding these n=7 participants, the group difference remained

significant (-10, 62, 12, t=4.24, FWE-corrected for the conjunction p_{peak}=.001; and -6, 56, 2 t=3.78, FWE-corrected for the conjunction p_{peak}=0.011).

Analyses of Covariance: Computational Modeling Analyses. Entering Smoking status as an additional between-factor in the MANOVA did not indicate a significant influence of the factor smoking (p=.18) and did not change the group difference in the double-update punishment learning-rate (p=.03). Gray matter density in fronto-limbic regions (as derived from the voxel-based morphometry analysis, VBM) as a covariate in the MANOVA on modeling parameters did not alter the observed results (group difference in double-update punishment learning p=.009). Including Gray matter density in the functional VOI (conjunction of RPE_{SU} and RPE_{DU} in the fMRI analysis) as a covariate in the reported analysis on modeling parameters did not change the observed results neither (group difference in double-update punishment learning p=.009). Likewise including the composite measure of cognitive functioning in the MANOVA on modeling parameters did not change the group difference in double-update punishment learning (p=.003). Controlling for depressive mood by entering the BDI score as a covariate did not affect the reported results (group difference on the DU-punishment learning rate (p=.03).

Analyses of Covariance: FMRI Analyses. Smoking status was entered as a covariate when testing for the RPE type x group interaction. Inclusion of the covariate did not alter the observed results in the medial prefrontal cortex (-10, 62, 12, t=4.0, FWE-corrected for the conjunction p_{peak}=.01); and posterior cingulate cortex (0, -40, 32, t=3.72, FWE-corrected for the conjunction p_{peak}=.025). Akinly, the post-hoc contrast was significant in the medial prefrontal cortex (-8, 62, 12, t=4.10, FWE-corrected for the conjunction p_{peak}=.007; and -6, 56, 2, t=3.55, FWE-corrected for the conjunction p_{peak}=0.04,).

We included gray matter density (as derived from VBM) in the anatomically predefined region of interest, in fronto-limbic parts of the brain as covariates when testing for the RPE type x group interaction. Including the covariate in the analysis did not affect the observed group differences in the medial prefrontal cortex (-10, 62, 12, t=3.96, FWE-corrected for the conjunction p_{peak}=.01); and posterior cingulate cortex (0, -40, 32, t=3.71, FWE-corrected for the conjunction p_{peak}=.026). Similarly, the post-hoc contrast remained significant in the medial prefrontal cortex (-8, 62, 12, t=4.19, FWE-corrected for the conjunction p_{peak}=.003; and -6, 56, 2, t=3.79, FWE-corrected for the conjunction p_{peak}=0.02).

Gray matter density values (as derived from VBM) in a functionally predefined region of interest (conjunction of RPE_{SU} and RPE_{DU} in the fMRI analysis) were likewise included as a covariate, when testing for the RPE type x group interaction. Inclusion of the covariate did not affect significance in medial prefrontal cortex (-10, 62, 12, t=3.97, FWE-corrected for the conjunction p_{peak}=.01); and posterior cingulate cortex (0, -40, 32, t=3.71, FWE-corrected for the conjunction p_{peak}=.026). Also the post-hoc contrast remained significant in the medial prefrontal cortex (-8, 62, 12, t=4.35, FWE-corrected for the conjunction p_{peak}=.003; and -6, 56, 2, t=3.76, FWE-corrected for the conjunction p_{peak}=0.02).

Next, we included the composite measure of cognitive functioning as a covariate when testing for the RPE type x group interaction. Inclusion of the covariate did not alter the findings in medial

prefrontal cortex (-10, 62, 12, t=3.96, FWE-corrected for the conjunction p_{peak}=.01); and posterior cingulate cortex (0, -40, 32, t=3.72, FWE-corrected for the conjunction p_{peak}=.021). Alike, the post-hoc contrast remained significant in the medial prefrontal cortex (-8, 62, 12, t=3.82, FWE-corrected for the conjunction p_{peak}=.02 and -6, 56, 2, t=3.59, FWE-corrected for the conjunction p_{peak}=.04).

To control for a possible influence of depressive mood on the observed results, we included the BDI score as a covariate when testing for the RPE type x group interaction. This did not affect significance in medial prefrontal cortex (-10, 62, 12, t=4.08, FWE-corrected for the conjunction p_{peak}=.01); and posterior cingulate cortex (0, -40, 32, t=3.72, FWE-corrected for the conjunction p_{peak}=.026). In line with our previous results, the post-hoc contrast remained significant in the medial prefrontal cortex (-8, 62, 12, t=4.38, FWE-corrected for the conjunction p_{peak}=.003, and -8, 56, 0, t=3.90, FWE-corrected for the conjunction p_{peak}=.01).

S-Table 1-1. Sample characteristics

	Healthy Controls	Alcohol-dependent patients	Test statistic
Demographic characteristics			
Age (35/43)	*M*=42.00 (SD=10.49)	*M*=44.42 (SD=10.21),	*t*=1.03, *p*=.307
Gender (male/female,35/43)	25 / 10	34 / 9	Chi^2=.434
Smokers (35/42)	16	33	Chi^2=.003
Handedness according to Edinburgh Handedness scale (right/both/left, 35/39)	32/3	33/5/1	Chi^2=.521
School leaving qualification (none/ basic secondary schooling/intermediate school certificate/university entrance qualification,35/41)	0/5/14/16	1/12/25/3	Chi^2=.001
Total years of unemployment (35/41)	*M*=0.9 (*SD*= 1.58)	*M*=4.54 (*SD*= 6.37)	*t* =3.27, *p*=.002
Neuropsychological Measurements			
Reasoning (Matrices) (35/41)	*M*=10.91(*SD*= 4.00)	*M*=6.71(*SD*= 3.64)	*t*=4.80, *p*<.001
Working Memory (Backward Digit Span) (35/42)	*M*=7.49 (*SD*= 2.50)	*M*=6.19(*SD*= 2.00)	*t*=2.54, *p*=.013
Cognitive Speed (Trail Making A) (35/42)	*M*=27.31 (*SD*= 14.44)	*M*=38.82 (*SD*= 18.10)	*t*=-3.04, *p*=.003
Complex attention, (Trail Making Test B) (35/42)	*M*=62.84(*SD*= 28.59)	*M*=101.82 (*SD*= 79.52)	*t*=2.75, *p*=.007
Cognitive speed (Digit Symbol Substitution Test) (35/41)	*M*=79.91(*SD*= 18.38)	*M*=60.85(*SD*= 16.14)	*t*=4.81, *p*<.001
Premorbid IQ (German vocabulary test) (35/41)	*M*=31.74 (*SD*= 3.38)	*M*=24.20(*SD*= 6.96)	*t*=5.85, *p*<.001
Barrat Impulsiveness Scale (35/42)	*M*=59.96 (*SD*= 10.03)	*M*=65.81 (*SD*= 9.18)	*t*=2.80, *p*=.007
Clinical characteristics			
Alcohol units (month before participation/ treatment begin, 35/38)	*M*=20.43 (SD=21.67)	*M*=301.61 (*SD*=294.06)	*t*= 5.64, *p*<.001
Obsessive Compulsive Drinking Scale (31/42)	*M*=3.65 (*SD*=3.86)	*M*=25.55 (*SD*=9.78)	*t*=11.80, *p*<.001
Alcohol Use Disorder Identification Test (35/42)	*M*=4.26 (*SD*=3.18)	*M*=26.24 (*SD*=8.72)	*t*=14.14, *p*<.001
Alcohol Craving Questionnaire (34/42)	*M*=1.3 (*SD*=0.38)	*M*=2.04 (*SD*=0.88)	*t*=4.42, *p*<.001
Duration of dependence (years) (36)	-	*M*=14.64 (*SD*=9.96)	-
Number of preceding detoxification treatments (35)	-	*M*=3.43 (*SD*=3.99)	-
Beck Depression Inventory (33/41)	*M* =5.09 (*SD*=6.32)	*M* =13.59 (*SD*=10.02)	*t*=4.24, *p*<.001

S-Table 1-2. Model Selection Results. All models were compared using Bayesian Model Selection. We report Exceedance Probabilities (XP), protected Exceedance Probabilities (pXP) and posterior model probabilities (exp_r)

	Single-Update			Double-Update			Hybrid			Sutton K1		
	XP	*pXP*	*exp_r*	*XP*	*pXP*	*exp_r*	*XP*	*pXP*	*exp_r*	*XP*	*pXP*	*exp_r*
Overall	.11	.19	.31	.00	.15	.08	.89	.51	.43	.00	.15	.18
Controls	.003	.22	.18	.00	.20	.12	.99	.39	.54	.00	.20	.16
Patients	.93	.58	.46	.00	.13	.07	.06	.16	.27	.01	.13	.20

S-Table 1-3. Best fitting parameters of the hybrid model

		β_{reward}	β_{punish}	α_{reward}	α_{punish}	*K*	*Initial Q*
25 th Percentile	All	3.53	0.80	0.27	0.19	0.03	-0.44
	HC	3.35	0.98	0.41	0.30	0.04	-0.33
	Patients	3.57	0.61	0.17	0.08	0.03	-0.51
Median	All	4.18	1.38	0.59	0.49	0.09	-0.23
	HC	4.52	1.74	0.59	0.47	0.12	-0.24
	Patients	4.08	1.22	0.49	0.53	0.08	-0.22
75th percentile	All	4.95	1.92	0.79	0.71	0.22	0.08
	HC	4.99	2.48	0.75	0.73	0.25	0.07
	patients	4.78	1.57	0.81	0.71	0.20	0.09

S-Table 1-4. Mean Predictive Probabilities per group and winning model

	Single-Update Model	Hybrid Model
Overall	m=0.71, SD=.10	m=0.73, SD=.11
Healthy Controls	m=0.76, SD=.10	m=0.77, SD=.10
Alcohol-Dependent Patients	m=0.68, SD=.09	m=0.69, SD=.10

S-Table 1-5. Neural signatures of single-update learning (RPE_{SU}) for both healthy controls and alcohol-dependent patients taken together at p<.05 FWE whole brain corrected.

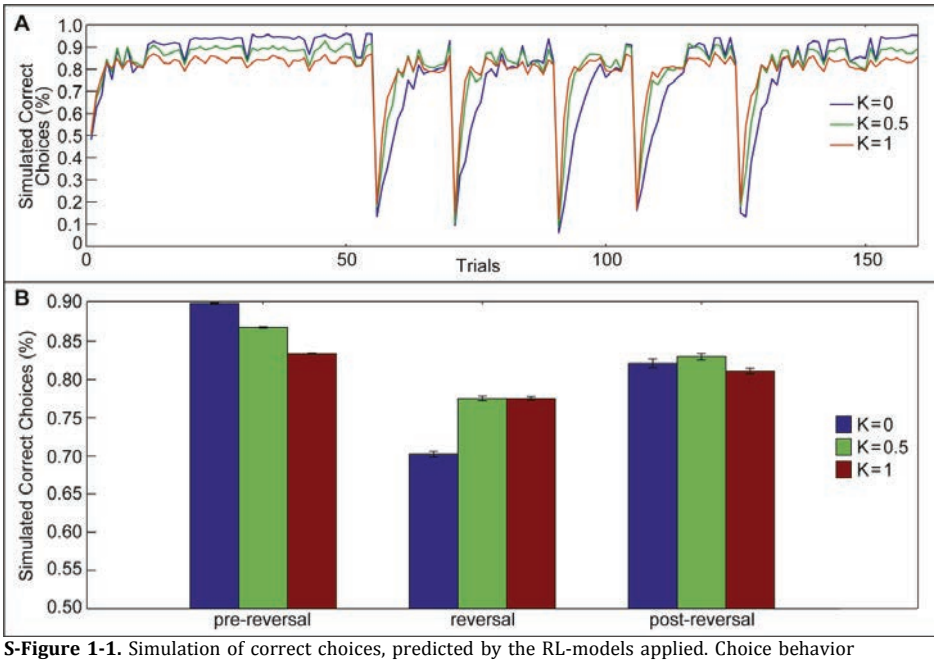
Single-Update Signals				
	MNI coordinate	Cluster size	T	p-FWE peak
Ventral Striatum	-8 8 -10	57	8.66	<0.001
Ventral Striatum	12 8 -10	82	8.54	<0.001
Middle Orbital Gyrus	6 42 -8	201	7.89	<0.001
Middle Orbital Gyrus	8 60 4		6.18	<0.001
Superior Medial Gyrus	-10 64 12	80	6.00	0.001
Middle Orbital Gyrus	-6 54 2		5.85	0.002
Anterior Cingulate Gyrus	-6 44 6	34	6.10	0.001
Anterior Cingulate Gyrus	-4 30 16	14	5.62	0.004
Middle Orbitofrontal Gyrus	-24 32 -16	20	5.55	0.006
Putamen	-26 -6 6	21	5.62	0.004
Putamen	26 0 2	17	5.53	0.006
Posterior Cingulate Gyrus	0 -34 34	68	6.30	<0.001
Precuneus	-4 -50 16	29	5.55	0.006
Angular Gyrus	-46 -70 34	17	5.46	0.008
Cerebellum	-44 -74 -34	161	6.80	<0.001
Cerebellum	36 -72 -40	93	6.20	<0.001
Cerebellum	44 -72 -32		5.49	0.008

S-Table 1-6. Neural signatures of double-update learning (RPE_{SU}) for both healthy controls and alcohol-dependent patients taken together at p<.05 FWE whole brain corrected.

Double-Update Signals				
	MNI coordinate	Cluster size	T	p-FWE peak
Middle Orbital Gyrus *Rectal Gyrus* *Inferior Frontal Gyrus*	-2 56 -4 -6 44 -10 -34 36 -10	2681	11.20 10.52 10.20	<0.001 <.0.001 <0.001
Inferior Frontal Gyrus	34 36 -12	86	6.77	<0.001
Superior Frontal Gyrus *Superior Frontal Gyrus* *Middle Frontal Gyrus*	-12 48 36 -18 38 50 -24 32 50	190	6.65 5.79 5.76	<0.001 0.002 0.002
Insula	-38 -2 14	37	5.88	0.001
Ventral Striatum. *Ventral Striatum* *Anterior Cingulate Cortex* 2 20 0	-6 8 -10 10 12 -8	386	7.43 7.01 6.66	<0.001 0.001 <0.001
Caudate	20 18 26	52	5.59	0.005
Hippocampus *Hippocampus* *Fusiform Gyrus*	-30 -12 -18 32 -28 -10 -32 -36 -14	188	7.61 5.88 5.5	<0.001 0.001 0.007
Hippocampus *Fusiform Area*	38 -24 -14 42 -18 -18	52	6.44 5.88	<0.001 0.001
Posterior Cingulate Gyrus *Precuneus* *Posterior Cingulate Gyrus*	-2 -42 32 -6 -54 18 -4 -52 26	1201	9.05 8.7 8.14	<0.001 <0.001 <0.001
Middle Temporal Gyrus *Angular Gyrus* *Middle Temporal Gyrus*	-50 -70 22 -46 -72 32 -44 -60 22	180	6.23 5.96 5.92	0 0.001 0.001
Middle Temporal Gyrus	58 -8 -22	29	6.25	<0.001
Middle Temporal Gyrus	-60 -10 -20	52	6.12	0.001
Superior Temporal Gyrus *Superior Temporal Gyrus*	68 -22 14 60 -24 16	17	5.3 5.19	0.016 0.025
Temporal Pole	56 2 6	22	5.79	0.002
Operculum	44 -20 20	40	5.67	0.004
Precentral Gyrus *Postcentral Gyrus*	42 -16 62 40 -26 58	69	6.14 5.07	<0.001 0.038

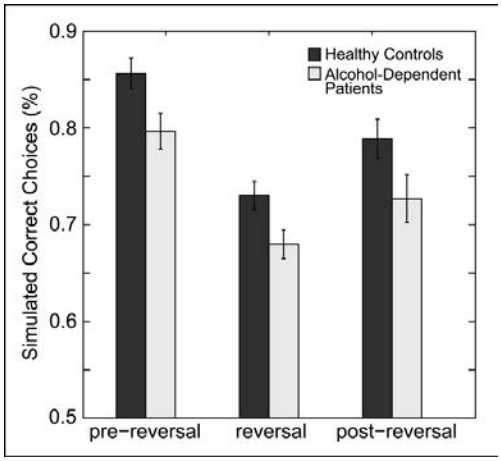
S-Table 1-7. Voxel-Based Morphometry: Group differences. Controls > alcohol-dependent patients at FWE whole brain corrected p<.05.

Voxel-Based Morphometry				
	MNI coordinate	Cluster size	T	p-FWE peak
Suppl Motor Cortex *Superior Medial Gyrus* *Middle Cingulate Cortex*	2 4 48 2 30 38 0 12 38	523	7.60 7.587.1 1	<0.001 <0.001 <0.001
Superior Medial Gyrus	4 52 38	26	7.44	<0.001
Middle Cingulate Cortex *Middle Cingulate Cortex* *Precuneus*	0 -32 46 2 -40 50 0 -46 38	378	7.25 6.53 6.44	<0.001 0.001 0.002
Precuneus	2 -78 40	22	6.41	0.002
Anterior Cingulate	-2 48 16	24	6.36	0.003
Frontal Pole	50 48 26	24	5.93	0.012

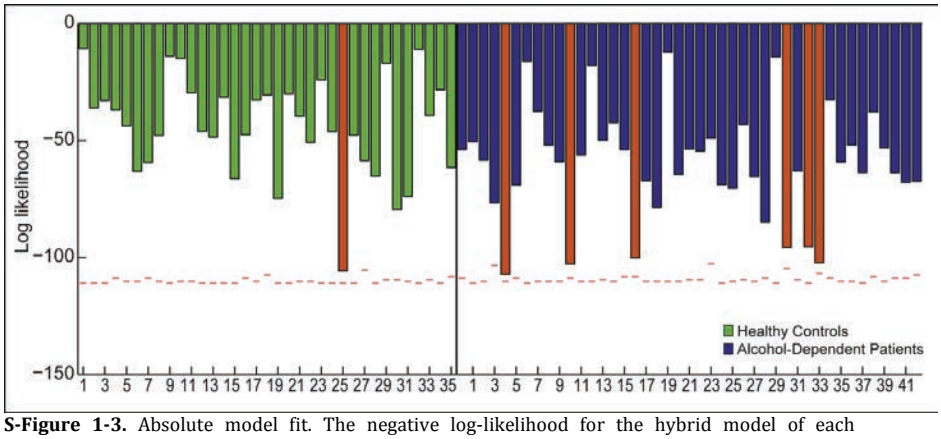


S-Figure 1-1. Simulation of correct choices, predicted by the RL-models applied. Choice behavior was simulated 1000 times per model (κ =0 (SU-model), κ =0.5 (hybrid model) and κ =1(DU-model). Learning rates α were fixed to .5 and the decision noise β to 3, respectively.

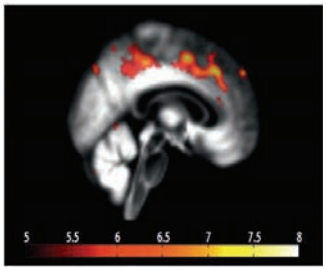
Note that the advantage of double-update weighted learning is prominent in blocks that require flexible behavioral adaptation. This is reflected by faster simulated adaptation after reversals (S-Figure 1-1A) and also a higher average of simulated correct responses (S-Figure 1-1B).



S-Figure 1-2. Simulation of Correct Responses on the basis of each individual's inferred parameters. Simulation of choice reproduced empirically observed choices (compare Figure 5-2), with the exception of not recapturing the group by phase interaction observed in the empirical data.



S-Figure 1-3. Absolute model fit. The negative log-likelihood for the hybrid model of each participant. Values closer to zero indicate better fit. Marked in red, we show participants that are fit by the model not significantly better than chance by the model (binominal test). Red lines indicate chance fit, which varies slightly based on the number of valid trials available.



S-Figure 1-4. Group differences (controls > alcohol-dependent patients) in Voxel-Based-Morphometry.

2. Supplementary Information for study 2 (SI-2)

Supplemental Methods

Participants. All subjects underwent the Structured Clinical Interview for DSM IV, Axis I disorders, SCID-I (First *et al.*, 2001). Healthy controls (HC) who were included did not report any current nor past psychiatric disorder. Patients suffering from binge eating disorder (BED) were diagnosed according to DSM-5 criteria by a psychologist using the German version of the structured Eating Disorder Examination Interview (Hilbert *et al.*, 2004). As Body-Mass-Index is not a diagnostic criterion according to DSM-5, patients were included irrespective of their BMI (Dingemans & van Furth, 2012). Participants who were included did not use any psychotropic medication. Due to raw data artifacts, fMRI datasets from two participants (n=1 BED, n=1 HC) were excluded. For demographic and clinical characteristics, see S-Table 2-1.

Neuropsychological testing. Behavioral control was shown to be linked to general cognitive capacities (Otto *et al.*, 2013b; Schad *et al.*, 2014; Otto *et al.*, 2015), which might relate to between-group differences in patient studies (Sebold *et al.*, 2014). Thus, participants underwent neuropsychological testing in an independent session on the following domains: working memory (Digit Span, Wechsler, 1955), cognitive speed (Digit-Symbol-Substitution Test, Wechsler, 1955), reasoning (Matrices Test, Amthauer, 1999), verbal IQ (German vocabulary test, Schmidt & Metzler, 1992) visual attention (Reitan Trail Making A, Reitan, 1955) and executive functioning (Reitan Trail Making B, Reitan, 1955). For results, see S-Table 2-2.

Task. Maximum response time was 1.5s. The displayed location (right vs. left) was randomized over trials. After choice of one stimulus by button press, the selected card was highlighted for 1.5s minus reaction time. Feedback (monetary win, 10 Eurocent vs. monetary loss, crossed 10 Eurocent) was displayed for 0.5s. During the inter-trial interval, a fixation cross was shown for a variable duration (jittered and exponentially distributed, range 1–12.5s, mean 2.5s). On average, trials were 4s long. If no response occurred in time, no outcome but the message "too slow" was presented. Mean number of missing trials was 1.14 (SD=2.06, maximum: 9). No significant group difference on missing trials was observed (t=0.29, p>.77).

Prior to the experiment (outside the MRI scanner), participants were instructed to opt for the card with the higher chance of winning. Depending on their choice, they could either win or lose 10 Eurocent per trial and the balance was paid to them at the end of the experiment. Participants were informed that reward probabilities might change over the course of the main experiment. These instruction slides did not provide details of reward probabilities, reversals nor of the task structure. The instruction session included 20 training trials with a different set of cards and without any reversal.

Computational Modeling of choice behavior. Different Reinforcement-Learning (RL-) models were fit to the observed choice data: a model-free Single-Update (SU-) model, a full Double-Update (DU-) model, and a third model that individually weights the degree of double-updating via a parameter (iDU). In the following, we describe each model in detail accompanied by equations.

First, the SU-algorithm updates a decision value $Q_{a,t}$ for the chosen stimulus via the RPE $\delta_{Q_{a,t}}$ which is defined as the difference between the received reward R_t and the anticipated reward for the chosen stimulus $Q_{a,t}$:

(1) $\delta_{Q_{a,t}} = R_t - Q_{a,t}$

The RPE $\delta_{Q_{a,t}}$ is used to iteratively update decision values of the chosen decision value trial-by-trial:

(2) $Q_{a,t+1} = Q_{a,t} + \alpha\delta_{Q_{a,t}}$

Here, α depicts the learning rate, which weights the influence of RPEs $\delta_{Q_{a,t}}$ on the updated values. α has natural boundaries between 0 and 1. Importantly, this model neglects the counterfactual task structure by only updating decision values for the chosen stimulus while the value of the unchosen stimulus $Q_{ua,t}$ remains unchanged:

(3) $Q_{ua,t+1} = Q_{ua,t}$

Second, the DU-algorithm updates chosen and unchosen decision values in each trial. This takes the counterfactual structure of the task into account. In our modeling approach, this is captured by additionally updating the unchosen decision The RPE for the DU-model is:

(4) $\delta_{Q_{ua,t}} = -R_t - Q_{ua,t}$

The same learning rate α is used for updating unchosen values, thus, equation 5 gives the same weight to the update of unchosen decision values as to that of chosen decision values:

(5) $Q_{ua,t+1} = Q_{ua,t} + \alpha\delta_{Q_{ua,t}}$

Third, the iDU-algorithm assumes that the degree of updating the alternative choice option differs across individuals. This is provided by the parameter κ, which weights the learning α for the unchosen RPE $\delta_{Q_{ua,t}}$:

(6) $Q_{ua,t+1} = Q_{ua,t} + \kappa\alpha\delta_{Q_{ua,t}}$

Please note that the three models described are nested. In the iDU-model, the RPE $\delta_{Q_{ua,t}}$ is weighted by the product of the learning rate for the chosen value and the parameter κ, where $\kappa = 0$ reduces to the SU-Model and $\kappa = 1$ to the DU-Model. This results in lower learning rates for DU-learning. For the task at hand, as double-updating depends on abstract inference derived from feedback actually experienced, updating the unchosen stimulus always relies on learning from feedback for the chosen stimulus, that is, it is rather unlikely to be a process which is independent from updating the chosen stimulus (compare (Li & Daw, 2011) for the identical implementation).

Additionally, we included a model with an adaptive learning rate, Sutton-K1 (Sutton, 1992). Sutton K-1 was discussed and used as a non-hierarchical approximation of a dynamic learning rate (Chumbley *et al.*, 2012; Kepecs & Mainen, 2012; Landy *et al.*, 2012; Iglesias *et al.*, 2013). By including it, we tested whether a dynamic learning rate captures the observed behavior generally better than algorithms with a fixed learning rate. In this model, values are also updated via prediction errors as in equations (1) and (2). Differently, learning rate α is dynamically updated as a function of the change in prediction errors encountered (Sutton, 1992). The dynamic learning rate is transformed with a logistic function to remain in boundaries between 0 and 1:

(7) $\alpha_t = \frac{1}{1+\exp(-\iota_i)}$

This is initialized with ι =0 corresponding to an initial learning rate of .5. The update of ι for the next trial depends on the change in reward prediction errors where

(8) $\iota_{t+1} = \iota(t) + \mu\delta_{Q_{a,t}} h_t$

and

(9) $h_{t+1} = (h_t + \alpha_t \delta Q_{a,t}) * \max((1 - \alpha_t), 0)$

μ given in (8) is a parameter which controls dynamic update of the learning rate. ι is a sensitivity parameter of the learning rate, controlling the influence of the RPE of the last trial on a trial-by-trial basis as a function of μ. Again, note that this model is nested with RL-models with a constant learning rate because setting μ=0 keeps α constant.

Decision model. For all models, decisions are transformed into action probabilities by applying a softmax equation. This includes a parameter β, which reflects the stochasticity of the choices and captures the exploration–exploitation dimension.

(7) $p(a) = \frac{\exp(\beta Q(a))}{\sum \exp(\beta Q(a'))}$

Model Fitting. Models were fitted using the HGF toolbox 2.0 (Poldrack, 2006; Mathys et al., 2011) as part of TNU Algorithms for Psychiatry-Advancing Science (TAPAS, http://www.translationalneuromodeling.org/tapas/). For priors on parameters of the learning algorithm and the observation model (softmax), please see S-Table 2-3. The negative variational free energy was maximized to gain a minimal Kullback-Leibler divergence, that is, the minimal divergence between true (exact) and approximate posterior distribution (Friston et al., 2007). For optimization, a quasi-Newton optimization algorithm was applied. As an approximation to the log-evidence, the negative variational free energy was subjected to random-effects Bayesian Model Selection (BMS) for each model and each individual (Stephan et al., 2009). After comparison of best-fitting modeling parameters between groups, we also controlled for the possibility that parameter comparisons can be confounded by poor absolute model fit (Akam et al., 2015), namely that a model cannot explain the data better than chance. This was done by looking at each individual's negative log-likelihood (the probability that the data is given by the parameters) relative to number the number of trials. If this "percentage of explained trials" did not exceed .55, a subject was classified as not fitting better than chance. This was the case for two patients.

MRI data acquisition. Functional imaging was conducted on a 3 Tesla Siemens Trio scanner to acquire gradient echo T2*-weighted echo-planar images with blood oxygenation level dependent contrast (40 slices at 20° to the AC-PC line, ascending order, 2.5-mm thickness, 3x3mm^2 in-plane voxel resolution, 0.5-mm gap, TR=2.09s, TE=22ms, flip angle α=90°). To account for individual homogeneity differences of the magnetic field, we acquired a field distortion map. T1-weighted anatomical images were collected for normalization purposes.

Preprocessing of fMRI data. Data were preprocessed and analyzed using SPM8. Images were corrected for delay of slice time acquisition. Voxel-displacement maps were estimated based on acquired field maps. For the purpose of motion correction, all images were realigned and additionally corrected for distortion and the interaction of distortion and motion. Normalization parameters were derived from the segmentation of the individual T1-weighted structural image (Ashburner & Friston, 2005) and used for spatial normalization of the functional images to the

Montreal Neurological Institute space. Normalized images were spatially smoothed (isotropic Gaussian kernel, 6mm full-width at half maximum).

Supplemental Results

Neuropsychological Measurements. We tested for group differences in general cognitive capacities by subjecting results of all neuropsychological tests (S-Table 2-2) to a Multivariate Analysis of Variance (MANOVA) with the between-subject factor group. No significant effect of group was observed ($F=1.52$, $p=.19$).

Brain-Behavior Relationships. For both groups separately, we tested for an association between BOLD activation in response to explorative trials at the peak coordinate of the between-group difference in aI/lPFC and behavioral performance (correct choices, switching behavior). Results did not indicate any significant correlation (Correlation with correct choices: $r_{HC}= .376$ $p_{HC}=.093$, $r_{BED} =.200$, $p_{BED.}=.383$; Correlation with switching: $r_{HC}= .054$ $p_{HC}=.814$,$r_{BED} .193$, $p_{BED.} =.40$).

Supplemental Tables

S-Table 2-4. Sample Characteristics. BMI=Body Mass Index, EHI=Edinburgh Handedness Inventory.

	Healthy Controls	Binge Eating Patients	Test statistic
Demographic characteristics			
Age (22/22)	*M*=27.8 (*SD*=4.54)	*M*=29.0 (*SD*=9.40)	*t=1.27, p=.57*
Gender (male/female,22/22)	7/15	6/16	Chi^2=,11 p=.74
Handedness according to EHI (right/both/left, 22/22)	20/0/2	17/2/3	Chi^2=2.44, p=.30
Smokers (22/22)	8	9	Chi^2=.10. p=.76
BMI (22/22)	*M*=26.06 (*SD*=4.35)	*M*=28.27 (*SD*=6.58),	*t*=1.31, *p*=.20
School leaving qualification (none/compulsory basic secondary schooling/intermediate school certificate/university entrance qualification, 22/22)	0/0/1/21	0/0/5/17	Chi^2=3.09, p=.08
Total years of unemployment (22/22)	*M*=0.5 (*SD*= 1.05)	*M*=0.7 (*SD*= 1.68)	*t* =0.55, *p*=.60
Clinical characteristics			
Yale Food Addiction Scale (17/22)	*M*=0.82 (*SD*=0.63)	*M*=4.68 (*SD*=2.21)	*t*=3.88, *p*<.001
Food Craving State (22/22)	*M*=27.55 (*SD*=12.41)	*M*=41.00 (*SD*=11.18)	*t*=7.78, *p*<.001
Food Craving Trait (17/22)	*M*=78.82 (*SD*=18.41)	*M*=158.14 (*SD*=34.10)	*t*=8.65, *p*<.001
Number of Objective Binge Eating Episodes (last 28 days) (20)	-	*M*=8.25 (*SD*=3.42)	-
Age of Onset (14)	-	*M*=18.93 (*SD*=9.01)	-
Duration of Disease (14)	-	*M*=8.29 (*SD*=7.01)	-

S-Table 2-5. Neuropsychological Tests. HC healthy controls BED binge eating patients.

Neuropsychological Measurements		
Measurement	*HC*	*BED*
Reasoning (Matrices) (21/21)	12.29 (*SD*=2.99)	10.38 (*SD*=3.64)
Working Memory (Backward Digit Span) (21/21)	8.38 (*SD*=7.52)	7.52 (*SD*=2.16)
Visual Attention (Trail Making A) Time (21/21)	21.74 (*SD*=6.88)	27.22 (*SD*=9.01)
Visual Attention (Trail Making A) Errors (21/21)	0 (*SD*=0)	.05 (*SD*=.22)
Complex Attention / Task Switching (Trail Making B) Time (21/21)	46.70 (*SD*=14.34)	69.46 (*SD*=56.22)
Complex Attention / Task Switching (Trail Making B) Errors (21/21)	0.14 (*SD*=.36)	0.57 (*SD*=1.17)
Cognitive Speed (Digit Symbol Substitution Test)	87.33 (*SD*=14.54)	81.90 (*SD*=13.65)
Verbal Intelligence (German Vocabulary Test) (21/21)	110.29 (*SD*=8.83)	102.24 (*SD*=11.80)

S-Table 2-6. Priors of parameters. The decision parameter of the observation model β was estimated in log-space and also parameters μ and h of Sutton K1; parameters α and κ of the learning algorithms were estimated in logit-space.

	Prior Mean	Prior Variance
Observation Model for all learning models		
Softmax		
β	1	16
Learning models		
SU		
κ	0	0
α	.55	1
SU-WL		
κ	0	0
α(w/l)	.5 / .6	1 / 1
DU		
κ	1	0
α	.25	1
DU-WL		
κ	1	0
α(w/l)	.4 / .1	1 / 1
iDU		
κ	.1	1
α	.25	1
iDU-WL		
κ	.1	1
α(w/l)	.55 / .45	1
Sutton-K1		
μ	1	1
H	.005	1

S-Table 2-7. Model Selection: Expected Posterior Probabilities (PP) and Exceedance Probabilities (XP) for all models. SU=single-update, DU=double-update, iDU=inidividually-weighted double-update, WL indicates that the model had separate learning rates for wins and losses HC healthy controls BED binge eating patiens

		SU	SU-WL	DU	DU-WL	iDU	iDU-WL	Sutton K-1
All (n=44)	PP	.160	.072	.052	.150	.271	.231	066
	XP	.054	.001	<.001	.039	601	.303	<.001
HC (n=22)	PP	.130	.073	.060	.180	.241	.280	.040
	XP	.035	.005	.003	.113	.321	.522	<.001
BED (n=22)	PP	.175	.117	.096	.099	.219	.158	.137
	XP	.210	.058	.030	.033	.426	.150	094

S-Table 2-8. Descriptive Statistics of parameters for iDU, the best-fitting model across HC and BED. The index c index indicates the learning rate for chosen stimulus. M=mean, SD=standard deviation.

	β	α_c	$\kappa^*\alpha_c$
Healthy Controls (n=22)	*M*=6.13, *SD*=2.67	*M*=0.54, *SD*=0.19	*M*=0.07, *SD*=0.04
Binge Eating Patients (n=22)	*M*=4.17, *SD*=2.50	*M*=0.49, *SD*=0.24	*M*=0.07, *SD*=0.06

S-Table 2-9. Distribution of inferred parameters for iDU, the best-fitting model across HC and BED

	β	α_c	$\kappa^*\alpha_c$
25th Percentile	2.78	.37	.03
50th Percentile	4.58	.56	.07
75th Percentile	7.73	.67	.09

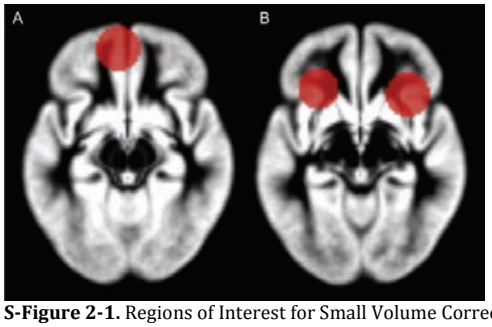
S-Table 2-10. Single-Update activation, Double-Update activation (only clusters k=>10) and conjunct activation of both.

Region	MNI coordinates	cluster size	T	p (FWE)
Single-Update RPE Activation				
	8 12 -10		6.18	<0.001
Caudate	14 6 -14	76	5.98	<0.001
	28 -4 4		6.58	<0.001
	30 -12 4		6.22	0.001
Putamen	30 6 4	101	5.65	0.009
	-12 6 -14		6.90	<0.001
	-18 -6 -18		6.63	<0.001
Amygdala	-26 -4 -18	141	6.07	<0.001
Hippocampus	28 -18 -18	12	5.65	0.009
	26 -6 -20		6.36	0.001
Hippocampus	18 -8 -18	24	5.98	0.002
Inferior Frontal Gyrus	-34 36 -16	18	5.58	0.001
	26 30 -18		5.86	0.004
Inferior Frontal Gyrus	30 36 -14	49	5.89	0.004
	-14 60 -2		6.20	0.001
Superior Medial Gyrus	-14 66 -6	24	5.20	0.042
Middle Orbital Gyrus	-6 60 -8	19	5.90	0.003
	-6 50 -14		5.41	0.020
	6 -48 30		6.02	0.002
Posterior Cingulate Cortex	-4 -36 38	67	5.19	0.043
	-60 -40 -12		6.36	<0.001
Middle Temporal Gyrus	-62 -32 -14	38	5.66	0.012
Inferior Temporal Gyrus	-56 -58 -8	14	5.34	0.004
	-40 -70 -42		7.29	<0.001
	-40 -78 -30		6.58	<0.001
Cerebellum	-28 -86 -28	204	6.05	<0.001
	44 -74 -32		6.81	<0.001
Cerebellum	44 -68 -42	101	6.34	0.001
Double-Update RPE Activation				
Caudate	-4 18 -8	13	6.27	<0.001
Angular Gyrus	-40 -68 30	24	5.79	<0.001
Middle Orbital Gyrus	-8 54 -2	19	5.62	0.009
Conjunction Single-Update and Double-Update RPE				
Middle Orbital Gyrus	-6 52 -12	3	5.27	0.033
Inferior Frontal Gyrus	32 36 -12	1	5.79	0.040

S-Table 2-11. Activation in Explorative vs. Exploitative Trials (F-contrast). vlPFC = ventrolateral prefrontal cortex.

Main Effect. Exploration vs. Exploitation				
Region	MNI coordinate	Cluster size	F	p(FWE-corr)
aI/vlPFC	-28 26 -2	41	59.95	<0.001
aI/vlPFC	32 24 -8	17	53.34	0.001

Supplemental Figure



S-Figure 2-1. Regions of Interest for Small Volume Correction of between-group differences. A) The peak voxel of the conjunction of both types of RPE (RPE_{SU} /RPE_{DU}) across the entire sample was surrounded by a 15mm sphere. B) For the neural analysis of exploration, the whole-brain peak voxels across all participants in left and right insula were each surrounded with 15mm spheres and combined in one volume.

3. Supplementary Information for study 4 (SI-3)

Computational modeling. As in previous studies (Daw et al., 2011; Wunderlich et al., 2012b; Deserno et al., 2015b; Voon et al., 2015), we adopted a modeling approach to disentangle influences of model-free and model-based control on participant's choice behavior. As described in the main manuscript, three types of algorithms were applied. All three algorithms learn values (Q) for each of the stimuli, which appear in the task as three pairs (sA, sB, sC). sA refers to the first-stage stimuli where values of model-free and model-based algorithms differ. sB and sC refer to the two pairs of second-stage stimuli. In the following equations, a indexes the chosen stimulus and index i denotes the stage (i=1 for SA at the first stage and i=2 for SB or SC at the second stage) and t signifies the trial.

First, the model-free algorithm was SARSA (λ) which learns values retrospectively after prediction errors occurred (Sutton & Barto, 1998):

[1] $Q_{MF}(s_{i,t+1}, a_{i,t+1}) = Q_{MF}(s_{i,t}, a_{i,t}) + \alpha_i \delta_{i,t}$

[2] $\delta_{i,t} = r_{i,t} + Q_{MF}(s_{i+1,t}, a_{i+1,t}) - Q_{MF}(s_{i,t}, a_{i,t})$

Here, α denotes learning rates for the first and second stage. Notably, $r_{1,t} = 0$ and $\mathrm{Q_{MF}}(\mathrm{s_{3,t}}, \mathrm{a_{3,t}}) = 0$ because no reward is delivered after a first-stage choice and $Q_{MF}(s_{3,t}, a_{3,t}) = 0\mathrm{r_{1,t}} = 0$ because the task has only two states. Further, we allow for an additional stage-skipping update of first-stage values by introducing the parameter λ. As part of the model-free algorithm, this parameter connects the two stages in way that reward prediction errors at the second stage can influence first-stage values:

[3] $Q_{MF}(s_{1,t}, a_{1,t}) = Q_{MF}(s_{1,t}, a_{1,t}) + \alpha_1 \lambda \delta_{2,t}$

Importantly, λ accounts for the main effect of reward as observed in the analysis of first-stage stay-switch behavior but not for an interaction of reward and state. Notably, a model-free temporal-difference algorithm, here SARSA (λ), could acquire the state transition given enough time and stationary reward probabilities at the second stage. In the applied task, second-stage rewards probabilities changed slowly and independently according Gaussian random walks, which were identical to Daw et al. (2011).

Second, the model-based algorithm learns values prospectively and computes first-stage values by multiplying the maximum values at the second stage with the explicitly instructed transition probabilities:

[4] $Q_{MB}(s_A, a_j) = \quad P(S_B|S_A, a_j)\ max\ Q_{MF}(s_B, a) + P(S_C|S_A, a_j)\ max\ Q_{MF}(s_c, a)$

In equation 4, second-stage values come from the model-free algorithm because the model-based algorithm converges with the model-free algorithm at the second stage. Note that this approach simplifies transition learning as this algorithm does not learn transition probabilities incrementally but this is in line with the task instructions and the training. In a simulation by Daw et al. (2011) it could be shown that this approach outperforms incremental learning of the transition probabilities; moreover, this was identically applied in other non-clinical (Wunderlich et al., 2012b; Deserno et al., 2015b) and clinical studies (Voon et al., 2015).

Third, the hybrid algorithm connects model-free and model-based values via the weighting parameter ω:

[5] $Q(s_A, a_j) = \omega Q_{MB}(s_A, a_j) + (1-\omega) Q_{MF}(s_A, a_j)$

Importantly, ω gives a weighing of the relative influence of model-free and model-based values. It represents the balance of the two control strategies.

Finally, we transform values into action probabilities using a softmax for values Q:

[6] $p(a_{i,t} = a | s_{i,t}) = \frac{exp(\beta_i [Q(s_{i,t},a) + \rho * rep(a)])}{\sum_{a'} \exp(\beta_i [Q((s_{i,t},a')) + \rho * rep(a')])}$

Here, β controls the stochasticity of the choices at the first and second stage separately. The additional parameter ρ captures first-stage choice perseveration and rep is an indicator function that equals 1 if the previous first-stage choice was the same (Daw et al., 2011).

Model fitting. Constrained parameters were transformed to a logistic (α, λ, ω) or exponential (β) distribution to enforce constraints and to render normally distributed parameter estimates. To infer the maximum-a-posteriori estimate MAP of parameters θ, we use a Gaussian prior with mean and variance μ and σ:

[7] $MAP_i = \text{argmax} \log p(Y | \theta)\ p(\theta | \mu, \sigma)$

where Y represents the data in terms of actions Ai per subject i. We set priors empirically to the maximum likelihood estimates ML of μ and σ given the data by all subjects:

[8] $ML_i = \text{argmax} \log p(Y | \theta)$

and achieve this by using Expectation-Maximisation. For an in-depth description please compare Huys et al., 2011; Huys et al., 2012. Inferred parameters were distributed similarly as observed in previous studies with the same task (e.g. Daw et al., 2011; Wunderlich et al., 2012b; Deserno et al., 2015b; Radenbach et al., 2015; Voon et al., 2015, S-Table 3-2). All modeling analyses were performed using Matlab 2010b. Code of the analyses is available from the authors upon request.

Model comparison. For all three models, we first report the negative log-likelihood and the Bayesian Information Criterion (BIC) based on the negative log-likelihood, S-Table 3-1). Second, we approximate the model evidence by integrating out the free parameters. The integral was approximated by sampling from the prior distribution and we therefore add the subscript 'int' to the BIC (S-Table 3-1; compare Huys et al., 2011; Huys et al., 2012). Third and reported in the main manuscript, we submit this integrated likelihood to a random-effects Bayesian model selection procedure (Stephan et al., 2009, spm_BMS contained in SPM8). We also show that best-fitting parameters nicely reproduce the observed behavior.

Relationship of the parameters ω *and* λ. In parallel to the analysis of first-state choice data, the parameter λ resembles the main effect of reward on first-stage behavior while ω relates to the interaction of reward and state. However, relatively low levels of ω could either result from a reduced influence of first-stage model-based values or from a stronger weighting of first-stage model-free values. Thus at certain levels of λ, ω will decrease and vice versa. Given relatively midrange levels for both parameters in a sample, no correlation would be expected. First, as we have previously published independent samples with this task including the identical modeling analysis, we did check the correlation of both parameters in all three samples (Deserno et al.,

2015b; Radenbach et al., 2015, for the former from the control condition only): Deserno et al. (2015): r=-.09, p=.65, mean ω=.53±.18 (SD), λ .71±.09; Radenbach et al. (2015): r=-.41, p=.01, mean ω=.68±.08, λ=.57±.11; presented study: r=-.34, p=.02, ω=.65±.09, λ=.60±.11; These correlations obtained across three independent samples demonstrate that a consistent correlation between the two parameters cannot be assumed. Importantly, the correlation in the present sample was not driven by one of the two groups (low-impulsivity r=-.40, p=.05, mean ω=.59±.11, λ=.62±.10, high-impulsivity r=-.39, p=.06, mean ω=.61±.11, λ=.69±.08). Given this relatively low to moderate correlations, we conclude that a change in one of the two does not simply imply a change in the other, which would in fact render an additional parameter redundant. This is also supported by the observation that the parameter ω can be well re-fitted from generated data based on the inferred parameters (please compare Deserno et al., 2015b). In the same vein, we were asked whether fitting λ as a free parameter could have concealed a group difference on ω. In fact, finding a difference on ω when fixing λ would be surprising because the raw data does not support a change in the overall balance of model-free and model-based control (while the reward x impulsivity in raw data supports an effect of impulsivity on λ). Thus, such an effect due to fixing λ would speak for redundancy of the two parameters, which is, as pointed out, not the case. In line with this reasoning, the suggested analysis did not reveal any difference in ω when keeping λ fixed to the sample mean of .65 and fitting the model (ω high-impulsive .61±.11, low-impulsive .60±.11, $T(1,48)=.38$, p=.70).

S-Table 3-1. Model Selection. -LL: negative log-likelihood; BIC: Bayesian Information Criterion, the subscript int refers to integrating out the free parameters; XP: Exceedance Probability; all=all participants n=50, high=high-impulsive n=24, low=low-impulsive n=26.

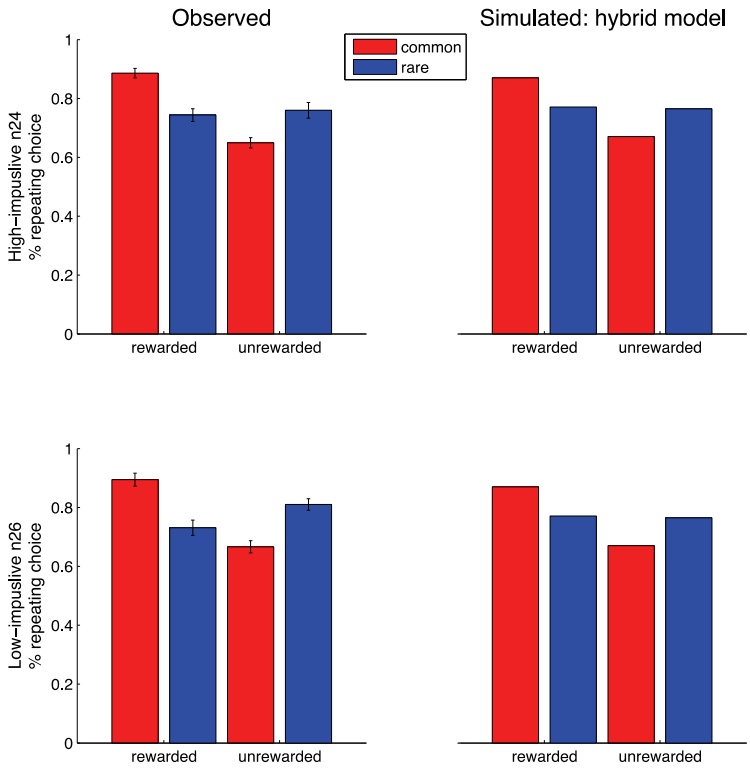
		-LL	BIC	BIC_{int}	XP
hybrid with λ	all	9261.94	18593.14	19326.75	.9999
	high	4578.76	9221.65	9560.81	.9950
	low	4683.18	9431.04	9825.49	.9997
		Δ -LL hybrid	Δ BIC hybrid	Δ BIC_{int} hybrid	
hybrid without λ	all	-151.14	292.39	265.54	0
	high	-83.09	157.01	147.86	3e-04
	low	-68.05	126.87	109.17	1e-05
model-based	all	-260.38	491.07	325.59	0
	high	-114.20	200.92	128.38	9e-05
	low	-146.18	264.63	171.68	3e-05
model-free with λ	all	-304.58	599.27	468.23	1e-05
	high	-120.77	232.37	183.13	.0050
	low	-183.81	358.38	276.60	.0003
model-free without λ	all	-549.17	1078.54	905.82	0
	high	-247.29	476.26	403.75	0
	low	-301.88	585.27	485.05	1e-05

S-Table 3-2. Distribution of best-fitting parameters. Hybrid Model. ω: weighting of model-free and model-based values; α1,α1: learning rates at the first and second stage; λ: stage-skipping update; Softmax Observation Model. β1,β2: stochasticity of first- and second-stage choices; -LL: negative log-likelihood; all=all participants n50, high=high-impulsive n=24, low=low-impulsive n=26.

		ω	α_1	α_2	λ	β_1	β_2	ρ	-LL
25th percentile	All	.55	.36	.49	.60	4.28	2.21	.10	222.24
	High	.57	.41	.43	.63	3.76	2.15	.13	227.61
	Low	.53	.30	.55	.55	5.01	2.34	.08	206.64
Median	All	.60	.49	.62	.65	7.20	3.08	.14	179.47
	High	.62	.52	.63	.69	6.13	3.04	.15	191.16
	Low	.59	.40	.62	.64	7.62	3.21	.12	176.77
75th percentile	All	.68	.65	.71	.71	8.46	3.89	.18	141.66
	high	.70	.66	.74	.73	7.99	3.85	.18	149.60
	Low	.68	.60	.68	.68	9.28	3.89	.18	141.62

S-Table 3-3 fMRI results. Whole-brain results for the conjunction of model-free and model-based learning signals across both groups (n=48, 23 high-impulsive and 25 low-impulsive participants)

Region	coordinates	t-value	p-FWE peak-level	K	p-FWE cluster-level
Conjunction of model-free and model-based					
Medial PFC	0, 50, 0	4.83	.059	919	2.55e-05
	0, 44, 8	4.59	.125		
	-4, 30, 8	4.02	.545		
Ventral Striatum	-12, 12, -8	5.51	.005	1451	2.49e-07
	12, 8, -8	5.42	.007		
	-8, 12, 2	5.23	.014		
Lateral PFC/OFC	20, 28, -16	4.43	.200	250	0.036
	38, 24, -16	3.98	.587		
	32,42, -14	3.77	.796		
Lateral Parietal Cortex	58, -44, 38	4.85	.054	856	4.55e-05
	60, -56, 38	4.62	.114		
	58, -52, 20	4.15	.417		
	-40, -70, 42	4.11	.453	319	.014
	-40, -62, 26	4.02	.543		
	-44, -56, 30	4.01	.561		
Posterior Cingulate Cortex	12, -44, 30	5.48	.004	1564	1.24e-07
	-4, -42, 36	4.51	.161		
	8, -44, 30	4.51	.161		



S-Figure 3-1. Observed choice and simulated data based on inferred parameters. Upper and lower left panels display data of high- and low-impulsive groups both showing aspects of model-free and model-based control in first-stage stay-switch behavior. Upper and lower right panels show simulated data based on inferred parameters of the hybrid model demonstrating that the hybrid model nicely recovers the actually observed behavior.

References

Abe, H. & Lee, D. (2011) Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. *Neuron*, **70**, 731-741.

Adams, C.D. (1981) Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Q J Exp Psychol*, **34B**, 77-98.

Adams, C.D. & Dickinson, A. (1981) Instrumental responding following reinforcer devaluation. *Q J Exp Psychol*, **33B**, 109-121.

Akam, T., Costa, R. & Dayan, P. (2015) Simple plans or sophisticated habits? State, transition and learning Interactions in the Two-step Task. *bioRxiv*.

Alexander, W.H. & Brown, J.W. (2011) Medial prefrontal cortex as an action-outcome predictor. *Nature neuroscience*, **14**, 1338-1344.

Allen, J.P., Litten, R.Z., Fertig, J.B. & Babor, T. (1997) A review of research on the Alcohol Use Disorders Identification Test (AUDIT). *Alcohol Clin Exp Res*, **21**, 613-619.

Amthauer, R.B., B.; Liepmann, D.; Beauducel, A. (1999) *Intelligenz-Struktur-Test 2000*. Horgrefe, Göttingen, Germany.

Anselme, P., Robinson, M.J. & Berridge, K.C. (2013) Reward uncertainty enhances incentive salience attribution as sign-tracking. *Behav Brain Res*, **238**, 53-61.

Anton, R.F., Moak, D.H. & Latham, P. (1995) The Obsessive Compulsive Drinking Scale: a self-rated instrument for the quantification of thoughts about alcohol and drinking behavior. *Alcohol Clin Exp Res*, **19**, 92-99.

Ashburner, J. & Friston, K.J. (2005) Unified segmentation. *Neuroimage*, **26**, 839-851.

American Psychiatric Association (2013). *The Diagnostic and Statistical Manual of Mental Disorder: DSM 5*. American Psychiatric Association, Washington D.C.

Badre, D. (2008) Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends Cogn Sci*, **12**, 193-200.

Balleine, B.W., Daw, N.D. & O'Doherty J, P. (2008a) Multiple Forms of Value Learning and the Function of Dopamine. In Glimcher, P.W., Camerer, C., Fehr, E., Poldrack, R.A. (eds) *Neuroeconomics: decision making and the brain*,. Academic Press.

Balleine, B.W., Daw, N.D. & O'Doherty, J.P. (2008b) Multiple forms of value learning and the function of dopamine. *Neuroeconomics: decision making and the brain*, Academic Press, 367-385.

Balleine, B.W. & Dickinson, A. (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology*, **37**, 407-419.

Balleine, B.W. & O'Doherty, J.P. (2010) Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, **35**, 48-69.

Bates, M.E., Bowden, S.C. & Barry, D. (2002) Neurocognitive impairment associated with alcohol use disorders: implications for treatment. *Experimental and clinical psychopharmacology*, **10**, 193-212.

Bechara, A. & Damasio, H. (2002) Decision-making and addiction (part I): impaired activation of somatic states in substance dependent individuals when pondering decisions with negative future consequences. *Neuropsychologia*, **40**, 1675-1689.

Bechara, A., Damasio, H., Tranel, D. & Damasio, A.R. (1997) Deciding advantageously before knowing the advantageous strategy. *Science*, **275**, 1293-1295.

Bechara, A., Tranel, D., Damasio, H. & Damasio, A.R. (1996) Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cerebral cortex*, **6**, 215-225.

Beck, A., Schlagenhauf, F., Wustenberg, T., Hein, J., Kienast, T., Kahnt, T., Schmack, K., Hagele, C., Knutson, B., Heinz, A. & Wrase, J. (2009) Ventral striatal activation during reward anticipation correlates with impulsivity in alcoholics. *Biol Psychiatry*, **66**, 734-742.

Beck, A., Wustenberg, T., Genauck, A., Wrase, J., Schlagenhauf, F., Smolka, M.N., Mann, K. & Heinz, A. (2012) Effect of brain structure, brain function, and brain connectivity on relapse in alcohol-dependent patients. *Arch Gen Psychiatry*, **69**, 842-852.

Becker, M.P., Nitsch, A.M., Miltner, W.H. & Straube, T. (2014) A single-trial estimation of the feedback-related negativity and its relation to BOLD responses in a time-estimation task. *J Neurosci*, **34**, 3005-3012.

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

Belin, D., Belin-Rauscent, A., Murray, J.E. & Everitt, B.J. (2013) Addiction: failure of control over maladaptive incentive habits. *Curr Opin Neurobiol*, **23**, 564-572.

Belin, D., Mar, A.C., Dalley, J.W., Robbins, T.W. & Everitt, B.J. (2008) High impulsivity predicts the switch to compulsive cocaine-taking. *Science*, **320**, 1352-1355.

Bellman, R.E. (1957) Dynamic Programming. Princeton University Press, Princeton, NJ.

Beveridge, T.J., Gill, K.E., Hanlon, C.A. & Porrino, L.J. (2008) Review. Parallel studies of cocaine-related neural and cognitive impairment in humans and monkeys. *Philos Trans R Soc Lond B Biol Sci*, **363**, 3257-3266.

Boorman, E.D., Behrens, T.E. & Rushworth, M.F. (2011) Counterfactual choice and learning in a neural network centered on human lateral frontopolar cortex. *PLoS Biol*, **9**, e1001093.

Boorman, E.D., Behrens, T.E., Woolrich, M.W. & Rushworth, M.F. (2009) How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron*, **62**, 733-743.

Braver, T.S. & Cohen, J.D. (1999) Dopamine, cognitive control, and schizophrenia: the gating model. *Prog Brain Res*, **121**, 327-349.

Brodersen, K.H., Deserno, L., Schlagenhauf, F., Lin, Z., Penny, W.D., Buhmann, J.M. & Stephan, K.E. (2014) Dissecting psychiatric spectrum disorders by generative embedding. *NeuroImage: Clinical*, **4**, 98-111.

Bromberg-Martin, E.S., Matsumoto, M., Hong, S. & Hikosaka, O. (2010) A pallidus-habenula-dopamine pathway signals inferred stimulus values. *J Neurophysiol*, **104**, 1068-1076.

Buckholtz, J.W. & Meyer-Lindenberg, A. (2012) Psychopathology and the human connectome: toward a transdiagnostic model of risk for mental illness. *Neuron*, **74**, 990-1004.

Buckholtz, J.W., Treadway, M.T., Cowan, R.L., Woodward, N.D., Li, R., Ansari, M.S., Baldwin, R.M., Schwartzman, A.N., Shelby, E.S., Smith, C.E., Kessler, R.M. & Zald, D.H. (2010) Dopaminergic network differences in human impulsivity. *Science*, **329**, 532.

Buxton, R.B. (2009) *Introduction to functional magnetic resonance imaging: principles and techniques*. Cambridge university press.

Callicott, J.H., Mattay, V.S., Verchinski, B.A., Marenco, S., Egan, M.F. & Weinberger, D.R. (2003) Complexity of prefrontal cortical dysfunction in schizophrenia: more than up or down. *Am J Psychiatry*, **160**, 2209-2215.

Campbell, J. & Ehlert, U. (2012) Acute psychosocial stress: does the emotional stress response correspond with physiological responses? *Psychoneuroendocrinology*, **37**, 1111-1134.

Casey, K.F., Benkelfat, C., Cherkasova, M.V., Baker, G.B., Dagher, A. & Leyton, M. (2014) Reduced dopamine response to amphetamine in subjects at ultra-high risk for addiction. *Biological psychiatry*, **76**, 23-30.

Cavagnaro, D.R., Gonzalez, R., Myung, J.I. & Pitt, M.A. (2013) Optimal Decision Stimuli for Risky Choice Experiments: An Adaptive Approach. *Management science*, **59**, 358-375.

Cavanagh, J.F. & Frank, M.J. (2013) Stop! Stay tuned for more information. *Exp Neurol*, **247**, 289-291.

Cavanagh, J.F. & Frank, M.J. (2014) Frontal theta as a mechanism for cognitive control. *Trends in cognitive sciences*, **18**, 414-421.

Cavanagh, J.F., Frank, M.J., Klein, T.J. & Allen, J.J. (2010) Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *Neuroimage*, **49**, 3198-3209.

Chase, H.W., Swainson, R., Durham, L., Benham, L. & Cools, R. (2011) Feedback-related negativity codes prediction error but not behavioral adjustment during probabilistic reversal learning. *J Cogn Neurosci*, **23**, 936-946.

Chiu, P.H., Lohrenz, T.M. & Montague, P.R. (2008) Smokers' brains compute, but ignore, a fictive error signal in a sequential investment task. *Nat Neurosci*, **11**, 514-520.

Chowdhury, R., Guitart-Masip, M., Lambert, C., Dayan, P., Huys, Q., Duzel, E. & Dolan, R.J. (2013) Dopamine restores reward prediction errors in old age. *Nat Neurosci*, **16**, 648-653.

Chumbley, J.R., Flandin, G., Bach, D.R., Daunizeau, J., Fehr, E., Dolan, R.J. & Friston, K.J. (2012) Learning and generalization under ambiguity: an fMRI study. *PLoS Comput Biol*, **8**, e1002346.

Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*. Erlbaum, Hillsdale, NJ.

Cohen, J.D., McClure, S.M. & Yu, A.J. (2007) Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philos Trans R Soc Lond B Biol Sci*, **362**, 933-942.

Cohen, M.X., Cavanagh, J.F. & Slagter, H.A. (2011a) Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: temporospatial principal components analysis and source localization of the feedback negativity: commentary. *Hum Brain Mapp*, **32**, 2270-2271.

Cohen, M.X., Wilmes, K.A. & van de Vijver, I. (2011b) Cortical electrophysiological network dynamics of feedback learning. *Trends in cognitive sciences*, **15**, 558-566.

Cohen, S., Kamarck, T. & Mermelstein, R. (1983) A global measure of perceived stress. *Journal of health and social behavior*, **24**, 385-396.

Collins, A.G. & Frank, M.J. (2012) How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *Eur J Neurosci*, **35**, 1024-1035.

Coltheart, M. (2006) What has functional neuroimaging told us about the mind (so far)? *Cortex*, **42**, 323-331.

Coltheart, M. (2013) How Can Functional Neuroimaging Inform Cognitive Theories? *Perspectives on psychological science: a journal of the Association for Psychological Science*, **8**, 98-103.

Cools, R. (2011) Dopaminergic control of the striatum for high-level cognition. *Curr Opin Neurobiol*, **21**, 402-407.

Cools, R., Blackwell, A., Clark, L., Menzies, L., Cox, S. & Robbins, T.W. (2005) Tryptophan depletion disrupts the motivational guidance of goal-directed behavior as a function of trait impulsivity. *Neuropsychopharmacology*, **30**, 1362-1373.

Cools, R., Clark, L., Owen, A.M. & Robbins, T.W. (2002) Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *J Neurosci*, **22**, 4563-4567.

Corbit, L.H. & Balleine, B.W. (2003) The role of prelimbic cortex in instrumental conditioning. *Behav Brain Res*, **146**, 145-157.

Corrigan, P.W., Kuwabara, S.A. & O'Shaughnessy, J. (2009) The public stigma of mental illness and drug addiction findings from a stratified random sample. *Journal of Social Work*, **9**, 139-147.

Critchley, D.R. (2005) Genetic, biochemical and structural approaches to talin function. *Biochem Soc T*, **33**, 1308-1312.

Dalbert, C. (1992) ASTS - Aktuelle Stimmungsskala, http://www.erzwiss.uni-halle.de/gliederung/paed/ppsych/sdasts.pdf.

Dalley, J.W., Everitt, B.J. & Robbins, T.W. (2011) Impulsivity, compulsivity, and top-down cognitive control. *Neuron*, **69**, 680-694.

Dalley, J.W., Fryer, T.D., Brichard, L., Robinson, E.S., Theobald, D.E., Laane, K., Pena, Y., Murphy, E.R., Shah, Y., Probst, K., Abakumova, I., Aigbirhio, F.I., Richards, H.K., Hong, Y., Baron, J.C., Everitt, B.J. & Robbins, T.W. (2007) Nucleus accumbens D2/3 receptors predict trait impulsivity and cocaine reinforcement. *Science*, **315**, 1267-1270.

Damasio, A.R. (1999) *The feeling of what happens: Body and Emotion in the Making of Consciousness*. Harcourt Incorporated, New York.

Danner, U.N., Ouwehand, C., van Haastert, N.L., Hornsveld, H. & de Ridder, D.T. (2012) Decision-making impairments in women with binge eating disorder in comparison with obese and normal weight women. *European eating disorders review : the journal of the Eating Disorders Association*, **20**, e56-62.

Davis, C., Patte, K., Curtis, C. & Reid, C. (2010) Immediate pleasures and future consequences. A neuropsychological study of binge eating and obesity. *Appetite*, **54**, 208-213.

Daw, N.D. (2009) Trial-by-trial data analysis using computational models. In Phelps, E.A., Robbins, T.W., Delgado, M. (eds) *In Affect, Learning and Decision Making, Attention and Performance XXIII*. Oxford University Press, New York, pp. 3-38.

Daw, N.D. (2011) Trial-by-trial data analysis using computational models. *Decision making, affect, and learning: Attention and performance XXIII*, **23**, 1.

Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P. & Dolan, R.J. (2011) Model-based influences on humans' choices and striatal prediction errors. *Neuron*, **69**, 1204-1215.

Daw, N.D., Niv, Y. & Dayan, P. (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci*, **8**, 1704-1711.

Daw, N.D., O'Doherty, J.P., Dayan, P., Seymour, B. & Dolan, R.J. (2006) Cortical substrates for exploratory decisions in humans. *Nature*, **441**, 876-879.

Daw, N.D. & O'Doherty, J.P. (2013) Multiple systems for value learning. *Neuroeconomics: Decision Making, and the Brain*.

Dayan, P. (2009a) Dopamine, reinforcement learning, and addiction. *Pharmacopsychiatry*, **42 Suppl 1**, S56-65.

Dayan, P. (2009b) Goal-directed control and its antipodes. *Neural Netw*, **22**, 213-219.

Dayan, P. & Daw, N.D. (2008) Decision theory, reinforcement learning, and the brain. *Cogn Affect Behav Neurosci*, **8**, 429-453.

Dayan, P. & Niv, Y. (2008) Reinforcement learning: the good, the bad and the ugly. *Curr Opin Neurobiol*, **18**, 185-196.

Dayan, P., Niv, Y., Seymour, B. & Daw, N.D. (2006) The misbehavior of value and the discipline of the will. *Neural Netw*, **19**, 1153-1160.

de Lissa, P., Sorensen, S., Badcock, N., Thie, J. & McArthur, G. (2015) Measuring the face-sensitive N170 with a gaming EEG system: A validation study. *J Neurosci Methods*, **253**, 47-54.

de Wit, S., Corlett, P.R., Aitken, M.R., Dickinson, A. & Fletcher, P.C. (2009) Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. *J Neurosci*, **29**, 11330-11338.

de Wit, S., Standing, H.R., Devito, E.E., Robinson, O.J., Ridderinkhof, K.R., Robbins, T.W. & Sahakian, B.J. (2012a) Reliance on habits at the expense of goal-directed control following dopamine precursor depletion. *Psychopharmacology (Berl)*, **219**, 621-631.

de Wit, S., Watson, P., Harsay, H.A., Cohen, M.X., van de Vijver, I. & Ridderinkhof, K.R. (2012b) Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. *J Neurosci*, **32**, 12066-12075.

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

den Ouden, H.E., Daw, N.D., Fernandez, G., Elshout, J.A., Rijpkema, M., Hoogman, M., Franke, B. & Cools, R. (2013) Dissociable effects of dopamine and serotonin on reversal learning. *Neuron*, **80**, 1090-1100.

Deserno, L., Beck, A., Huys, Q.J., Lorenz, R.C., Buchert, R., Buchholz, H., Plotkin, M., Kumakara, Y., Cumming, P., Heinze, H., Grace, A.A., Rapp, M.A., Schlagenhauf, F. & Heinz, A. (2014) Chronic alcohol intake abolishes the relationship between dopamine synthesis capacity and learning signals in the ventral striatum. *Eur J Neurosci*.

Deserno, L., Beck, A., Huys, Q.J., Lorenz, R.C., Buchert, R., Buchholz, H.G., Plotkin, M., Kumakara, Y., Cumming, P., Heinze, H.J., Grace, A.A., Rapp, M.A., Schlagenhauf, F. & Heinz, A. (2015a) Chronic alcohol intake abolishes the relationship between dopamine synthesis capacity and learning signals in the ventral striatum. *Eur J Neurosci*, **41**, 477-486.

Deserno, L., Boehme, R., Heinz, A. & Schlagenhauf, F. (2013) Reinforcement learning and dopamine in schizophrenia: dimensions of symptoms or specific features of a disease group? *Front Psychiatry*, **4**, 172.

Deserno, L., Huys, Q., Boehme, R., Buchert, R., Heinze, H.J., Grace, A.A., Dolan, R.J., Heinz, A. & Schlagenhauf, F. (2015b) Ventral striatal presynaptic dopamine reflects behavioral and neural signatures of model-based control during sequential decision-making. *Proc Natl Acad Sci U S A*.

Deserno, L., Wilbertz, T., Reiter, A.M.F., Horstmann, A., Neumann, J., Villringer, A., Heinze, H.J. & Schlagenhauf, F. (in press) Lateral prefrontal model-based signals are reduced in healthy individuals with high trait impulsivity. *Transl Psychiatry*.

Dias-Ferreira, E., Sousa, J.C., Melo, I., Morgado, P., Mesquita, A.R., Cerqueira, J.J., Costa, R.M. & Sousa, N. (2009) Chronic stress causes frontostriatal reorganization and affects decision-making. *Science*, **325**, 621-625.

Dickinson, A. & Balleine, B. (2002) The role of learning in the operation of motivational systems. In Gallistel, R. (ed) *Stevens' handbook of experimental psychology*. Wiley, pp. 497-534.

Dickinson, A.D. (1985) Action and Habits: The Development of Behavioural Autonomy. *Philos Trans R Soc Lond B Biol Sci*, **308**, 67-78.

Diener, E. (2010) Neuroimaging: Voodoo, New Phrenology, or Scientific Breakthrough? Introduction to Special Section on fMRI. *Perspectives on psychological science : a journal of the Association for Psychological Science*, **5**, 714-715.

Dietrich, A., Federbusch, M.G., Grellmann, C., Villringer, A. & Horstmann, A. (2014) Body weight status, eating behavior, sensitivity to reward/punishment, and gender: relationships and interdependencies. *Frontiers Psychology*, **5**, 1073.

Dingemans, A.E. & van Furth, E.F. (2012) Binge eating disorder psychopathology in normal weight and obese individuals. *International Journal of Eating Disorders*, **45**, 135-138.

Dolan, R.J. (2008) Neuroimaging of cognition: past, present, and future. *Neuron*, **60**, 496-502.

Dolan, R.J. & Dayan, P. (2013) Goals and habits in the brain. *Neuron*, **80**, 312-325.

Doll, B.B., Duncan, K.D., Simon, D.A., Shohamy, D. & Daw, N.D. (2015) Model-based choices involve prospective neural activity. *Nat Neurosci*, **18**, 767-772.

Doll, B.B., Simon, D.A. & Daw, N.D. (2012) The ubiquity of model-based reinforcement learning. *Curr Opin Neurobiol*, **22**, 1075-1081.

Doya, K., Samejima, K., Katagiri, K. & Kawato, M. (2002) Multiple model-based reinforcement learning. *Neural Comput*, **14**, 1347-1369.

Eppinger, B., Walter, M., Heekeren, H.R. & Li, S.C. (2013) Of goals and habits: age-related and individual differences in goal-directed decision-making. *Front Neurosci*, **7**, 253.

Epstein, D.H. & Shaham, Y. (2010) Cheesecake-eating rats and the question of food addiction. *Nat Neurosci*, **13**, 529-531.

Ersche, K.D., Jones, P.S., Williams, G.B., Smith, D.G., Bullmore, E.T. & Robbins, T.W. (2013) Distinctive personality traits and neural correlates associated with stimulant drug use versus familial risk of stimulant dependence. *Biol Psychiatry*, **74**, 137-144.

Ersche, K.D., Jones, P.S., Williams, G.B., Turton, A.J., Robbins, T.W. & Bullmore, E.T. (2012a) Abnormal brain structure implicated in stimulant drug addiction. *Science*, **335**, 601-604.

Ersche, K.D., Roiser, J.P., Abbott, S., Craig, K.J., Muller, U., Suckling, J., Ooi, C., Shabbir, S.S., Clark, L., Sahakian, B.J., Fineberg, N.A., Merlo-Pich, E.V., Robbins, T.W. & Bullmore, E.T. (2011) Response perseveration in stimulant dependence is associated with striatal dysfunction and can be ameliorated by a D(2/3) receptor agonist. *Biol Psychiatry*, **70**, 754-762.

Ersche, K.D., Roiser, J.P., Robbins, T.W. & Sahakian, B.J. (2008) Chronic cocaine but not chronic amphetamine use is associated with perseverative responding in humans. *Psychopharmacology (Berl)*, **197**, 421-431.

Ersche, K.D., Turton, A.J., Chamberlain, S.R., Muller, U., Bullmore, E.T. & Robbins, T.W. (2012b) Cognitive dysfunction and anxious-impulsive personality traits are endophenotypes for drug dependence. *Am J Psychiatry*, **169**, 926-936.

Ersche, K.D., Turton, A.J., Pradhan, S., Bullmore, E.T. & Robbins, T.W. (2010) Drug addiction endophenotypes: impulsive versus sensation-seeking personality traits. *Biol Psychiatry*, **68**, 770-773.

Etkin, A. & Wager, T.D. (2007) Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *Am J Psychiatry*, **164**, 1476-1488.

Everitt, B.J., Belin, D., Economidou, D., Pelloux, Y., Dalley, J.W. & Robbins, T.W. (2008) Review. Neural mechanisms underlying the vulnerability to develop compulsive drug-seeking habits and addiction. *Philos Trans R Soc Lond B Biol Sci*, **363**, 3125-3135.

Everitt, B.J. & Robbins, T.W. (2005) Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat Neurosci*, **8**, 1481-1489.

Farr, O.M., Hu, S., Zhang, S. & Li, C.S. (2012) Decreased saliency processing as a neural measure of Barratt impulsivity in healthy adults. *Neuroimage*, **63**, 1070-1077.

Farrell, S. & Lewandowsky, S. (2010) Computational Models as Aids to Better Reasoning in Psychology. *Current Directions in Psychological Science*, **19**, 329-335.

Farrell, S. & Lewandowsky, S. (2015) An Introduction to Cognitive Modeling. In Forstmann, B.U., Wagenmakers, E.J. (eds) *Model-Based Cognitive Neuroscience*. Springer, New York, pp. 3-24.

Featherstone, R.E. & McDonald, R.J. (2004) Dorsal striatum and stimulus-response learning: lesions of the dorsolateral, but not dorsomedial, striatum impair acquisition of a simple discrimination task. *Behav Brain Res*, **150**, 15-23.

First, M.B., Spitzer, R.L., Gibbon, M. & Williams, J. (2001) *Structured Clinical interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition With Psychotic Screen (SCID-I/P W/ PSY SCREEN)*. New York State Psychiatric Institute New York.

First, M.B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (1997) *User's guide for the Structured clinical interview for DSM-IV axis I disorders SCID-I: clinician version.* Biometrics Research Dept, New York State Psychiatric Institute, New York, N.Y.

Fischer, A.G. & Ullsperger, M. (2013) Real and fictive outcomes are processed differently but converge on a common adaptive mechanism. *Neuron*, **79**, 1243-1255.

Forstmann, B.U., Dutilh, G., Brown, S., Neumann, J., von Cramon, D.Y., Ridderinkhof, K.R. & Wagenmakers, E.J. (2008) Striatum and pre-SMA facilitate decision-making under time pressure. *Proc Natl Acad Sci U S A*, **105**, 17538-17542.

Forstmann, B.U. & Wagenmakers, E.J. (2015) Model-Based Cognitive Neuroscience: A Conceptual Introduction *An Introduction to Model-Based Cognitive Neuroscience*. Springer, New York, pp. 139-156.

Forstmann, B.U., Wagenmakers, E.J., Eichele, T., Brown, S. & Serences, J.T. (2011) Reciprocal relations between cognitive neuroscience and formal cognitive models: opposites attract? *Trends Cogn Sci*, **15**, 272-279.

Foti, D., Weinberg, A., Dien, J. & Hajcak, G. (2011a) Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: response to commentary. *Hum Brain Mapp*, **32**, 2267-2269.

Foti, D., Weinberg, A., Dien, J. & Hajcak, G. (2011b) Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: temporospatial principal components analysis and source localization of the feedback negativity. *Hum Brain Mapp*, **32**, 2207-2216.

Frank, M.J., Doll, B.B., Oas-Terpstra, J. & Moreno, F. (2009) Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nat Neurosci*, **12**, 1062-1068.

Frank, M.J., Seeberger, L.C. & O'Reilly R, C. (2004) By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science*, **306**, 1940-1943.

Freedman, R., Lewis, D.A., Michels, R., Pine, D.S., Schultz, S.K., Tamminga, C.A., Gabbard, G.O., Gau, S.S., Javitt, D.C., Oquendo, M.A., Shrout, P.E., Vieta, E. & Yager, J. (2013) The initial field trials of DSM-5: new blooms and old thorns. *Am J Psychiatry*, **170**, 1-5.

Friedel, E., Koch, S.P., Wendt, J., Heinz, A., Deserno, L. & Schlagenhauf, F. (2014) Devaluation and sequential decisions: linking goal-directed and model-based behavior. *Front Hum Neurosci*, **8**, 587.

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J. & Penny, W. (2007) Variational free energy and the Laplace approximation. *Neuroimage*, **34**, 220-234.

Friston, K.J. (2009) Modalities, modes, and models in functional neuroimaging. *Science*, **326**, 399-403.

Garavan, H. & Stout, J.C. (2005) Neurocognitive insights into substance abuse. *Trends Cogn Sci*, **9**, 195-201.

Garbusow, M., Schad, D.J., Sebold, M., Friedel, E., Bernhardt, N., Koch, S.P., Steinacher, B., Kathmann, N., Geurts, D.E. & Sommer, C. (2015) Pavlovian-to-instrumental transfer effects in the nucleus accumbens relate to relapse in alcohol dependence. *Addiction biology*.

Gehring, W.J. & Willoughby, A.R. (2002) The medial frontal cortex and the rapid processing of monetary gains and losses. *Science*, **295**, 2279-2282.

Gershman, S.J., Markman, A.B. & Otto, A.R. (2014) Retrospective revaluation in sequential decision making: a tale of two systems. *Journal of experimental psychology. General*, **143**, 182-194.

Gillan, C.M., Otto, A.R., Phelps, E.A. & Daw, N.D. (2015) Model-based learning protects against forming habits. *Cogn Affect Behav Neurosci*.

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

Glascher, J., Hampton, A.N. & O'Doherty, J.P. (2009) Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cereb Cortex*, **19**, 483-495.

Gluck, M.E. (2006) Stress response and binge eating disorder. *Appetite*, **46**, 26-30.

Gluck, M.E., Geliebter, A., Hung, J. & Yahav, E. (2004) Cortisol, hunger, and desire to binge eat following a cold stress test in obese women with binge eating disorder. *Psychosomatic medicine*, **66**, 876-881.

Goldstein, R.Z., Craig, A.D., Bechara, A., Garavan, H., Childress, A.R., Paulus, M.P. & Volkow, N.D. (2009) The neurocircuitry of impaired insight in drug addiction. *Trends Cogn Sci*, **13**, 372-380.

Goldstein, R.Z., Leskovjan, A.C., Hoff, A.L., Hitzemann, R., Bashan, F., Khalsa, S.S., Wang, G.J., Fowler, J.S. & Volkow, N.D. (2004) Severity of neuropsychological impairment in cocaine and alcohol addiction: association with metabolism in the prefrontal cortex. *Neuropsychologia*, **42**, 1447-1458.

Goldstein, R.Z. & Volkow, N.D. (2011) Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nat Rev Neurosci*, **12**, 652-669.

Goto, Y. & Grace, A.A. (2005) Dopaminergic modulation of limbic and cortical drive of nucleus accumbens in goal-directed behavior. *Nat Neurosci*, **8**, 805-812.

Gottesman, II & Shields, J. (1973) Genetic theorizing and schizophrenia. *Br J Psychiatry*, **122**, 15-30.

Grant, B.F., Stinson, F.S., Dawson, D.A., Chou, S.P., Dufour, M.C., Compton, W., Pickering, R.P. & Kaplan, K. (2004) Prevalence and co-occurrence of substance use disorders and independentmood and anxiety disorders: Results from the national epidemiologic survey on alcohol and relatedconditions. *Archives of general psychiatry*, **61**, 807-816.

Grawe, K. (2005) Empirisch validierte Wirkfaktoren statt Therapiemethoden. *Report Psychologie*, **7**, 311.

Grawe, K., Caspar, F. & Ambuhl, H. (1990) Die Berner Therapievergleichsstudie: Wirkungsvergleich und differentielle Indikation. *Zeitschrift für Klinische Psychologie*, **19**, 338-361.

Grinband, J., Hirsch, J. & Ferrera, V.P. (2006) A neural representation of categorization uncertainty in the human brain. *Neuron*, **49**, 757-763.

Gruner, P., Anticevic, A., Lee, D. & Pittenger, C. (2015) Arbitration between action strategies in obsessive-compulsive disorder. *The Neuroscientist*, 1073858414568317.

Haber, S.N. & Behrens, T.E. (2014) The neural network underlying incentive-based learning: implications for interpreting circuit disruptions in psychiatric disorders. *Neuron*, **83**, 1019-1039.

Hampton, A.N., Bossaerts, P. & O'Doherty, J.P. (2006) The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J Neurosci*, **26**, 8360-8367.

Hare, T. (2014) Neuroscience. Exploiting and exploring the options. *Science*, **344**, 1446-1447.

Hauser, T.U., Iannaccone, R., Ball, J., Mathys, C., Brandeis, D., Walitza, S. & Brem, S. (2014a) Role of the medial prefrontal cortex in impaired decision making in juvenile attention-deficit/hyperactivity disorder. *JAMA Psychiatry*, **71**, 1165-1173.

Hauser, T.U., Iannaccone, R., Stampfli, P., Drechsler, R., Brandeis, D., Walitza, S. & Brem, S. (2014b) The feedback-related negativity (FRN) revisited: new insights into the localization, meaning and network organization. *Neuroimage*, **84**, 159-168.

Hawkins, J.D., Graham, J.W., Maguin, E., Abbott, R., Hill, K.G. & Catalano, R.F. (1997) Exploring the effects of age of alcohol use initiation and psychosocial risk factors on subsequent alcohol misuse. *Journal of studies on alcohol*, **58**, 280.

Hayes, A.F. & Matthes, J. (2009) Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. *Behav Res Methods*, **41**, 924-936.

Heathcote, A., Brown, S.D. & Wagenmakers, E.-J. (2015) An introduction to good practices in cognitive modeling *An Introduction to Model-Based Cognitive Neuroscience*. Springer, pp. 25-48.

Heinz, A. (2002) Dopaminergic dysfunction in alcoholism and schizophrenia--psychopathological and behavioral correlates. *Eur Psychiatry*, **17**, 9-16.

Heinz, A. & Batra, A. (2003) *Neurobiologie der Alkohol-und Nikotinabhängigkeit*. Kohlhammer.

Heinz, A., Siessmeier, T., Wrase, J., Buchholz, H.G., Grunder, G., Kumakura, Y., Cumming, P., Schreckenberger, M., Smolka, M.N., Rosch, F., Mann, K. & Bartenstein, P. (2005) Correlation of alcohol craving with striatal dopamine synthesis capacity and D2/3 receptor availability: a combined [18F]DOPA and [18F]DMFP PET study in detoxified alcoholic patients. *Am J Psychiatry*, **162**, 1515-1520.

Heinz, A., Siessmeier, T., Wrase, J., Hermann, D., Klein, S., Grusser, S.M., Flor, H., Braus, D.F., Buchholz, H.G., Grunder, G., Schreckenberger, M., Smolka, M.N., Rosch, F., Mann, K. & Bartenstein, P. (2004) Correlation between dopamine D(2) receptors in the ventral striatum and central processing of alcohol cues and craving. *Am J Psychiatry*, **161**, 1783-1789.

Heinz, A.J., Beck, A., Meyer-Lindenberg, A., Sterzer, P. & Heinz, A. (2011) Cognitive and neurobiological mechanisms of alcohol-related aggression. *Nat Rev Neurosci*, **12**, 400-413.

Hewig, J., Trippe, R., Hecht, H., Coles, M.G., Holroyd, C.B. & Miltner, W.H. (2007) Decision-making in Blackjack: an electrophysiological analysis. *Cerebral Cortex*, **17**, 865-877.

Hikosaka, O. & Isoda, M. (2010) Switching from automatic to controlled behavior: cortico-basal ganglia mechanisms. *Trends in cognitive sciences*, **14**, 154-161.

Hilbert, A., Tuschen-Caffier, B. & Ohms, M. (2004) Eating disorder examination: a German version of the structured eating disorder interview. *Diagnostica*, **50**, 98-106.

Hodgins, D.C., Maticka-Tyndale, E., el-Guebaly, N. & West, M. (1993) The cast-6: development of a short-form of the Children of Alcoholics Screening Test. *Addict Behav*, **18**, 337-345.

Hogarth, L., Attwood, A.S., Bate, H.A. & Munafo, M.R. (2012a) Acute alcohol impairs human goal-directed action. *Biol Psychol*, **90**, 154-160.

Hogarth, L., Balleine, B.W., Corbit, L.H. & Killcross, S. (2013) Associative learning mechanisms underpinning the transition from recreational drug use to addiction. *Ann N Y Acad Sci*, **1282**, 12-24.

Hogarth, L., Chase, H.W. & Baess, K. (2012b) Impaired goal-directed behavioural control in human impulsivity. *Quarterly journal of experimental psychology*, **65**, 305-316.

Hogarth, L., Dickinson, A., Wright, A., Kouvaraki, M. & Duka, T. (2007) The role of drug expectancy in the control of human drug seeking. *J Exp Psychol Anim Behav Process*, **33**, 484-496.

Holmes, A. & Friston, K. (1998) Generalisability, Random Effects & Population Inference. *Neuroimage*, **7**, S754.

Holmes, T.H. & Rahe, R.H. (1967) The Social Readjustment Rating Scale. *Journal of psychosomatic research*, **11**, 213-218.

Holroyd, C.B. & Coles, M.G. (2002) The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol Rev*, **109**, 679-709.

Huettel, S.A., Song, A.W. & McCarthy, G. (2004) *Functional magnetic resonance imaging*. Sinauer Associates Sunderland.

Huys, Q.J., Cools, R., Golzer, M., Friedel, E., Heinz, A., Dolan, R.J. & Dayan, P. (2011) Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS Comput Biol*, **7**, e1002028.

Huys, Q.J., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P. & Roiser, J.P. (2012) Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Comput Biol*, **8**, e1002410.

Huys, Q.J., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S.J., Dayan, P. & Roiser, J.P. (2015a) Interplay of approximate planning strategies. *Proc Natl Acad Sci U S A*, **112**, 3098-3103.

Huys, Q.J., Tobler, P.N., Hasler, G. & Flagel, S.B. (2014) The role of learning-related dopamine signals in addiction vulnerability. *Prog Brain Res*, **211**, 31-77.

Huys, Q.J.M., Guitart-Masip, M., Dolan, R. & Dayan, P. (2015b) Decision-Theoretic Psychiatry. *Clinical Psychologial Science*, **3**, 400-421.

Hyman, S.E. (2005) Addiction: a disease of learning and memory. *Am J Psychiatry*, **162**, 1414-1422.

Hyman, S.E. (2012) Revolution stalled. *Science translational medicine*, **4**, 155cm111.

Iglesias, S., Mathys, C., Brodersen, K.H., Kasper, L., Piccirelli, M., den Ouden, H.E. & Stephan, K.E. (2013) Hierarchical Prediction Errors in Midbrain and Basal Forebrain during Sensory Learning. *Neuron*, **80**, 519-530.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D.S., Quinn, K., Sanislow, C. & Wang, P. (2010) Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry*, **167**, 748-751.

Jensen, R. (2006) Behaviorism, latent learning, and cognitive maps: needed revisions in introductory psychology textbooks. *The Behavior analyst / MABA*, **29**, 187-209.

Jentsch, J.D., Olausson, P., De La Garza, R., 2nd & Taylor, J.R. (2002) Impairments of reversal learning and response perseveration after repeated, intermittent cocaine administrations to monkeys. *Neuropsychopharmacology*, **26**, 183-190.

Johnson, P.M. & Kenny, P.J. (2010) Dopamine D2 receptors in addiction-like reward dysfunction and compulsive eating in obese rats. *Nat Neurosci*, **13**, 635-641.

Kalivas, P.W. & Volkow, N.D. (2005) The neural basis of addiction: a pathology of motivation and choice. *Am J Psychiatry*, **162**, 1403-1413.

Kappenman, E.S. & Luck, S.J. (2012) ERP components: The ups and downs of brainwave recordings. In Luck, S.J., Kappenman, E.S. (eds) *The Oxford Handbook of Event-Related Potential Components*. Oxford University Press, Oxford, pp. 3-30.

Kapur, S., Phillips, A.G. & Insel, T.R. (2012) Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry*, **17**, 1174-1179.

Kelley, A.E. (2004) Ventral striatal control of appetitive motivation: role in ingestive behavior and reward-related learning. *Neurosci Biobehav Rev*, **27**, 765-776.

Kennedy, D. (2005) Neuroimaging: revolutionary research tool or a post-modern phrenology? *The American journal of bioethics : AJOB*, **5**, 19; discussion W13-14.

Kepecs, A. & Mainen, Z.F. (2012) A computational framework for the study of confidence in humans and animals. *Philos Trans R Soc Lond B Biol Sci*, **367**, 1322-1337.

Keramati, M., Dezfouli, A. & Piray, P. (2011) Speed/accuracy trade-off between the habitual and the goal-directed processes. *PLoS Comput Biol*, **7**, e1002055.

Kim, W., Pitt, M.A., Lu, Z.L., Steyvers, M. & Myung, J.I. (2014) A hierarchical adaptive approach to optimal experimental design. *Neural Comput*, **26**, 2465-2492.

Kirschbaum, C., Pirke, K.M. & Hellhammer, D.H. (1993) The 'Trier Social Stress Test'--a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, **28**, 76-81.

Klimesch, W., Sauseng, P., Hanslmayr, S., Gruber, W. & Freunberger, R. (2007) Event-related phase reorganization may explain evoked neural dynamics. *Neurosci Biobehav Rev*, **31**, 1003-1016.

Klimesch, W., Schack, B., Schabus, M., Doppelmayr, M., Gruber, W. & Sauseng, P. (2004) Phase-locked alpha and theta oscillations generate the P1-N1 complex and are related to memory performance. *Brain Res Cogn Brain Res*, **19**, 302-316.

Kloppel, S., Abdulkadir, A., Jack, C.R., Jr., Koutsouleris, N., Mourao-Miranda, J. & Vemuri, P. (2012) Diagnostic neuroimaging across diseases. *Neuroimage*, **61**, 457-463.

Knutson, B. & Gibbs, S.E. (2007) Linking nucleus accumbens dopamine and blood oxygenation. *Psychopharmacology (Berl)*, **191**, 813-822.

Koechlin, E., Ody, C. & Kouneiher, F. (2003) The architecture of cognitive control in the human prefrontal cortex. *Science*, **302**, 1181-1185.

Koob, G.F. (2008) A role for brain stress systems in addiction. *Neuron*, **59**, 11-34.

Kraus, L. & Augustin, R. (2001) Repräsentativerhebung zum Gebrauch psychoaktiver Substanzen bei Erwachsenen in Deutschland 2000. *Sucht*, **47**, 3-85.

Krugel, L.K., Biele, G., Mohr, P.N., Li, S.C. & Heekeren, H.R. (2009) Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. *Proc Natl Acad Sci U S A*, **106**, 17951-17956.

Kudielka, B.M., Hellhammer, D.H. & Kirschbaum, C. (2007) Ten years of research with the Trier Social Stress Test—revisited. In Harmon-Jones E, W.P. (ed) *Social neuroscience: Integrating biological and psychological explanations of social behavior*. The Guilford Press, New York, pp. 56-83.

Kudielka, B.M., Schommer, N.C., Hellhammer, D.H. & Kirschbaum, C. (2004) Acute HPA axis responses, heart rate, and mood changes to psychosocial stress (TSST) in humans at different times of day. *Psychoneuroendocrinology*, **29**, 983-992.

Kurth-Nelson, Z. & Redish, A.D. (2012) Don'T let me do that! - models of precommitment. *Front Neurosci*, **6**, 138.

Landy, M.S., Trommershauser, J. & Daw, N.D. (2012) Dynamic estimation of task-relevant variance in movement under risk. *J Neurosci*, **32**, 12702-12711.

Laux, L., Glanzmann, P., Schaffner, P. & Spielberger, C. (1981) *Das State-Trait-Angstinventar*. Beltz, Weinheim.

Lee, S.W., Shimojo, S. & O'Doherty, J.P. (2014) Neural Computations Underlying Arbitration between Model-Based and Model-free Learning. *Neuron*, **81**, 687-699.

Lehmann, D. & Skrandies, W. (1980) Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and clinical neurophysiology*, **48**, 609-621.

Lewandowsky, S. & Farrell, S. (2011) *Computational Modeling in Cognition*. SAGE Publications, Inc., Thousand Oaks.

Li, J. & Daw, N.D. (2011) Signals in human striatum are appropriate for policy update rather than value prediction. *J Neurosci*, **31**, 5504-5511.

Li, J., Schiller, D., Schoenbaum, G., Phelps, E.A. & Daw, N.D. (2011) Differential roles of human striatum and amygdala in associative learning. *Nat Neurosci*, **14**, 1250-1252.

Liston, C., McEwen, B.S. & Casey, B.J. (2009) Psychosocial stress reversibly disrupts prefrontal processing and attentional control. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 912-917.

Logothetis, N.K. (2007) The ins and outs of fMRI signals. *Nat Neurosci*, **10**, 1230-1232.

Logothetis, N.K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. (2001) Neurophysiological investigation of the basis of the fMRI signal. *Nature*, **412**, 150-157.

Lohrenz, T., McCabe, K., Camerer, C.F. & Montague, P.R. (2007) Neural signature of fictive learning signals in a sequential investment task. *Proc Natl Acad Sci U S A*, **104**, 9493-9498.

Lorenz, R.C., Gleich, T., Kuhn, S., Pohland, L., Pelz, P., Wustenberg, T., Raufelder, D., Heinz, A. & Beck, A. (2015) Subjective illusion of control modulates striatal reward anticipation in adolescence. *Neuroimage*.

Lucantonio, F., Stalnaker, T.A., Shaham, Y., Niv, Y. & Schoenbaum, G. (2012) The impact of orbitofrontal dysfunction on cocaine addiction. *Nat Neurosci*, **15**, 358-366.

Lucantonio, F., Takahashi, Y.K., Hoffman, A.F., Chang, C.Y., Bali-Chaudhary, S., Shaham, Y., Lupica, C.R. & Schoenbaum, G. (2014) Orbitofrontal activation restores insight lost after cocaine use. *Nat Neurosci*, **17**, 1092-1099.

Luce, R.D. (1995) Four tensions concerning mathematical modeling in psychology. *Annu Rev Psychol*, **46**, 1-26.

Luck, S.J. (2012) Event-related potentials. In Cooper, H., Camic, P.M., Long, D.L., Panter, A.T., Rindskopf, D., Sher, K.J. (eds) *APA handbook of research methods in psychology*. American Psychological Association, Washington, DC, pp. 523-546.

Luine, V., Villegas, M., Martinez, C. & McEwen, B.S. (1994) Repeated stress causes reversible impairments of spatial memory performance. *Brain research*, **639**, 167-170.

Lupien, S.J., McEwen, B.S., Gunnar, M.R. & Heim, C. (2009) Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nat Rev Neurosci*, **10**, 434-445.

Luthi, A. & Luscher, C. (2014) Pathological circuit function underlying addiction and anxiety disorders. *Nat Neurosci*, **17**, 1635-1643.

Maia, T.V. (2015) Introduction to the Series on Computational Psychiatry. *Clinical Psychologial Science*, **3**, 374-377.

Maia, T.V. & Frank, M.J. (2011) From reinforcement learning models to psychiatric and neurological disorders. *Nat Neurosci*, **14**, 154-162.

Maier, S.U., Makwana, A.B. & Hare, T.A. (2015) Acute Stress Impairs Self-Control in Goal-Directed Choice by Altering Multiple Functional Connections within the Brain's Decision Circuits. *Neuron*, **87**, 621-631.

Makeig, S. & Onton, J. (2012) ERP Features and EEG Dynamics: An ICA Perspective. In Luck, S.J., Kappenman, E.S. (eds) *The Oxford Handbook of Event-Related Potential Components*. Oxford University Press, Oxford, pp. 51-86.

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

Mars, R.B., Shea, N.J., Kolling, N. & Rushworth, M.F. (2012) Model-based analyses: Promises, pitfalls, and example applications to the study of cognitive control. *Quarterly journal of experimental psychology*, **65**, 252-267.

Martinez, D., Gil, R., Slifstein, M., Hwang, D.R., Huang, Y., Perez, A., Kegeles, L., Talbot, P., Evans, S., Krystal, J., Laruelle, M. & Abi-Dargham, A. (2005) Alcohol dependence is associated with blunted dopamine transmission in the ventral striatum. *Biol Psychiatry*, **58**, 779-786.

Martinez, D., Narendran, R., Foltin, R.W., Slifstein, M., Hwang, D.R., Broft, A., Huang, Y., Cooper, T.B., Fischman, M.W., Kleber, H.D. & Laruelle, M. (2007) Amphetamine-induced dopamine release: markedly blunted in cocaine dependence and predictive of the choice to self-administer cocaine. *Am J Psychiatry*, **164**, 622-629.

Martinez, D., Saccone, P.A., Liu, F., Slifstein, M., Orlowska, D., Grassetti, A., Cook, S., Broft, A., Van Heertum, R. & Comer, S.D. (2012) Deficits in dopamine D(2) receptors and presynaptic dopamine in heroin dependence: commonalities and differences with other types of addiction. *Biol Psychiatry*, **71**, 192-198.

Mathys, C., Daunizeau, J., Friston, K.J. & Stephan, K.E. (2011) A bayesian foundation for individual learning under uncertainty. *Front Hum Neurosci*, **5**, 39.

Mathys, C.D., Lomakina, E.I., Daunizeau, J., Iglesias, S., Brodersen, K.H., Friston, K.J. & Stephan, K.E. (2014) Uncertainty in perception and the Hierarchical Gaussian Filter. *Front Hum Neurosci*, **8**, 825.

McEwen, B.S. (2004) Protection and damage from acute and chronic stress: allostasis and allostatic overload and relevance to the pathophysiology of psychiatric disorders. *Annals of the New York Academy of Sciences*, **1032**, 1-7.

McGee, T.J., King, C., Tremblay, K., Nicol, T.G., Cunningham, J. & Kraus, N. (2001) Long-term habituation of the speech-elicited mismatch negativity. *Psychophysiology*, **38**, 653-658.

McNair, D., Lorr, M. & Droppleman, L. (1971) *Manual for the Profile of Mood States*. Educational and Industrial Testing Services, San Diego.

Mendelson, J.H., Sholar, M.B., Goletiani, N., Siegel, A.J. & Mello, N.K. (2005) Effects of low- and high-nicotine cigarette smoking on mood states and the HPA axis in men. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, **30**, 1751-1763.

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

Merikangas, K.R., Stevens, D.E., Fenton, B., Stolar, M., O'Malley, S., Woods, S.W. & Risch, N. (1998a) Co-morbidity and familial aggregation of alcoholism and anxiety disorders. *Psychol Med*, **28**, 773-788.

Miltner, W.H., Braun, C.H. & Coles, M.G. (1997) Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a "generic" neural system for error detection. *Journal of cognitive neuroscience*, **9**, 788-798.

Montague, P.R., Dayan, P. & Sejnowski, T.J. (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci*, **16**, 1936-1947.

Montague, P.R., Dolan, R.J., Friston, K.J. & Dayan, P. (2012) Computational psychiatry. *Trends Cogn Sci*, **16**, 72-80.

Morgan, D., Grant, K.A., Gage, H.D., Mach, R.H., Kaplan, J.R., Prioleau, O., Nader, S.H., Buchheimer, N., Ehrenkaufer, R.L. & Nader, M.A. (2002) Social dominance in monkeys: dopamine D2 receptors and cocaine self-administration. *Nat Neurosci*, **5**, 169-174.

Morris, L.S., Baek, K., Kundu, P., Harrison, N.A., Frank, M.J. & Voon, V. (2015) Biases in the Explore-Exploit Tradeoff in Addictions: the Role of Avoidance of Uncertainty. *Neuropsychopharmacology*, **accepted article preview**.

Myung, J.I. & Pitt, M.A. (2009) Optimal experimental design for model discrimination. *Psychol Rev*, **116**, 499-518.

Nidal, K. & Malik, A.S. (2014) *EEG/ERP Analysis: Methods and Applications*. CRC Press.

Norris, D. (2005) How do computational models help us build better theories? In Cutler, A. (ed) *Twenty-first century psycholinguistics: four cornerstones*. Lawrence Erlbaum, Mahwah, pp. 331–346.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K. & Dolan, R.J. (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, **304**, 452-454.

O'Doherty, J.P. (2014) The problem with value. *Neurosci Biobehav Rev*, **43**, 259-268.

O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H. & Dolan, R.J. (2003) Temporal difference models and reward-related learning in the human brain. *Neuron*, **38**, 329-337.

O'Doherty, J.P., Hampton, A. & Kim, H. (2007) Model-based fMRI and its application to reward learning and decision making. *Ann N Y Acad Sci*, **1104**, 35-53.

Ogawa, S., Lee, T.M., Kay, A.R. & Tank, D.W. (1990) Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci U S A*, **87**, 9868-9872.

Ostlund, S.B. & Balleine, B.W. (2005) Lesions of medial prefrontal cortex disrupt the acquisition but not the expression of goal-directed learning. *J Neurosci*, **25**, 7763-7770.

Otto, A.R., Gershman, S.J., Markman, A.B. & Daw, N.D. (2013a) The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychol Sci*, **24**, 751-761.

Otto, A.R., Raio, C.M., Chiang, A., Phelps, E.A. & Daw, N.D. (2013b) Working-memory capacity protects model-based learning from stress. *Proc Natl Acad Sci U S A*, **110**, 20941-20946.

Otto, A.R., Skatova, A., Madlon-Kay, S. & Daw, N.D. (2015) Cognitive control predicts use of model-based reinforcement learning. *J Cogn Neurosci*, **27**, 319-333.

Park, S.Q., Kahnt, T., Beck, A., Cohen, M.X., Dolan, R.J., Wrase, J. & Heinz, A. (2010) Prefrontal cortex fails to learn from reward prediction errors in alcohol dependence. *J Neurosci*, **30**, 7749-7753.

Parvaz, M.A., Konova, A.B., Proudfit, G.H., Dunning, J.P., Malaker, P., Moeller, S.J., Maloney, T., Alia-Klein, N. & Goldstein, R.Z. (2015) Impaired Neural Response to Negative Prediction Errors in Cocaine Addiction. *The Journal of Neuroscience*, **35**, 1872-1879.

Patton, J.H., Stanford, M.S. & Barratt, E.S. (1995) Factor structure of the Barratt impulsiveness scale. *J Clin Psychol*, **51**, 768-774.

Pauling, L. & Coryell, C.D. (1936) The Magnetic Properties and Structure of Hemoglobin, Oxyhemoglobin and Carbonmonoxyhemoglobin. *Proc Natl Acad Sci U S A*, **22**, 210-216.

Paulus, M.P., Lovero, K.L., Wittmann, M. & Leland, D.S. (2008) Reduced behavioral and neural activation in stimulant users to different error rates during decision making. *Biol Psychiatry*, **63**, 1054-1060.

Paulus, M.P., Rogalsky, C., Simmons, A., Feinstein, J.S. & Stein, M.B. (2003) Increased activation in the right insula during risk-taking decision making is related to harm avoidance and neuroticism. *Neuroimage*, **19**, 1439-1448.

Penny, W.D., Friston, K.J., Ashburner, J.T., Kiebel, S.J. & Nichols, T.E. (2011) *Statistical parametric mapping: the analysis of functional brain images: the analysis of functional brain images*. Academic press.

Petzold, A., Plessow, F., Goschke, T. & Kirschbaum, C. (2010) Stress reduces use of negative feedback in a feedback-based learning task. *Behavioral neuroscience*, **124**, 248-255.

Philiastides, M.G., Biele, G., Vavatzanidis, N., Kazzer, P. & Heekeren, H.R. (2010) Temporal dynamics of prediction error processing during reward-based decision making. *Neuroimage*, **53**, 221-232.

Pitt, M.A. & Myung, I.J. (2002) When a good fit can be bad. *Trends Cogn Sci*, **6**, 421-425.

Pizzagalli, D.A. (2007) Electroencephalography and high-density electrophysiological source localization. *Handbook of psychophysiology*, **3**, 56-84.

Plessow, F., Fischer, R., Kirschbaum, C. & Goschke, T. (2011) Inflexibly focused under stress: acute psychosocial stress increases shielding of action goals at the expense of reduced cognitive flexibility with increasing time lag to the stressor. *Journal of cognitive neuroscience*, **23**, 3218-3227.

Plessow, F., Kiesel, A. & Kirschbaum, C. (2012) The stressed prefrontal cortex and goal-directed behaviour: acute psychosocial stress impairs the flexible implementation of task goals. *Experimental brain research*, **216**, 397-408.

Poldrack, R.A. (2006) Can cognitive processes be inferred from neuroimaging data? *Trends Cogn Sci*, **10**, 59-63.

Popper, K.R. (1982) *Logik der Forschung*. JCB Mohr (Paul Siebeck).

Preuschoff, K., Quartz, S.R. & Bossaerts, P. (2008) Human insula activation reflects risk prediction errors as well as risk. *J Neurosci*, **28**, 2745-2752.

Price, C.J. & Friston, K.J. (1999) Scanning patients with tasks they can perform. *Human brain mapping*, **8**, 102-108.

Radenbach, C., Reiter, A.M., Engert, V., Sjoerds, Z., Villringer, A., Heinze, H.J., Deserno, L. & Schlagenhauf, F. (2015) The interaction of acute and chronic stress impairs model-based behavioral control. *Psychoneuroendocrinology*, **53**, 268-280.

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

Rangel, A. & Hare, T. (2010) Neural computations associated with goal-directed choice. *Curr Opin Neurobiol*, **20**, 262-270.

Ratcliff, R. & Rouder, J.N. (1998) Modeling response times for two-choice decisions. *Psychological Science*, **9**, 347-356.

Redish, A.D. (2004) Addiction as a computational process gone awry. *Science*, **306**, 1944-1947.

Redish, A.D., Jensen, S. & Johnson, A. (2008) A unified framework for addiction: vulnerabilities in the decision process. *Behav Brain Sci*, **31**, 415-437; discussion 437-487.

Reitan, R.M. (1955) The relation of the trail making test to organic brain damage. *J Consult Psychol*, **19**, 393-394.

Reiter, A.M.F, Deserno, L., Kallert, T., Heinz, A., Heinze, H.J. & Schlagenhauf, F. (under review). Neglecting what might have happened – disturbed inference on alternative choices in alcohol-dependent patients.

Reiter, A.M.F., Deserno, L., Wilbertz, T., Heinze, H.J. & Schlagenhauf, F. (under review). Risk factors for addiction and their association with model-based behavioral control.

Reiter, A.M.F., Heinze, H.J., Schlagenhauf, F. & Deserno, L. (under review). Impaired flexible reward-based decision-making in Binge Eating Disorder: evidence from computational modeling and functional neuroimaging.

Reiter, A.M.F., Koch, S.P., Schröger, E., Hinrichs, H., Heinze, H.J., Deserno, L. & Schlagenhauf, F. (in revision) The Feedback-Related Negativity codes components of abstract inference during reward-based decision-making *Journal of Cognitive Neuroscience*.

Robbins, T.W. & Clark, L. (2015) Behavioral addictions. *Curr Opin Neurobiol*, **30**, 66-72.

Robbins, T.W. & Everitt, B.J. (1999) Drug addiction: bad habits add up. *Nature*, **398**, 567-570.

Robbins, T.W., Gillan, C.M., Smith, D.G., de Wit, S. & Ersche, K.D. (2012) Neurocognitive endophenotypes of impulsivity and compulsivity: towards dimensional psychiatry. *Trends Cogn Sci*, **16**, 81-91.

Roesch, M.R., Esber, G.R., Li, J., Daw, N.D. & Schoenbaum, G. (2012) Surprise! Neural correlates of Pearce-Hall and Rescorla-Wagner coexist within the brain. *Eur J Neurosci*, **35**, 1190-1200.

Roth, A. & Fonagy, P. (2013) *What works for whom?: a critical review of psychotherapy research*. Guilford Publications.

Rushworth, M.F., Noonan, M.P., Boorman, E.D., Walton, M.E. & Behrens, T.E. (2011) Frontal cortex and reward-guided learning and decision-making. *Neuron*, **70**, 1054-1069.

Russell, J.A., Weiss, A. & Mendelsohn, G.A. (1989) Affect Grid - a Single-Item Scale of Pleasure and Arousal. *J Pers Soc Psychol*, **57**, 493-502.

Salamone, J.D. & Correa, M. (2013) Dopamine and food addiction: lexicon badly needed. *Biol Psychiatry*, **73**, e15-24.

Schad, D.J., Junger, E., Sebold, M., Garbusow, M., Bernhardt, N., Javadi, A.H., Zimmermann, U.S., Smolka, M.N., Heinz, A., Rapp, M.A. & Huys, Q.J. (2014) Processing speed enhances model-based over model-free reinforcement learning in the presence of high working memory functioning. *Front Psychol*, **5**, 1450.

Schlagenhauf, F., Huys, Q.J., Deserno, L., Rapp, M.A., Beck, A., Heinze, H.J., Dolan, R. & Heinz, A. (2014) Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. *Neuroimage*, **89**, 171-180.

Schlagenhauf, F., Rapp, M.A., Huys, Q.J., Beck, A., Wustenberg, T., Deserno, L., Buchholz, H.G., Kalbitzer, J., Buchert, R., Bauer, M., Kienast, T., Cumming, P., Plotkin, M., Kumakura, Y., Grace, A.A., Dolan, R.J. & Heinz, A. (2013) Ventral striatal prediction error signaling is associated with dopamine synthesis capacity and fluid intelligence. *Hum Brain Mapp*, **34**, 1490-1499.

Schluter, T., Winz, O., Henkel, K., Prinz, S., Rademacher, L., Schmaljohann, J., Dautzenberg, K., Cumming, P., Kumakura, Y., Rex, S., Mottaghy, F.M., Grunder, G. & Vernaleken, I. (2013) The impact of dopamine on aggression: an [18F]-FDOPA PET Study in healthy males. *J Neurosci*, **33**, 16889-16896.

Schmidt, K.-H. & Metzler, P. (1992) *Wortschatztest (WST)*. Beltz Test GmbH., Weinheim.

Schoenbaum, G. & Setlow, B. (2005) Cocaine makes actions insensitive to outcomes but not extinction: implications for altered orbitofrontal-amygdalar function. *Cereb Cortex*, **15**, 1162-1169.

Schoenbaum, G. & Shaham, Y. (2008) The role of orbitofrontal cortex in drug addiction: a review of preclinical studies. *Biol Psychiatry*, **63**, 256-262.

Schommer, N.C., Hellhammer, D.H. & Kirschbaum, C. (2003) Dissociation between reactivity of the hypothalamus-pituitary-adrenal axis and the sympathetic-adrenal-medullary system to repeated psychosocial stress. *Psychosomatic medicine*, **65**, 450-460.

Schultz, W. (2013) Updating dopamine reward signals. *Curr Opin Neurobiol*, **23**, 229-238.

Schultz, W., Dayan, P. & Montague, P.R. (1997) A neural substrate of prediction and reward. *Science*, **275**, 1593-1599.

Schumann, G., Loth, E., Banaschewski, T., Barbot, A., Barker, G., Buchel, C., Conrod, P.J., Dalley, J.W., Flor, H., Gallinat, J., Garavan, H., Heinz, A., Itterman, B., Lathrop, M., Mallik, C., Mann, K., Martinot, J.L., Paus, T., Poline, J.B., Robbins, T.W., Rietschel, M., Reed, L., Smolka, M., Spanagel, R., Speiser, C., Stephens, D.N., Strohle, A., Struve, M. & consortium, I. (2010) The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Mol Psychiatry*, **15**, 1128-1139.

Schwabe, L. & Wolf, O.T. (2009) Stress prompts habit behavior in humans. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, **29**, 7191-7198.

Schwabe, L. & Wolf, O.T. (2011) Stress-induced modulation of instrumental behavior: from goal-directed to habitual control of action. *Behavioural brain research*, **219**, 321-328.

Schwabe, L. & Wolf, O.T. (2013) Stress and multiple memory systems: from 'thinking' to 'doing'. *Trends in cognitive sciences*, **17**, 60-68.

Schwartenbeck, P., FitzGerald, T.H., Mathys, C., Dolan, R., Wurst, F., Kronbichler, M. & Friston, K. (2015) Optimal inference with suboptimal models: addiction and active Bayesian inference. *Med Hypotheses*, **84**, 109-117.

Schwarz, G. (1978) Estimating the dimension of a model. *The annals of statistics*, **6**, 461-464.

Seamans, J.K. & Yang, C.R. (2004) The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Prog Neurobiol*, **74**, 1-58.

Sebold, M., Deserno, L., Nebe, S., Schad, D.J., Garbusow, M., Hagele, C., Keller, J., Junger, E., Kathmann, N., Smolka, M., Rapp, M.A., Schlagenhauf, F., Heinz, A. & Huys, Q.J. (2014) Model-based and model-free decisions in alcohol dependence. *Neuropsychobiology*, **70**, 122-131.

Shohamy, D. & Wagner, A.D. (2008) Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. *Neuron*, **60**, 378-389.

Simon, D.A. & Daw, N.D. (2011) Neural correlates of forward planning in a spatial decision task in humans. *J Neurosci*, **31**, 5526-5539.

Simon, D.A. & Daw, N.D. (2012) Dual-system learning models and drugs of abuse *Computational Neuroscience of Drug Addiction*. Springer, pp. 145-161.

Singer, T., Critchley, H.D. & Preuschoff, K. (2009) A common role of insula in feelings, empathy and uncertainty. *Trends Cogn Sci*, **13**, 334-340.

Sinha, R. (2008) Chronic stress, drug use, and vulnerability to addiction. *Ann N Y Acad Sci*, **1141**, 105-130.

Sjoerds, Z., de Wit, S., van den Brink, W., Robbins, T.W., Beekman, A.T., Penninx, B.W. & Veltman, D.J. (2013) Behavioral and neuroimaging evidence for overreliance on habit learning in alcohol-dependent patients. *Transl Psychiatry*, **3**, e337.

Smith, D.G. & Robbins, T.W. (2013) The neurobiological underpinnings of obesity and binge eating: a rationale for adopting the food addiction model. *Biol Psychiatry*, **73**, 804-810.

Smittenaar, P. (2015) Action control in uncertain environments. *Wellcome Trust Centre for Neuroimaging*. University College London, pp. 272.

Smittenaar, P., FitzGerald, T.H., Romei, V., Wright, N.D. & Dolan, R.J. (2013) Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. *Neuron*, **80**, 914-919.

Soares, J.M., Sampaio, A., Ferreira, L.M., Santos, N.C., Marques, F., Palha, J.A., Cerqueira, J.J. & Sousa, N. (2012) Stress-induced changes in human decision-making are reversible. *Translational psychiatry*, **2**, e131.

Sobell, L.C.S., M.B. (1992) Timeline follow-back: A technique for assessing self-reported alcohol consumption. In Litten, R.Z.A., J. (ed) *Measuring alcohol consumption: Psychosocial and biological methods*. Humana Press, New Jersey.

Sokol-Hessner, P., Hutcherson, C., Hare, T. & Rangel, A. (2012) Decision value computation in DLPFC and VMPFC adjusts to the available decision time. *Eur J Neurosci*, **35**, 1065-1074.

Stahl, S.M. (2008) *Stahl's Essential Psychopharmacology: Neuroscientific Basis and Practical Applications*. Cambridge University Press.

Stanford, M.S., Mathias, C.W., Dougherty, D.M., Lake, S.L., Anderson, N.E. & Patton, J.H. (2009) Fifty years of the Barratt Impulsiveness Scale: An update and review. *Personality and Individual Differences*, **47**, 385-395.

Starcke, K. & Brand, M. (2012) Decision making under stress: a selective review. *Neuroscience and biobehavioral reviews*, **36**, 1228-1248.

Starcke, K., Polzer, C., Wolf, O.T. & Brand, M. (2011) Does stress alter everyday moral decision-making? *Psychoneuroendocrinology*, **36**, 210-219.

Stephan, K.E., Iglesias, S., Heinzle, J. & Diaconescu, A.O. (2015) Translational Perspectives for Computational Neuroimaging. *Neuron*, **87**, 716-732.

Stephan, K.E. & Mathys, C. (2014) Computational approaches to psychiatry. *Curr Opin Neurobiol*, **25**, 85-92.

Stephan, K.E., Penny, W.D., Daunizeau, J., Moran, R.J. & Friston, K.J. (2009) Bayesian model selection for group studies. *Neuroimage*, **46**, 1004-1017.

Sutton, R.S. (1992) Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. *Proceedings of the seventh international conference on machine learning*, p. 216-224.

Sutton, R.S. (1992) Gain adaptation beats least squares? *Proceedings of the 7th Yale Workshop on Adaptive and Learning Systems*, pp. 161-166.

Sutton, R.S. & Barto, A.G. (1998) *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA.

Svaldi, J., Brand, M. & Tuschen-Caffier, B. (2010) Decision-making impairments in women with binge eating disorder. *Appetite*, **54**, 84-92.

Takahashi, Y.K., Chang, C.Y., Lucantonio, F., Haney, R.Z., Berg, B.A., Yau, H.J., Bonci, A. & Schoenbaum, G. (2013) Neural estimates of imagined outcomes in the orbitofrontal cortex drive behavior and learning. *Neuron*, **80**, 507-518.

Talmi, D., Atkinson, R. & El-Deredy, W. (2013) The feedback-related negativity signals salience prediction errors, not reward prediction errors. *The Journal of Neuroscience*, **33**, 8264-8269.

Talmi, D., Fuentemilla, L., Litvak, V., Duzel, E. & Dolan, R.J. (2012) An MEG signature corresponding to an axiomatic model of reward prediction error. *Neuroimage*, **59**, 635-645.

Talmi, D., Seymour, B., Dayan, P. & Dolan, R.J. (2008) Human pavlovian-instrumental transfer. *J Neurosci*, **28**, 360-368.

Tan, H.Y., Sust, S., Buckholtz, J.W., Mattay, V.S., Meyer-Lindenberg, A., Egan, M.F., Weinberger, D.R. & Callicott, J.H. (2006) Dysfunctional prefrontal regional specialization and compensation in schizophrenia. *Am J Psychiatry*, **163**, 1969-1977.

Tanabe, J., Reynolds, J., Krmpotich, T., Claus, E., Thompson, L.L., Du, Y.P. & Banich, M.T. (2013) Reduced neural tracking of prediction error in substance-dependent individuals. *Am J Psychiatry*, **170**, 1356-1363.

Thorndike, E.L. (1911) *Animal intelligence: Experimental studies*. Macmillan.

Tiffany, S.T., Carter, B.L. & Singleton, E.G. (2000) Challenges in the manipulation, assessment and interpretation of craving relevant variables. *Addiction*, **95 Suppl 2**, S177-187.

Tolman, E.C. (1948) Cognitive maps in rats and men *Psychol Rev*, **55**, 189-208.

Trantham-Davidson, H., Burnett, E.J., Gass, J.T., Lopez, M.F., Mulholland, P.J., Centanni, S.W., Floresco, S.B. & Chandler, L.J. (2014) Chronic alcohol disrupts dopamine receptor activity and the cognitive function of the medial prefrontal cortex. *J Neurosci*, **34**, 3706-3718.

Tricomi, E., Balleine, B.W. & O'Doherty, J.P. (2009) A specific role for posterior dorsolateral striatum in human habit learning. *Eur J Neurosci*, **29**, 2225-2232.

Tse, Y.C., Montoya, I., Wong, A.S., Mathieu, A., Lissemore, J., Lagace, D.C. & Wong, T.P. (2014) A longitudinal study of stress-induced hippocampal volume changes in mice that are susceptible or resilient to chronic social defeat. *Hippocampus*, **24**, 1120-1128.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B. & Joliot, M. (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, **15**, 273-289.

Ullsperger, M., Fischer, A.G., Nigbur, R. & Endrass, T. (2014) Neural mechanisms and temporal dynamics of performance monitoring. *Trends in cognitive sciences*, **18**, 259-267.

Valentin, V.V., Dickinson, A. & O'Doherty, J.P. (2007) Determining the neural substrates of goal-directed learning in the human brain. *J Neurosci*, **27**, 4019-4026.

Van Cauter, E. & Refetoff, S. (1985) Evidence for two subtypes of Cushing's disease based on the analysis of episodic cortisol secretion. *The New England journal of medicine*, **312**, 1343-1349.

van der Meer, M.A., Johnson, A., Schmitzer-Torbert, N.C. & Redish, A.D. (2010) Triple dissociation of information processing in dorsal striatum, ventral striatum, and hippocampus on a learned spatial decision task. *Neuron*, **67**, 25-32.

van Maanen, L., Forstmann, B.U., Keuken, M.C., Wagenmakers, E.J. & Heathcote, A. (2015) The impact of MRI scanner environment on perceptual decision-making. *Behav Res Methods*.

Verbruggen, F., McLaren, I.P. & Chambers, C.D. (2014) Banishing the Control Homunculi in Studies of Action Control and Behavior Change. *Perspectives on psychological science : a journal of the Association for Psychological Science*, **9**, 497-524.

Verdejo-Garcia, A., Lawrence, A.J. & Clark, L. (2008) Impulsivity as a vulnerability marker for substance-use disorders: review of findings from high-risk research, problem gamblers and genetic association studies. *Neurosci Biobehav Rev*, **32**, 777-810.

Volkow, N., Fowler, J., Wang, G., Baler, R. & Telang, F. (2009) Imaging dopamine's role in drug abuse and addiction. *Neuropharmacology*, **56**, 3-8.

Volkow, N.D., Fowler, J.S., Wang, G.J. & Swanson, J.M. (2004) Dopamine in drug abuse and addiction: results from imaging studies and treatment implications. *Mol Psychiatry*, **9**, 557-569.

Volkow, N.D., Fowler, J.S., Wolf, A.P., Schlyer, D., Shiue, C.Y., Alpert, R., Dewey, S.L., Logan, J., Bendriem, B., Christman, D. & et al. (1990) Effects of chronic cocaine abuse on postsynaptic dopamine receptors. *Am J Psychiatry*, **147**, 719-724.

Volkow, N.D., Wang, G.J., Begleiter, H., Porjesz, B., Fowler, J.S., Telang, F., Wong, C., Ma, Y., Logan, J., Goldstein, R., Alexoff, D. & Thanos, P.K. (2006) High levels of dopamine D2 receptors in unaffected members of alcoholic families: possible protective factors. *Arch Gen Psychiatry*, **63**, 999-1008.

Volkow, N.D., Wang, G.J., Fowler, J.S., Logan, J., Hitzemann, R., Ding, Y.S., Pappas, N., Shea, C. & Piscani, K. (1996) Decreases in dopamine receptors but not in dopamine transporters in alcoholics. *Alcohol Clin Exp Res*, **20**, 1594-1598.

Volkow, N.D., Wang, G.J., Tomasi, D. & Baler, R.D. (2013) The addictive dimensionality of obesity. *Biol Psychiatry*, **73**, 811-818.

Volkow, N.D. & Wise, R.A. (2005) How can drug addiction help us understand obesity? *Nat Neurosci*, **8**, 555-560.

Von Aster, M., Neubauer, A. & Horn, R. (2006) *Wechsler Intelligenztest für Erwachsene*. Harcourt Test Services, Frankfurt.

Voon, V., Derbyshire, K., Ruck, C., Irvine, M.A., Worbe, Y., Enander, J., Schreiber, L.R., Gillan, C., Fineberg, N.A., Sahakian, B.J., Robbins, T.W., Harrison, N.A., Wood, J., Daw, N.D., Dayan, P., Grant, J.E. & Bullmore, E.T. (2015) Disorders of compulsivity: a common bias towards learning habits. *Mol Psychiatry*, **20**, 345-352.

Voon, V., Irvine, M.A., Derbyshire, K., Worbe, Y., Lange, I., Abbott, S., Morein-Zamir, S., Dudley, R., Caprioli, D., Harrison, N.A., Wood, J., Dalley, J.W., Bullmore, E.T., Grant, J.E. & Robbins, T.W. (2014) Measuring "waiting" impulsivity in substance addictions and binge eating disorder in a novel analogue of rodent serial reaction time task. *Biol Psychiatry*, **75**, 148-155.

Walsh, M.M. & Anderson, J.R. (2012) Learning from experience: event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neurosci Biobehav Rev*, **36**, 1870-1884.

Wang, X.J. & Krystal, J.H. (2014) Computational psychiatry. *Neuron*, **84**, 638-654.

Watkins, C.J. & Dayan, P. (1992) Q-learning. *Machine learning*, **8**, 279-292.

Wechsler, D. (1945) A standardized memory scale for clinical use. *Journal of Psychology*, **19**, 87-95.

Wechsler, D. (1955) *Wechsler Adult Intelligence Scale Manual*. Psychological Corporation, New York.

Weisberg, D.S., Keil, F.C., Goodstein, J., Rawson, E. & Gray, J.R. (2008) The seductive allure of neuroscience explanations. *J Cogn Neurosci*, **20**, 470-477.

White, C.N. & Poldrack, R.A. (2013) Using fMRI to Constrain Theories of Cognition. *Perspectives on psychological science : a journal of the Association for Psychological Science*, **8**, 79-83.

Wiecki, T., Poland, J. & Frank, M.J. (2015) Model-Based Cognitive Neuroscience Approaches to Computational Psychiatry Clustering and Classification. *Clinical Psychologial Science*, **3**, 378-399.

Wilbertz, T., Deserno, L., Horstmann, A., Neumann, J., Villringer, A., Heinze, H.J., Boehler, C.N. & Schlagenhauf, F. (2014) Response inhibition and its relation to multidimensional impulsivity. *Neuroimage*, **103C**, 241-248.

Williamson, D.A., Netemeyer, R.G., Jackman, L.P., Anderson, D.A., Funsch, C.L. & Rabalais, J.Y. (1995) Structural equation modeling of risk factors for the development of eating disorder symptoms in female athletes. *International Journal of Eating Disorders*, **17**, 387-393.

Wimmer, G.E., Daw, N.D. & Shohamy, D. (2012) Generalization of value in reinforcement learning by humans. *Eur J Neurosci*, **35**, 1092-1104.

Wrase, J., Schlagenhauf, F., Kienast, T., Wustenberg, T., Bermpohl, F., Kahnt, T., Beck, A., Strohle, A., Juckel, G., Knutson, B. & Heinz, A. (2007) Dysfunction of reward processing correlates with alcohol craving in detoxified alcoholics. *Neuroimage*, **35**, 787-794.

Wunderlich, K., Dayan, P. & Dolan, R.J. (2012a) Mapping value based planning and extensively trained choice in the human brain. *Nat Neurosci*, **15**, 786-791.

Wunderlich, K., Smittenaar, P. & Dolan, R.J. (2012b) Dopamine enhances model-based over model-free choice behavior. *Neuron*, **75**, 418-424.

Wunderlich, K., Symmonds, M., Bossaerts, P. & Dolan, R.J. (2011) Hedging your bets by learning reward correlations in the human brain. *Neuron*, **71**, 1141-1152.

Yin, H.H., Knowlton, B.J. & Balleine, B.W. (2004) Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur J Neurosci*, **19**, 181-189.

Yin, H.H., Ostlund, S.B., Knowlton, B.J. & Balleine, B.W. (2005) The role of the dorsomedial striatum in instrumental conditioning. *Eur J Neurosci*, **22**, 513-523.

Ziauddeen, H., Farooqi, I.S. & Fletcher, P.C. (2012) Obesity and the brain: how convincing is the addiction model? *Nat Rev Neurosci*, **13**, 279-286.

List of Tables

List of Figures

Abbreviations

ACQ	Alcohol Craving Questionnaire
aI	anterior Insula
ANOVA	Analysis of Variance
AUDIT	Alcohol Use Disorder Test
BED	Binge Eating Disorder
BIC	Bayesian Information Criterion
BIS	Barratt Impulsiveness Scale
BMS	Bayesian Model Selection
BOLD	Blood Oxygen Level Dependence
DA	Dopamine
DSM	Diagnostic Statistical Manual
DS	Digit Span
DU	Double-Update
e.g.	exempli gratia (for example)
EEG	Electroencephalography
ERP	Event-Related Potential
et al.	and others
fMRI	Functional magnetic resonance imaging
FRN	Feedback-Related Negativity
GLM	General Linear Model
Hb	Hemoglobin
HC	Healthy control
i.e.	id est (that is)
ICA	Independent component analysis
iDU	individually weighte double-updating
ISI	Interstimulus interval
ITI	Intertrial interval
MANOVA	Multivariate analysis of variance
MFC	Media frontal cortex
MNI	Montreal Neurological Institute
mPFC	Medial prefrontal cortex
ms	Milliseconds
OCD	Obsessive Compulsive Disorder
OCDS	Obsessive compulsive drinking scale
PE	Prediction Error
PET	Positions Emissions Tomography
RF pulse	Radio frequency
RL	Reinforcement Learning
RPE	Reward Prediction Error
s	Second
SD	Standard Deviation
SI	Supplementary Information
Sig.	Significance
S-R	Stimulus-Response
SU	Single-Update
TD	Temporal difference

TE	Echo Time
TLFB	Time-Line-Follow-Back-Questionnaire
TMS	Transcranial Magnetic Stimulation
TR	Repetition Time
VBM	Voxel-based Morphometry
vlPFC	Ventro-lateral prefrontal cortex
vmPFC	Ventro-medial prefrontal cortex
WST	Wortschatztest
XP	Exceedance Probabilities

Curriculum Vitae

Andrea Maria Franziska Reiter
Date of Birth: 17.05.1986 in Augsburg

Education

Since 2012: Studies of Psychotherapy (Master of Advanced Sciences, Cognitive Behavioral Therapy), University of Bern, Switzerland

Since 2012: Postgraduate Training in Psychotherapy (Cognitive Behavioral Therapy), Deutsche Gesellschaft für Verhaltenstherapie, Tübingen, Germany

2012: Diploma in Psychology

2009: Studies of Psychology, University of Lisbon, Portugal

2008 – 2012: Studies of Literature, Linguistics and Cultural Studies of Modern Languages (Romance and English studies), University of Würzburg, Germany

2006 – 2012: Studies of Psychology, University of Würzburg, Germany

2005: University Entrance Qualification (Abitur), Gymnasium Königsbrunn, Germany

Career

Since 2015: Research Position, Lifespan Developmental Neuroscience, Technische Universität Dresden, Germany

2012 – 2015: PhD Student, International Max Planck Research School for the Neuroscience of Communication, Research Group: Cognitive and Affective Control of Behavioral Adaptation, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

2012 – 2014: Lecturer for Psychology of Emotion and Motivation, Department of Psychology II, Julius-Maximilians-Universität Würzburg, Germany

2012: Clinical Psychologist, Department of Psychiatry, Hospital Tauberbischofsheim, Germany

Publications

Reiter, A.M.F, Deserno, L., Kallert, T., Heinz, A., Heinze, H.J., Schlagenhauf, F. (under review). Neglecting what might have happened – disturbed inference on alternative choices in alcohol-dependent patients.

Reiter, A.M.F.*, Deserno, L.*, Wilbertz, T., Heinze, H.J., Schlagenhauf, F. (under review). Risk factors for addiction and their association with model-based behavioral control.
*equal contribution

Reiter, A.M.F., Heinze, H.J., Schlagenhauf, F., Deserno, L. (under review). Impaired flexible reward-based decision-making – evidence from computational modeling and functional neuroimaging.

Mussel, P., **Reiter, A. M.F.,** Schmitt, B., Albrecht, B. Osinsky, R., Hewig, J. (under review). Approaching the construct of greed.

Reiter, A.M.F., Koch, S.P., Schroeger, E., Hinrichs, H., Heinze H.J., Deserno, L., Schlagenhauf, F. (in revision). The Feedback-Related Negativity codes components of abstract inference during reward-based decision-making.

Deserno, L., Wilbertz, T., **Reiter, A.,** Horstmann, A., Neumann, J., Villringer, A., Heinze,H.-J., Schlagenhauf, F. (in press). Lateral prefrontal model-based signatures are reduced in healthy individuals with high trait impulsivity. *Translational Psychiatry*

Radenbach, C.*. **Reiter, A.M.F.***, Engert, V., Sjoerds, Z., Heinze, H.-J., Deserno, L., Schlagenhauf, F. (2015). The interaction of chronic and acute stress impairs model-based behavioral control. *Psychoneuroendocrinology. 53*, 268-280.
*equal contribution

Mussel, P., **Reiter, A.M.F**., Osinksy, R., Hewig, J. (2014). State- and trait-greed, its impact on risky decision-making and underlying neural mechanisms. *Social Neuroscience 10*(2), 126-134.

Reiter, A.M.F. (2013). Vom Homo Oeconomicus zum Homo Emotionalis. *Marketing intern 3,* 32-33.

Talks

Reiter, A.M.F, Deserno, L., Heinze, HJ, Schlagenhauf, F. (2015). Goals and habits in addictive disorders. Talk presented at the 57th Conference of Experimental Psychologists (Tagung experimentell arbeitender Psychologen, TeaP) University of Hildesheim, Germany, 2015-03-09

Reiter, A.M.F & Schlagenhauf, F. (2014). Störungen basaler Lernmechanismen bei Alkoholabhängigkeit, Talk presented at HELIOS Park-Klinikum Leipzig, Germany, 2014-12-03

Reiter, A.M.F. (2014). Behaviors running out of control? (Neuro-)clinical investigations on behavioral adaptation. Invited Talk presented at colloquium, Research Group Motivation and Emotion, Julius-Maximilians-Universität Würzburg, Germany, 2014-02-09

Poster Presentations

Schaare, L., Gaebler, M., Kumral, D., Reinelt, J., Erbey, M., **Reiter, A.,** Röbbig, J., Babayan, A. & Villringer, A. (2015). Higher blood pressure is associated with lower regional grey matter density in healthy, young adults. Poster presented at the Conference of the International Society for Autonomic Neuroscience, Stresa, Italy.

Gaebler, M., Kumral, D., Reinelt, J., Erbey, M., **Reiter, A**., Röbbig, J., Babayan, A. & Villringer, A. (2015): Allostatic load and its connection to the brain. Poster presented at the 45th meeting of the International Society for Psychoneuroendocrinology, Edinburgh, Scotland.

Reinelt, J., Kumral, D., Erbey, M., Röbbig, J., **Reiter, A.,** Schaare, H. L., Babayan, A., Villringer, A. & Gaebler, M. (2015): Neural dynamics of stress recovery and their relation to hormonal, cardiac, and subjective changes. Poster presented at the 45th meeting of the International Society for Psychoneuroendocrinology, Edinburgh, Scotland.

Reiter, A.M.F., Radenbach, C., Sjoerds, Z., Engert, V., Kallert, T., Heinz, A., Villringer, A., Heinze H.J., Deserno L., Schlagenhauf, F. (2015): Failure Modes of the Will - from goals to habits in addictive disorders and their risk factors. Poster presented at 5th IMPRS NeuroCom Summer School, Leipzig, Germany

Reiter, A.M.F., Deserno, L., Kallert, T., Heinze, H.J., Heinz, A., Schlagenhauf F. (2015): Neglecting what might have happened – disturbed inference on alternative choices in alcohol-dependent patients. Poster presented at the 21st annual meeting of the Organization of Human Brain Mapping, Honolulu, Hawaii.

Deserno L, Wilbertz T, **Reiter A**., Horstmann A, Neumann J, Villringer A, Heinze HJ, Schlagenhauf F (2015): Lateral prefrontal model-based signatures are reduced in healthy individuals with high trait impulsivity. Poster presented at 21st Annual Meeting of the Organization for Human Brain Mapping (OHBM), Honolulu, Hawaii.

Reiter, A.M.F., Deserno, L., Schlagenhauf, F. (2014). Dissecting Learning from Reward and Punishment. Poster presented at Summer School on Computational Modeling of Cognition, Laufen, Germany.

Reiter, A.M.F., Deserno, L. Sjoerds, Z., Radenbach, C.D., Heinze, H.J., Schlagenhauf F.(2014). Learning to Crave: Appetitive Pavlovian Conditioning in Addictive Disorders. Poster presented at 4th IMPRS NeuroCom Summer School, London, United Kingdom.

Schaare, H. L., Rohr, C., Mueller, K., Margulies, D. S., Pampel, A., Erbey, M., Gaebler, M., Reinelt, J., **Reiter, A.,** Röbbig, J., Sacher, J., Dreyer, M., Okon-Singer, H., Babayan, A., & Villringer, A. (2014). Less grey matter density in young adults' frontal lobes is associated with higher blood pressure. Poster presented at 4th IMPRS NeuroCom Summer School, London, United Kingdom.

Reiter, A.M.F., Deserno, L. Sjoerds, Z., Wilbertz, T., Heinze, H.J., Schlagenhauf F.(2014): Transdiagnostic Investigation of Learning Mechanisms in Patients with Failure of Behavioral Adaptation. Poster presented at Society of Biological Psychiatry's 69th Annual Meeting, New York City, USA.

Reiter, A.M.F., Deserno, L. Sjoerds, Z., Wilbertz, T., Heinze, H.J., Schlagenhauf F. (2014): Transdiagnostic Investigation of Learning Mechanisms in Patients with Failure of Behavioral Control. Poster presented at 20th Annual Meeting of the Organization for Human Brain Mapping (OHBM), Hamburg, Germany.

Schaare, L., Rohr, C., Margulies, D., Pampel, A., Erbey, M., **Reiter, A**., Roebbig, J., Dreyer, D., Babayan, A., Villringer, A. (2014). Resting State Functional Connectivity Patterns of Blood Pressure: Links to Emotional Dampening? Poster presented at 20th Annual Meeting of the Organization for Human Brain Mapping (OHBM), Hamburg, Germany

Erbey, M., Nierhaus, T., Schaare, L., **Reiter, A**., Roebbig, J., Rohr, C. Reinelt, J., Sacher, J., Babayan, A., Villringer, A. (2014). An eye tracker study on the positivity effect and its cognitive correlates. Poster presented at the Mind, Brain & Body Symposium, Berlin, Germany.

Radenbach, C., **Reiter, A.M.F**., Deserno, L., Engert, V., Wilbertz, T., Villringer, A., Heinze, H.-J., Schlagenhauf, F. (2013). Influences of acute stress on model-based versus model-free choice behavior. Poster presented at the annual conference of Deutsche Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenheilkunde, Berlin, Germany.

Reiter, A.M.F., Koch, S.P., Deserno, L., Klein, T.A., Heinze, H.-J., Schlagenhauf, F. (2013).The Feedback Related Negativity – An Electrophysiological Correlate of Model-Based Decision-Making. Poster presented at the 3rd IMPRS Summerschool, Leipzig, Germany.

Erklärung gemäß § 8 der Promotionsordnung

Hiermit versichere ich, dass die vorliegende Arbeit ohne unzulässige Hilfsmittel und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt wurde. Aus fremden Quellen direkt oder indirekt übernommene Gedanken sind in der Arbeit als solche kenntlich gemacht. Diese Arbeit enthält eine Liste, in der alle Ko-autoren Art und Umfang ihres Anteils an der wissenschaftlichen Leistung der in dieser Promotion enthaltenen Publikationen angeben. Ich versichere, dass außer den hier genannten Personen keine weiteren Personen bei der geistigen Herstellung dieser Arbeit beteiligt waren, insbesondere auch nicht die Hilfe eines Promotionsberaters in Anspruch genommen wurde. Ich versichere, dass kein Dritter unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten hat, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die vorgelegte Arbeit wurde nicht in gleicher oder ähnlicher Form einer anderen wissenschaftlichen Einrichtung zum Zwecke der Promotion oder eines anderen Prüfungsverfahrens vorgelegt. Es gab keine früheren erfolglosen Promotionsversuche.

Die Promotionsordnung vom 29.04.2015 ist bekannt und wird anerkannt.

Leipzig, den 16.10.2015 Andrea Maria Franziska Reiter

MPI Series in Human Cognitive and Brain Sciences:

1 Anja Hahne
Charakteristika syntaktischer und semantischer Prozesse bei der auditiven Sprachverarbeitung: Evidenz aus ereigniskorrelierten Potentialstudien

2 Ricarda Schubotz
Erinnern kurzer Zeitdauern: Behaviorale und neurophysiologische Korrelate einer Arbeitsgedächtnisfunktion

3 Volker Bosch
Das Halten von Information im Arbeitsgedächtnis: Dissoziationen langsamer corticaler Potentiale

4 Jorge Jovicich
An investigation of the use of Gradient- and Spin-Echo (GRASE) imaging for functional MRI of the human brain

5 Rosemary C. Dymond
Spatial Specificity and Temporal Accuracy in Functional Magnetic Resonance Investigations

6 Stefan Zysset
Eine experimentalpsychologische Studie zu Gedächtnisabrufprozessen unter Verwendung der funktionellen Magnetresonanztomographie

7 Ulrich Hartmann
Ein mechanisches Finite-Elemente-Modell des menschlichen Kopfes

8 Bertram Opitz
Funktionelle Neuroanatomie der Verarbeitung einfacher und komplexer akustischer Reize: Integration haemodynamischer und elektrophysiologischer Maße

9 Gisela Müller-Plath
Formale Modellierung visueller Suchstrategien mit Anwendungen bei der Lokalisation von Hirnfunktionen und in der Diagnostik von Aufmerksamkeitsstörungen

10 Thomas Jacobsen
Characteristics of processing morphological structural and inherent case in language comprehension

11 Stefan Kölsch
Brain and Music
A contribution to the investigation of central auditory processing with a new electrophysiological approach

12 Stefan Frisch
Verb-Argument-Struktur, Kasus und thematische Interpretation beim Sprachverstehen

13 Markus Ullsperger
The role of retrieval inhibition in directed forgetting – an event-related brain potential analysis

14 Martin Koch
Measurement of the Self-Diffusion Tensor of Water in the Human Brain

15 Axel Hutt
Methoden zur Untersuchung der Dynamik raumzeitlicher Signale

16 Frithjof Kruggel
Detektion und Quantifizierung von Hirnaktivität mit der funktionellen Magnetresonanztomographie

17 Anja Dove
Lokalisierung an internen Kontrollprozessen beteiligter Hirngebiete mithilfe des Aufgabenwechselparadigmas und der ereigniskorrelierten funktionellen Magnetresonanztomographie

18 Karsten Steinhauer
Hirnphysiologische Korrelate prosodischer Satzverarbeitung bei gesprochener und geschriebener Sprache

19 Silke Urban
Verbinformationen im Satzverstehen

20 Katja Werheid
Implizites Sequenzlernen bei Morbus Parkinson

21 Doreen Nessler
Is it Memory or Illusion? Electrophysiological Characteristics of True and False Recognition

22 Christoph Herrmann
Die Bedeutung von 40-Hz-Oszillationen für kognitive Prozesse

23 Christian Fiebach
Working Memory and Syntax during Sentence Processing.
A neurocognitive investigation with event-related brain potentials and functional magnetic resonance imaging

24 Grit Hein
Lokalisation von Doppelaufgabendefiziten bei gesunden älteren Personen und neurologischen Patienten

25 Monica de Filippis
Die visuelle Verarbeitung unbeachteter Wörter. Ein elektrophysiologischer Ansatz

26 Ulrich Müller
Die katecholaminerge Modulation präfrontaler kognitiver Funktionen beim Menschen

27 Kristina Uhl
Kontrollfunktion des Arbeitsgedächtnisses über interferierende Information

28 Ina Bornkessel
The Argument Dependency Model: A Neurocognitive Approach to Incremental Interpretation

29 Sonja Lattner
Neurophysiologische Untersuchungen zur auditorischen Verarbeitung von Stimminformationen

30 Christin Grünewald
Die Rolle motorischer Schemata bei der Objektrepräsentation: Untersuchungen mit funktioneller Magnetresonanztomographie

31 Annett Schirmer
Emotional Speech Perception: Electrophysiological Insights into the Processing of Emotional Prosody and Word Valence in Men and Women

32 André J. Szameitat
Die Funktionalität des lateral-präfrontalen Cortex für die Verarbeitung von Doppelaufgaben

33 Susanne Wagner
Verbales Arbeitsgedächtnis und die Verarbeitung ambiger Wörter in Wort- und Satzkontexten

34 Sophie Manthey
Hirn und Handlung: Untersuchung der Handlungsrepräsentation im ventralen prämotorischen Cortex mit Hilfe der funktionellen Magnet-Resonanz-Tomographie

35 Stefan Heim
Towards a Common Neural Network Model of Language Production and Comprehension: fMRI Evidence for the Processing of Phonological and Syntactic Information in Single Words

36 Claudia Friedrich
Prosody and spoken word recognition: Behavioral and ERP correlates

37 Ulrike Lex
Sprachlateralisierung bei Rechts- und Linkshändern mit funktioneller Magnetresonanztomographie

38 Thomas Arnold
Computergestützte Befundung klinischer Elektroenzephalogramme

39 Carsten H. Wolters
Influence of Tissue Conductivity Inhomogeneity and Anisotropy on EEG/MEG based Source Localization in the Human Brain

40 Ansgar Hantsch
Fisch oder Karpfen? Lexikale Aktivierung von Benennungsalternative bei der Objektbenennung

41 Peggy Bungert
Zentralnervöse Verarbeitung akustischer Informationen Signalidentifikation, Signallateralisation und zeitgebundene Informationsverarbeitung bei Patienten mit erworbenen Hirnschädigungen

42 Daniel Senkowski
Neuronal correlates of selective attention: An investigation of electrophysiological brain responses in the EEG and MEG

43 Gert Wollny
Analysis of Changes in Temporal Series of Medical Images

S 1 Markus Ullsperger & Michael Falkenstein
Errors, Conflicts, and the Brain Current Opinions on Performance Monitoring

44 Angelika Wolf
Sprachverstehen mit Cochlea-Implantat: EKP-Studien mit postlingual ertaubten erwachsenen CI-Trägern

45 Kirsten G. Volz
Brain correlates of uncertain decisions: Types and degrees of uncertainty

46 Hagen Huttner
Magnetresonanztomographische Untersuchungen über die anatomische Variabilität des Frontallappens des menschlichen Großhirns

47 Dirk Köster
Morphology and Spoken Word Comprehension: Electrophysiological Investigations of Internal Compound Structure

48 Claudia A. Hruska
Einflüsse kontextueller und prosodischer Informationen in der auditorischen Satzverarbeitung: Untersuchungen mit ereigniskorrelierten Hirnpotentialen

49 Hannes Ruge
Eine Analyse des raum-zeitlichen Musters neuronaler Aktivierung im Aufgabenwechselparadigma zur Untersuchung handlungssteuernder Prozesse

50 Ricarda I. Schubotz
Human premotor cortex: Beyond motor performance

51 Clemens von Zerssen
Bewusstes Erinnern und falsches Wiedererkennen: Eine funktionelle MRT Studie neuroanatomischer Gedächtniskorrelate

52 Christiane Weber
Rhythm is gonna get you.
Electrophysiological markers of rhythmic processing in infants with and without risk for Specific Language Impairment (SLI)

53 Marc Schönwiesner
Functional Mapping of Basic Acoustic Parameters in the Human Central Auditory System

54 Katja Fiehler
Temporospatial characteristics of error correction

55 Britta Stolterfoht
Processing Word Order Variations and Ellipses: The Interplay of Syntax and Information Structure during Sentence Comprehension

56 Claudia Danielmeier
Neuronale Grundlagen der Interferenz zwischen Handlung und visueller Wahrnehmung

57 Margret Hund-Georgiadis
Die Organisation von Sprache und ihre Reorganisation bei ausgewählten, neurologischen Erkrankungen gemessen mit funktioneller Magnetresonanztomographie – Einflüsse von Händigkeit, Läsion, Performanz und Perfusion

58 Jutta L. Mueller
Mechanisms of auditory sentence comprehension in first and second language: An electrophysiological miniature grammar study

59 Franziska Biedermann
Auditorische Diskriminationsleistungen nach unilateralen Läsionen im Di- und Telenzephalon

60 Shirley-Ann Rüschemeyer
The Processing of Lexical Semantic and Syntactic Information in Spoken Sentences: Neuroimaging and Behavioral Studies of Native and Non-Native Speakers

61 Kerstin Leuckefeld
The Development of Argument Processing Mechanisms in German. An Electrophysiological Investigation with School-Aged Children and Adults

62 Axel Christian Kühn
Bestimmung der Lateralisierung von Sprachprozessen unter besondere Berücksichtigung des temporalen Cortex, gemessen mit fMRT

63 Ann Pannekamp
Prosodische Informationsverarbeitung bei normalsprachlichem und deviantem Satzmaterial: Untersuchungen mit ereigniskorrelierten Hirnpotentialen

64 Jan Derrfuß
Functional specialization in the lateral frontal cortex: The role of the inferior frontal junction in cognitive control

65 Andrea Mona Philipp
The cognitive representation of tasks – Exploring the role of response modalities using the task-switching paradigm

66 Ulrike Toepel
Contrastive Topic and Focus Information in Discourse – Prosodic Realisation and Electrophysiological Brain Correlates

67 Karsten Müller
Die Anwendung von Spektral- und Waveletanalyse zur Untersuchung der Dynamik von BOLD-Zeitreihen verschiedener Hirnareale

68 Sonja A.Kotz
The role of the basal ganglia in auditory language processing: Evidence from ERP lesion studies and functional neuroimaging

69 Sonja Rossi
The role of proficiency in syntactic second language processing: Evidence from event-related brain potentials in German and Italian

70 Birte U. Forstmann
Behavioral and neural correlates of endogenous control processes in task switching

71 Silke Paulmann
Electrophysiological Evidence on the Processing of Emotional Prosody: Insights from Healthy and Patient Populations

72 Matthias L. Schroeter
Enlightening the Brain – Optical Imaging in Cognitive Neuroscience

73 Julia Reinholz
Interhemispheric interaction in object- and word-related visual areas

74 Evelyn C. Ferstl
The Functional Neuroanatomy of Text Comprehension

75 Miriam Gade
Aufgabeninhibition als Mechanismus der Konfliktreduktion zwischen Aufgabenrepräsentationen

76 Juliane Hofmann
Phonological, Morphological, and Semantic Aspects of Grammatical Gender Processing in German

77 Petra Augurzky
Attaching Relative Clauses in German – The Role of Implicit and Explicit Prosody in Sentence Processing

78 Uta Wolfensteller
Habituelle und arbiträre sensomotorische Verknüpfungen im lateralen prämotorischen Kortex des Menschen

79 Päivi Sivonen
Event-related brain activation in speech perception: From sensory to cognitive processes

80 Yun Nan
Music phrase structure perception: the neural basis, the effects of acculturation and musical training

81 Katrin Schulze
Neural Correlates of Working Memory for Verbal and Tonal Stimuli in Nonmusicians and Musicians With and Without Absolute Pitch

82 Korinna Eckstein
Interaktion von Syntax und Prosodie beim Sprachverstehen: Untersuchungen anhand ereigniskorrelierter Hirnpotentiale

83 Florian Th. Siebörger
Funktionelle Neuroanatomie des Textverstehens: Kohärenzbildung bei Witzen und anderen ungewöhnlichen Texten

84 Diana Böttger
Aktivität im Gamma-Frequenzbereich des EEG: Einfluss demographischer Faktoren und kognitiver Korrelate

85 Jörg Bahlmann
Neural correlates of the processing of linear and hierarchical artificial grammar rules: Electrophysiological and neuroimaging studies

86 Jan Zwickel
Specific Interference Effects Between Temporally Overlapping Action and Perception

87 Markus Ullsperger
Functional Neuroanatomy of Performance Monitoring: fMRI, ERP, and Patient Studies

88 Susanne Dietrich
Vom Brüllen zum Wort – MRT-Studien zur kognitiven Verarbeitung emotionaler Vokalisationen

89 Maren Schmidt-Kassow
What's Beat got to do with ist? The Influence of Meter on Syntactic Processing: ERP Evidence from Healthy and Patient populations

90 Monika Lück
Die Verarbeitung morphologisch komplexer Wörter bei Kindern im Schulalter: Neurophysiologische Korrelate der Entwicklung

91 Diana P. Szameitat
Perzeption und akustische Eigenschaften von Emotionen in menschlichem Lachen

92 Beate Sabisch
Mechanisms of auditory sentence comprehension in children with specific language impairment and children with developmental dyslexia: A neurophysiological investigation

93 Regine Oberecker
Grammatikverarbeitung im Kindesalter: EKP-Studien zum auditorischen Satzverstehen

94 Şükrü Barış Demiral
Incremental Argument Interpretation in Turkish Sentence Comprehension

95 Henning Holle
The Comprehension of Co-Speech Iconic Gestures: Behavioral, Electrophysiological and Neuroimaging Studies

96 Marcel Braß
Das inferior frontale Kreuzungsareal und seine Rolle bei der kognitiven Kontrolle unseres Verhaltens

97 Anna S. Hasting
Syntax in a blink: Early and automatic processing of syntactic rules as revealed by event-related brain potentials

98 Sebastian Jentschke
Neural Correlates of Processing Syntax in Music and Language – Influences of Development, Musical Training and Language Impairment

99 Amelie Mahlstedt
The Acquisition of Case marking Information as a Cue to Argument Interpretation in German
An Electrophysiological Investigation with Pre-school Children

100 Nikolaus Steinbeis
Investigating the meaning of music using EEG and fMRI

101 Tilmann A. Klein
Learning from errors: Genetic evidence for a central role of dopamine in human performance monitoring

102 Franziska Maria Korb
Die funktionelle Spezialisierung des lateralen präfrontalen Cortex: Untersuchungen mittels funktioneller Magnetresonanztomographie

103 Sonja Fleischhauer
Neuronale Verarbeitung emotionaler Prosodie und Syntax: die Rolle des verbalen Arbeitsgedächtnisses

104 Friederike Sophie Haupt
The component mapping problem: An investigation of grammatical function reanalysis in differing experimental contexts using eventrelated brain potentials

105 Jens Brauer
Functional development and structural maturation in the brain's neural network underlying language comprehension

106 Philipp Kanske
Exploring executive attention in emotion: ERP and fMRI evidence

107 Julia Grieser Painter
Music, meaning, and a semantic space for musical sounds

108 Daniela Sammler
The Neuroanatomical Overlap of Syntax Processing in Music and Language - Evidence from Lesion and Intracranial ERP Studies

109 Norbert Zmyj
Selective Imitation in One-Year-Olds: How a Model's Characteristics Influence Imitation

110 Thomas Fritz
Emotion investigated with music of variable valence – neurophysiology and cultural influence

111 Stefanie Regel
The comprehension of figurative language: Electrophysiological evidence on the processing of irony

112 Miriam Beisert
Transformation Rules in Tool Use

113 Veronika Krieghoff
Neural correlates of Intentional Actions

114 Andreja Bubić
Violation of expectations in sequence processing

115 Claudia Männel
Prosodic processing during language acquisition: Electrophysiological studies on intonational phrase processing

116 Konstanze Albrecht
Brain correlates of cognitive processes underlying intertemporal choice for self and other

117 Katrin Sakreida
Nicht-motorische Funktionen des prämotorischen Kortex: Patientenstudien und funktionelle Bildgebung

118 Susann Wolff
The interplay of free word order and pro-drop in incremental sentence processing: Neurophysiological evidence from Japanese

119 Tim Raettig
The Cortical Infrastructure of Language Processing: Evidence from Functional and Anatomical Neuroimaging

120 Maria Golde
Premotor cortex contributions to abstract and action-related relational processing

121 Daniel S. Margulies
Resting-State Functional Connectivity fMRI: A new approach for assessing functional neuroanatomy in humans with applications to neuroanatomical, developmental and clinical questions

122 Franziska Süß
The interplay between attention and syntactic processes in the adult and developing brain: ERP evidences

123 Stefan Bode
From stimuli to motor responses: Decoding rules and decision mechanisms in the human brain

124 Christiane Diefenbach
Interactions between sentence comprehension and concurrent action: The role of movement effects and timing

125 Moritz M. Daum
Mechanismen der frühkindlichen Entwicklung des Handlungsverständnisses

126 Jürgen Dukart
Contribution of FDG-PET and MRI to improve Understanding, Detection and Differentiation of Dementia

127 Kamal Kumar Choudhary
Incremental Argument Interpretation in a Split Ergative Language: Neurophysiological Evidence from Hindi

128 Peggy Sparenberg
Filling the Gap: Temporal and Motor Aspects of the Mental Simulation of Occluded Actions

129 Luming Wang
The Influence of Animacy and Context on Word Order Processing: Neurophysiological Evidence from Mandarin Chinese

130 Barbara Ettrich
Beeinträchtigung frontomedianer Funktionen bei Schädel-Hirn-Trauma

131 Sandra Dietrich
Coordination of Unimanual Continuous Movements with External Events

132 R. Muralikrishnan
An Electrophysiological Investigation Of Tamil Dative-Subject Constructions

133 Christian Obermeier
Exploring the significance of task, timing and background noise on gesture-speech integration

134 Björn Herrmann
Grammar and perception: Dissociation of early auditory processes in the brain

135 Eugenia Solano-Castiella
In vivo anatomical segmentation of the human amygdala and parcellation of emotional processing

136 Marco Taubert
Plastizität im sensomotorischen System – Lerninduzierte Veränderungen in der Struktur und Funktion des menschlichen Gehirns

137 Patricia Garrido Vásquez
Emotion Processing in Parkinson's Disease: The Role of Motor Symptom Asymmetry

138 Michael Schwartze
Adaptation to temporal structure

139 Christine S. Schipke
Processing Mechanisms of Argument Structure and Case-marking in Child Development: Neural Correlates and Behavioral Evidence

140 Sarah Jessen
Emotion Perception in the Multisensory Brain

141 Jane Neumann
Beyond activation detection: Advancing computational techniques for the analysis of functional MRI data

142 Franziska Knolle
Knowing what's next: The role of the cerebellum in generating predictions

143 Michael Skeide
Syntax and semantics networks in the developing brain

144 Sarah M. E. Gierhan
Brain networks for language
Anatomy and functional roles of neural pathways supporting language comprehension and repetition

145 Lars Meyer
The Working Memory of Argument-Verb Dependencies
Spatiotemporal Brain Dynamics during Sentence Processing

146 Benjamin Stahl
Treatment of Non-Fluent Aphasia through Melody, Rhythm and Formulaic Language

147 Kathrin Rothermich
The rhythm's gonna get you: ERP and fMRI evidence on the interaction of metric and semantic processing

148 Julia Merrill
Song and Speech Perception – Evidence from fMRI, Lesion Studies and Musical Disorder

149 Klaus-Martin Krönke
Learning by Doing?
Gesture-Based Word-Learning and its Neural Correlates in Healthy Volunteers and Patients with Residual Aphasia

150 Lisa Joana Knoll
When the hedgehog kisses the frog
A functional and structural investigation of syntactic processing in the developing brain

151 Nadine Diersch
Action prediction in the aging mind

152 Thomas Dolk
A Referential Coding Account for the Social Simon Effect

153 Mareike Bacha-Trams
Neurotransmitter receptor distribution in Broca's area and the posterior superior temporal gyrus

154 Andrea Michaela Walter
The role of goal representations in action control

155 Anne Keitel
Action perception in development: The role of experience

156 Iris Nikola Knierim
Rules don't come easy: Investigating feedback-based learning of phonotactic rules in language.

157 Jan Schreiber
Plausibility Tracking: A method to evaluate anatomical connectivity and microstructural properties along fiber pathways

158 Katja Macher
Die Beteiligung des Cerebellums am verbalen Arbeitsgedächtnis

159 Julia Erb
The neural dynamics of perceptual adaptation to degraded speech

160 Philipp Kanske
Neural bases of emotional processing in affective disorders

161 David Moreno-Dominguez
Whole-brain cortical parcellation: A hierarchical method based on dMRI tractography

162 Maria Christine van der Steen
Temporal adaptation and anticipation mechanisms in sensorimotor synchronization

163 Antje Strauß
Neural oscillatory dynamics of spoken word recognition

164 Jonas Obleser
The brain dynamics of comprehending degraded speech

165 Corinna E. Bonhage
Memory and Prediction in Sentence Processing

S 2 Tania Singer, Bethany E. Kok, Boris Bornemann, Matthias Bolz, and Christina A. Bochow
The Resource Project
Background, Design, Samples, and Measurements

166 Anna Wilsch
Neural oscillations in auditory working memory

167 Dominique Goltz
Sustained Spatial Attention in Touch: Underlying Brain Areas and Their Interaction

168 Juliane Dinse
A Model-Based Cortical Parcellation Scheme for High-Resolution 7 Tesla MRI Data

169 Gesa Schaadt
Visual, Auditory, and Visual-Auditory Speech Processing in School Children with Writing Difficulties

170 Laura Verga
Learning together or learning alone: Investigating the role of social interaction in second language word learning

171 Eva Maria Quinque
Brain, mood and cognition in hypothyroidism

172 Malte Wöstmann
Neural dynamics of selective attention to speech in noise

173 Charles-Étienne Benoit
Music-based gait rehabilitation in Parkinson's disease

174 Anja Fengler
How the Brain Attunes to Sentence Processing Relating Behavior, Structure, and Function

175 Emiliano Zaccarella
Breaking Down Complexity: The Neural Basis of the Syntactic Merge Mechanism in the Human Brain

S 2 Tania Singer, Bethany E. Kok, Boris Bornemann, Sandra Zurborg, Matthias Bolz, and Christina Bochow
The Resource Project
Background, Design, Samples, and Measurements
Second Edition

176 Manja Attig
Handlungsverständnis in den ersten Lebensjahren: retrospektive und prospektive Verarbeitung