Aus dem Zentralinstitut für seelische Gesundheit
der Medizinischen Fakultät Mannheim
Direktor: Prof. Dr. med. Andreas Meyer-Lindenberg

Cognitive Flexibility in Humans: a translational magnetoencephalographic study

Inauguraldissertation
zur Erlangung des medizinischen Doktorengrades
der
Medizinischen Fakultät Mannheim
der Ruprecht-Karls-Universität
zu
Heidelberg

vorgelegt von
Hélène Laura Valentina Zingone
aus Courbevoie

2024

Dekan: Prof. Dr. med. Sergij Goerdt
Referent: Prof. Dr. med. Andreas Meyer-Lindenberg

Abstract

Cognitive flexibility is characterized by the ability to adapt cognition and behavior to an ever-changing environment. While much is known about the importance of cognitive flexibility and disorders implicated with its dysfunction, its underlying mechanisms remain vastly unrevealed. The goal of this thesis is to uncover a part of the neural and behavioral mechanisms involved in the correct functioning of cognitive flexibility in humans. This will be achieved using a novel multidimensional rule-switching learning paradigm to analyze behavioral mechanisms and simultaneously recorded magnetoencephalographic data. Further, this thesis will aim to investigate the relationship between selective attention markers and efficient learning as well as validate neural and behavioral results by demonstrating correlations with externally validated neuropsychological assessments.

Table of Contents

Acronyms **1**

1 Introduction **3**
1.1 Cognitive Flexibility . . . 3
1.1.1 Definition . . . 4
1.1.2 Implicated Brain Regions . . . 5
1.1.3 Measuring Cognitive Flexibility . . . 7
1.1.4 Reinforcement Learning . . . 10
1.1.5 *Eureka!* - Sudden Insight in Learning . . . 11
1.2 Magnetoencephalography . . . 12
1.2.1 Basic Principles of *MEG* . . . 12
1.2.2 *MEG* Design . . . 12
1.2.3 Spatial and Temporal range of *MEG* . . . 15
1.2.4 Advantages and Disadvantages of *MEG* . . . 16
1.2.5 Current Clinical Applications of *MEG* . . . 17
1.3 Temporal Generalization Matrix . . . 17
1.3.1 *MVPA* with *MRI* data . . . 18
1.3.2 *MPVA* and *MEG* data . . . 20
1.3.3 Time x Time Generalization . . . 20
1.4 Goals of this Thesis . . . 21

2 Method and Materials **23**
2.1 Method Outline . . . 23
2.1.1 Pilot Experiment . . . 23
2.1.2 Main Experiment Combining Neural and Behavioral Data . . . 25
2.1.3 *MEG* Data Aquisition . . . 27
2.1.4 *CANTAB* Data Aquisition . . . 28

2.2 Analysis Outline . . . 29
2.2.1 Behavioral Data Analysis . . . 29
2.2.2 *MEG* Data Analysis . . . 34
2.2.3 *Biochannel* Data Analysis . . . 35
2.2.4 Population *MEG* Data Analyses . . . 36
2.2.5 Correlation Analyses of Topographical Accuracy Maps with Behavioral Variables . . . 37

3 Results **39**
3.1 Behavioral Analysis . . . 39
3.1.1 Qualitative Results . . . 39
3.1.2 Strategy Detection . . . 40
3.1.3 Change Points . . . 41
3.1.4 Correlation of behavioral variables with *CANTAB* performance parameters . . . 45
3.2 *MEG* data Analysis . . . 46
3.2.1 Single-participant level . . . 47
3.2.2 Group-level Analyses . . . 50
3.2.3 Conclusion - Decoding Analyses . . . 52
3.3 Correlation of Behavioral Variables with *MEG* Data . . . 53
3.3.1 Reaction Time *RTI* . . . 57
3.3.2 Rapid Visual Processing *RVP* . . . 58
3.3.3 Spatial Working Memory *SWM* . . . 60
3.3.4 One Touch Stockings of Cambridge *OTS* . . . 63
3.3.5 Intra- Extra- Dimensional Set Shift *IED* . . . 63

4 Discussion **65**
4.1 Do humans use strategies? . . . 65
4.1.1 Qualitative Behavioral Analysis . . . 65
4.1.2 Quantitative Results . . . 66
4.2 Are learning processes shaped by abrupt transitions? . . . 67
4.2.1 Abrupt Behavioral Shifts . . . 67
4.3 How is learning guided by attention using strategies? . . . 68
4.3.1 Behavioral Results . . . 68
4.3.2 Neural decoding accross time . . . 69
4.3.3 Sensor-level analysis of attention-specific strategy classes . . . 70

4.4 Does decoding accuracy correlate with validated neuropsychological assessment in coherent sensor spaces? 71
4.4.1 Reaction Time *(RTI)* 71
4.4.2 Rapid Visual Processing *(RVP)* 72
4.4.3 Spatial Working Memory *(SWM)* 73
4.4.4 One Touch Stockings of Cambridge *(OTS)* 73
4.4.5 Intra- Extra- Dimensional Set Shift *(IED)* 74
4.4.6 Conclusions to draw from the correlation of topographical decoding accuracies and *CANTAB* performance 75
4.5 Translation of rodent results and outlook 78

5 Summary **79**

Figures **81**

Tables **83**

Bibliography **85**

Curriculum Vitae

Acknowledgements

Acronyms

CANTAB	Cambridge Neuropsychological Test Automated Battery
CT	Computed Tomography
dlPFC	Dorsolateral Prefrontal Cortex
ECG	Electrocardiography
EEG	Electroencephalography
EOG	Electrooculography
ICA	Independent Component Analysis
IED	Intra/ Extradimensional Set Shift
FEF	Frontal Eye Field
fMRI	Functional Magnetic Resonance Imaging
IPS	Intraparietal Sulcus
LDA	Linear Discriminant Analysis
MEG	Magnetoencephalography
MVPA	Multivariate Pattern Analysis
MRI	Magnetic Resonance Imaging
OTS	One Touch Stocking of Cambridge
PAL	Paired Associates Learning
PET	Positron Emission Tomography

RTI	Reaction Time
RVP	Rapid Visual Processing
SD	Standard Deviation
SSP	Spatial Span
SSS	Signal Space Separation
SST	Stop Signal Task
TPJ	Temporoparietal Junction
TGM	Temporal Generalization Matrix
SVM	Support Vector Machine
SWM	Spatial Working Memory
vlPFC	Ventrolateral Prefrontal Cortex
WCST	Wisconsin Card Sorting Test

Chapter 1

Introduction

1.1 Cognitive Flexibility

Cognitive flexibility is essential to survival, as it permits adaptation of cognition and behavior to an ever-changing environment. It can be characterized as a complex contruct comprising multiple executive functions. These enable generation and subsequent evaluation of concurrent concepts about the best course of action, as well as shifting between concepts to carry out actions most likely yielding success when confronted with continuously changing environments.

Benefits of its correct functioning have been associated with a multitude of dayly life facets. These include better health, better quality of life (Brown & Landgraf, 2010; Davis, Marra, Najafzadeh, & Liu-Ambrose, 2010), better school (de Santana, Roazzi, & Nobre, 2022; Engel de Abreu et al., 2014) and job outcomes (Bailey, 2007), better marital outcomes (Eakin et al., 2004), emotion regulation (Pruessner, Barnow, Holt, Joormann, & Schulze, 2020), and an overall better response to stress (Kruczek, Basińska, & Janicka, 2020; Borzyszkowska & Basińska, 2020). It is further seen as an important feature of creativity (Preiss, 2022), understanding other individuals' emotions (Bock, Gallaway, & Hund, 2015), as well as contributing to better mood and happiness (Hirt, Devers, & McCrea, 2008) and higher resilience (Ram, Chandran, Sadar, & Gowdappa, 2019).

These qualities do not only benefit individuals but also society. By enforcing traits such as resilience to stressful factors, benefiting flexible and efficient learning, as well as promoting creativity, emotional, and intelectual intelligence, cognitive flexibility

enables an overall well-being. Together, these qualities allow for a better coexistence in, and better contribution to society (Beddington et al., 2008).

Much as cognitive flexibility is pertinent to mental health and well-being, its dysfunction has been repeatedly shown to affect a range of neurological and psychological disorders. These include schizophrenia (Manoach, 2017; Ravizza, Moua, Long, & Carter, 2009), obsessive compulsive disorder (Gruner & Pittenger, 2016; Chamberlain, Solly, Hook, Vaghi, & Robbins, 2021), anxiety (C. Wilson, Nusbaum, Whitney, & Hinson, 2018; Park & Moghaddam, 2016), depression (Stange, Alloy, & Fresco, 2017), attention deficit hyperactivity disorder (Diamond, 2005), autism spectrum disorder (D'Cruz et al., 2013), eating disorders such as anorexia nervosa (Brockmeyer et al., 2022) or obesity (Perpiñá, Segura, & Sanchez-Reales, 2017), addiction (Baler & Volkow, 2007), post-traumatic stess disorder (Ben-Zion et al., 2018), amyotrophic lateral sclerosis (Evans et al., 2015), parkinson's disease (Dirnberger & Jahanshahi, 2013), mild cognitive impairment (Kirova, Bays, & Lagalwar, 2015) and alzheimer's disease (Guarino et al., 2019). Cognitive flexibility has also been show to behave dynamically across individuals' life spans, developing during childhood and adolescence (Davidson, Amso, Anderson, & Diamond, 2006; Hsu & Jaeggi, 2013), and detering in aging individuals (Verhaeghen & Cerella, 2002; Magnusson & Brim, 2014).

1.1.1 Definition

Cognitive flexibility is a broad construct. It can be defined as a product of reciprocal interactions of multiple intertwined executive functions. Executive functions underlying correct functioning of cognitive flexibility include salience detection, attention, working memory, and inhibition (Dajani & Uddin, 2015; Miyake et al., 2000).

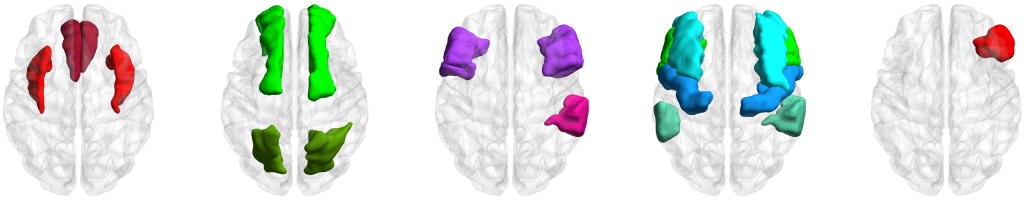
Salience detection is understood as the ability to notice potentially behaviorally relevant stimuli in order to allocate neural resources such as attention (Menon & Uddin, 2010). Attention is the ability to filter out the most goal-relevant stimuli by actively enhancing their perception (Parr & Friston, 2019). It can be divided into two entities, top-down and bottom-up attention. Top-down attention refers to intrinsic, goal-driven attention, while bottom-up attention is specified as extrinsic, stimulus-driven attention (Katsuki & Constantinidis, 2013). Working memory in-

vokes representing and manipulating information "no longer perceptually present" relevant to action planning (Diamond, 2020). Lastly, inhibition allows for previously correct responses to be inhibited, permitting updated responses to be carried out. These funcitons allow flexibly and effectively switching between behavioral responses according to context (Aron, Robbins, & Poldrack, 2014).

1.1.2 Implicated Brain Regions

Traditionally, the prefrontal cortex has been fundamentally implicated as a core brain region underlying cognitive flexibility (Milner, 1963; Alvarez & Emory, 2006; Jones & Graff-Radford, 2021). Research has further identified a range of brain regions specifically involved in correct functioning of executive functions indispensable to cognitive flexibility.

A salience network (also referred to as the midcingulo-insular network) based in the dorsal anterior cingulate cortex and the anterior insula (Uddin, 2014; Seeley et al., 2007; Uddin, Yeo, & Spreng, 2019) has been shown to influence large-scale networks allocating attention (Menon & Uddin, 2010), thus implementing saliency detection. Top-down and bottom-up attention rely on different, but functionally intertwined networks. Top-down attention is based on the dorsal attention network, primarily comprising the intraparietal sulcus and the frontal eye field. The middle temporal complex, and the inferior frontal junction have further been associated to it (Vossel, Geng, & Fink, 2013; Majerus, Peters, Bouffier, Cowan, & Phillips, 2017; Uddin et al., 2019). The ventral attention network is made up of the ventrolateral prefrontal cortex and the right temporoparietal junction (Corbetta & Shulman, 2002; Uddin et al., 2019). The executive control network (also referred to as the lateral frontoparietal network) has been liked to working memory (Uddin et al., 2019; Uddin, 2021). It consists of the ventrolateral and dorsolateral prefrontal cortices, as well as the anterior inferior parietal cortex and the premotor cortex (Thomason et al., 2008; Dajani & Uddin, 2015). Inhibition is closely linked to the right inferior frontal cortex (Aron et al., 2014). Figure 1.1 gives orientation points to localize key regions to these networks.



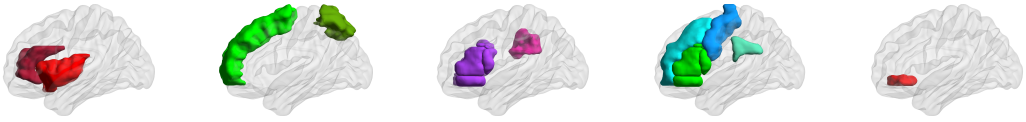
(a) Midcingulo-insular network for salience processing, axial

(b) Dorsal attention network for top-down attention, axial

(c) Ventral attention network for bottom-up attention, axial

(d) lateral frontoparietal network for executive control and working memory

(e) Right inferior cortex (red) involved in inhibition, axial



(f) 1.1a in sagittal, insula (light red) and cingulate cortex (dark red) comprising the anterior cingulate cortex

(g) 1.1b in sagittal, superior frontal gyrus (light green) with FEF in posterior part, superior parietal gyrus (dark green) with IPS in caudal part

(h) 1.1c in sagittal, vlPFC (purple) and right TPJ (pink)

(i) 1.1d in sagittal, vlPFC (green), dlPFC (dorsal parts of middle frontal gyrus in turquoise), anterior inferior parietal cortex (teal) and premotor cortex (blue)

(j) 1.1e in sagittel, right inferior frontal cortex

Figure 1.1: Key regions of networks involved in cognitive processes: Figures 1.1a and 1.1f show key regions of the midcingulo-insular network, involved in salience detection. Figures 1.1b, 1.1g, 1.1c, and 1.1h show key regions of respectively, the dorsal and ventral attention networks. Figures 1.1d and 1.1i show key regions of the lateral frontoparietal network, also referred to as the executive control network, involved in working memory. The right inferior cortex mainly involved in inhibition is shown in figures 1.1e and 1.1j. BrainNet Viewer (Xia et al., 2013) was used in Matlab 2022b to plot regions of interest. The automated anatomical labelling atlas version 3 (Rolls et al., 2019) was used as mapping file for all plots except for 1.1a and 1.1f, where version 2 was used (Rolls et al., 2015). ICBM152 (JC et al., 2001) was used as surface file. *(FEF: frontal eye field; IPS: intraparietal sulcus; vlPFC: ventrolateral prefrontal cortex; dlPFC: dorsolateral prefrontal cortex; TPJ: temporoparietal junction)*

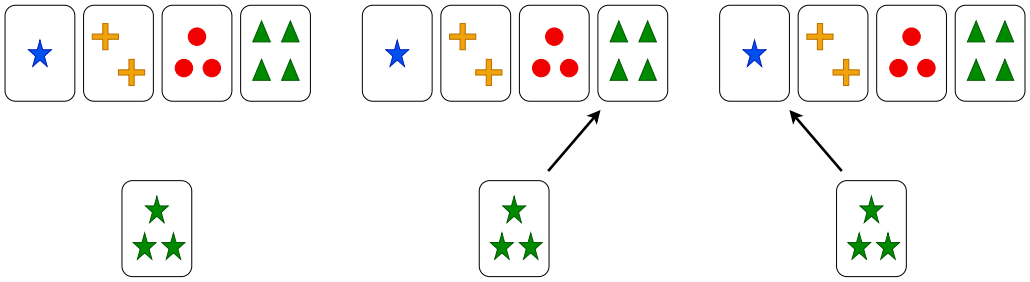
As cognitive flexibility is a broad concept, the task of defining brain regions functionally relevant to it seems more logical in defining these independently per underlying executive functions. Of course, both executive functions and brain regions described above functionally overlap and dynamically interact to produce cognitive computations (Niendam et al., 2012). Further research may be needed to pinpoint how neural dynamics and specific brain regions interact to produce cognitive computations. Current research identifies mainly the lateral frontoparietal network and the midcingulo-insular networks as most prominent in the correct functioning of cognitive flexibility (Seeley et al., 2007; Uddin, 2021).

1.1.3 Measuring Cognitive Flexibility

To measure cognitive flexibility, rule-shifting paradigms are used. Typical behavioral processes assessed are the ability to:

1. Infer a correct course of action in accordance with feedback following responses made in a given environment;
2. notice when such course of action does not yield enough positive feedback anymore;
3. incorporate new feedback to reinvestigate newly feedback-relevant stimuli and subsequent action planning;
4. update behavior to the current environment in order to maintain positive feedback yield by shifting responses and inhibting previously correct responses.

As such, set-sihfting and task-switching learning tasks, in which individuals are presented with stimuli, have to respond, obtain feedback, and must adapt behavioral responses to changing environments are used to measure these abilities. Multiple rule-switching learning paradigms are developed and used in cognitive research. Standardized and established tests are, i.e. the Wisconsin Card Sorting Test *(WCST)* (Berg, 1948; Grant & Berg, 1948; Kopp, Lange, & Steinke, 2019) and the intra-extra-dimensional task shift test of the Cambridge Neuropsychological Test Automated Battery (Downes et al., 1989), shown to be an analoge of the *WCST*.



(a) Participants are asked to match a card to one of the four presented cards.

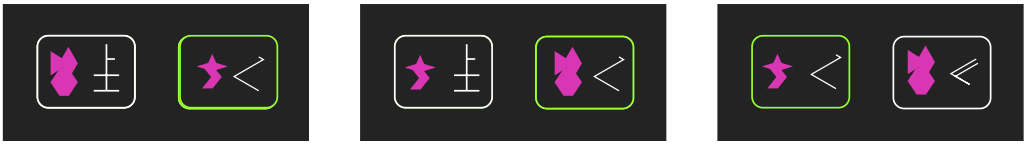
(b) The presented card could, for example, be sorted according to color.

(c) It could also be sorted according to shape, or to number (not depicted).

Figure 1.2: Wisconsin Card Sorting Test

Wisconsin Card Sorting Test Figure 1.2 exemplarily depicts the *WCST*. Participants are asked to sort a card to one of four presented cards. The card to be sorted in Figure 1.2a could be sorted according to the dimension "color" (1.2b), to "shape" (1.2c) or to "number" (not depicted). Feedback obtained to participants' sorting responses can be used to infer the correct sorting rule. After 10 correctly sorted cards, the underlying rule changes and participants have to learn a newly correct rule.

Intra- Extra- Dimensional Set Shift *CANTAB IED* presents participants with simple stimuli such as colored shapes or white lines; or compound stimuli comprising both in later trials. Using the obtained feedbach after choosing a stimulus, they must determine the rule-relevant stimulus. After six correctly selected shapes, a rule switch occurs, which can either be intra-dimensional (i.e., white lines remain the correct stimulus class) or extra-dimensional (i.e. the correct stimulus class switches from white lines to pink shapes). Figure 1.3 exemplarily shows three trials correctly completed; correctly chosen compound stimuli are marked in green. In this example, using the obtained feedback, this would infer the rule-relevant stimulus to be the isometric shape, found fourth from left in figure 1.3a).

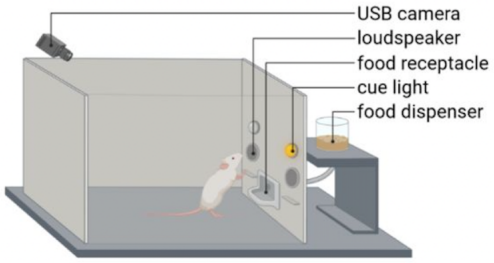


(a) Trial 1 (b) Trial 2 (c) Trial 3

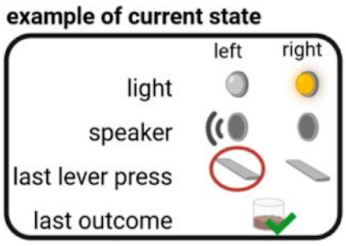
Figure 1.3: Exemplary Depiction of the *CANTAB IED* Test. Participants are asked to choose one of the two compound stimuli, the correctly chosen compound stimuli are indicated by the green outline.

Novel Rule Learning Task The novel rule learning task employed in this thesis has been adapted to humans from a rule learning task in rats. Rats were presented with simultaneously appearing auditory and visual stimuli. The stimuli were a sound and a light, either on the left- or right-hand side. The side of their appearance was pseudoramized and independent from one another. Rats could respond in form of a right- or left-hand lever press. Positive feedback according to the underlying rule was given in form of a treat after correct responses. Rules switched without additional indication other than a change in feedback. Outcome measures included reaction time, response pattern, and head inclination as a measure of attention allocation. Results showed rats learn rules on a trial-and-error basis by consecutively trying out behavioral strategies to test for the correct rule instead of learning one action each for all possible stimuli combinations. Strategies are defined here as hypotheses about the correct behavioral rule. An example of such a rule in the learning paradigm employed here could be: *"always press the lever oposite of the visual stimulus"*.

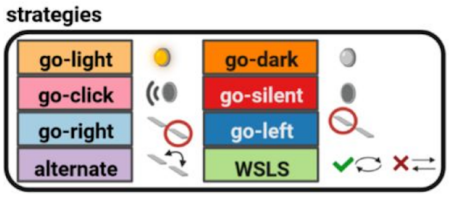
Figure 1.4 from Bähner et al., 2022 illustrates the setup of the employed paradigm in rats (1.4a), the state exemplarily resulting from the combination of used cues, last lever press, and last outcome (1.4b), as well as possible strategies (1.4c).



(a) Experiment setup



(b) Exemplary state



(c) Possible strategies in rats and their pictographic description

Figure 1.4: Rule-learning paradigm in rats, from Bähner et al., 2022

1.1.4 Reinforcement Learning

Understanding processes underlying learning is key to correctly understanding cognitive flexibility. A better comprehension of learning and crucial cognitive mechanisms may further achieve a better understanding of neuropsychological traits of a considerable number of neuropsychiatric diseases.

Learning processes are often modeled by reinforcement learning. Reinforcement learning functions by computing a representation of the environment, a state, and inferring the corresponding best course of action. Learning progresses on a trial-and-error basis. Ensuing expected and obtained reward or punishment are taken into account to update behavioral responses (Sutton & Barto, 1998; Kaelbling, Littman, & Moore, 1996), while a changing environment is also integrated into new states.

Albeit accurately explaining learning in lowdimensional environments, classic reinforcement learning frameworks get exceedingly inefficient as the number of potentially relevant dimensions increases (Niv, 2019). As such, they do not accurately

model human learning behavior in highly dimensional and ever-changing environments as our outer world, as humans and animals are well capable of completing high-dimensional learning and switching tasks.

It is often proposed humans reduce task dimensionality to maintain effective computations in high-dimensional environments. Selective attention and structure learning are suggested as a means of effectively reducing task dimensionality (R. Wilson & Niv, 2011; Leong, Radulescu, Daniel, DeWoskin, & Niv, 2017; Blakeman & Mareschal, 2022). Selective attention allocates neural and computational resources to the preselected stimulus most probably relevant to optimal action and task resolution. Structure learning implies humans find structure in sensory input through the use of episodic memory. It assumes similar tasks share similar relevant dimensions. Transferring knowledge about task structure enables efficient computations in previously unknown environments (Gershman & Niv, 2010). This shows structure learning is closely linked to the use of selective attention in learning.

1.1.5 *Eureka!* - Sudden Insight in Learning

Furthermore, classic reinforcement learning frameworks often imply a gradual learning curve. Gradual learning curves do not account for sudden behavioral transitions, sometimes referred to as moments of "sudden insight". Evidence of sudden insight shaping the learning process is found across species, with moments of sudden insights reflected by jumps in reaction time and performance metrics in mice (Rosenberg, Zhang, Perona, & Meister, 2021), rats (Durstewitz, Vittoz, Floresco, & Seamans, 2010), macaques (Bartolo & Averbeck, 2020) and in humans (Aziz-Zadeh, Kaplan, & Iacoboni, 2009; Kounios & Beeman, 2014; Donoso, Collins, & Koechlin, 2014). Sudden transitions are also explicitely found during rule learning (Durstewitz et al., 2010). Consequently, shortcomings of classic reinforcement learning frameworks include inefficiency in high-dimensional environments, resulting in a slow and gradual learning and thus contrasting with humans' efficiency in these environments, as well as the inability to explain sudden transitions in performance. Classic reinforcement learning is also unable to explain transfer effects from previously experienced learning environments to similar new ones (Bähner et al., 2022).

1.2 Magnetoencephalography

The following paragraphs introduce the neuroimaging technique used in this thesis.

1.2.1 Basic Principles of *MEG*

Magnetoencephalography *(MEG)* is a functional neuroimaging technique developed in the 1960s by David Cohen and James E. Zimmerman based on measuring the magnetic field induced by the electric current of post-synaptic potentials using sensors called magnetometers (Wöhrle, 2020). Neurons with neurites oriented in the same direction firing simultaneously elicit an electric current. This electric current induces a magnetic field as illustrated in figure 1.5, which in turn is picked up by nearby magnetometers (Baillet, 2017). The metric obtained from measurements is a magnetic field value *(Tesla)* per magnetometer per time point. It has been shown than a minimum of 10.000 to 50.000 pyramidal neurons simultaneously firing in the same direction are needed to elicit a signal (Murakami & Okada, 2006; Williamson & Kaufman, 1989).

1.2.2 *MEG* Design

One challenge in *MEG* data acquisition is the low signal-to-noise ratio. The electromagnetic background noise constituted by the earth's geomagnetic field (10^{-4} T), by surrounding magnetic sources such as cars, elevators, hospital beds, but also magnetically induced fields by electrophysiological processes such as subjects' eye blinks and heart activity is magnitudes higher than the magnetic signal of interest emitted by the brain (10^{-14}-10^{-12} T) (Singh, 2014). In addition, numerous ferromagnetic materials are able to perturb magnetic fields measures, such as i.e. dental implants, clothing articles, belts, metal bras, specific make-up or tattoo types. As such, meticulous subject selection, subject preparation, and specially developed and installed technology are needed to ensure optimal measurement. To assure a valid measurement of the comparatively weak brain signal surrounded by the strong background noise, supraconducting quantum interference devices *(SQUIDs)* able to pick up the comparably weak magnetic neural signals are combined with an electric and magnetic shielding room excluding background noise, as well as active shielding processes (Proudfoot, Woolrich, Nobre, & Turner, 2014). Further, preprocessing methods such as signal-space-separation (Taulu & Kajola, 2005) and independent component analysis (Hyvärinen, Karhunen, & Oja, 2001) are included to amplify

the sensitivity to signals of interest.

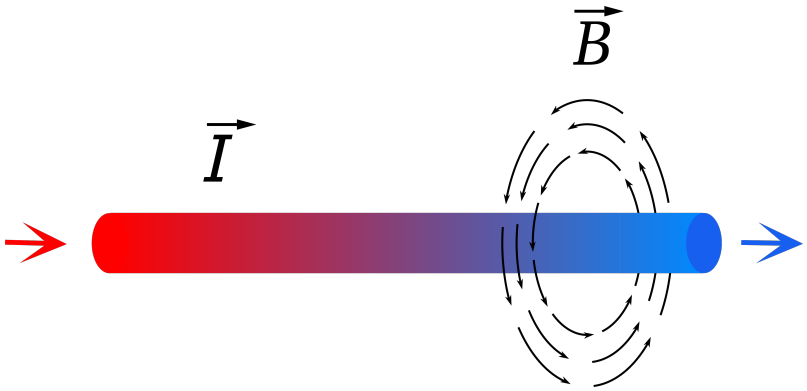
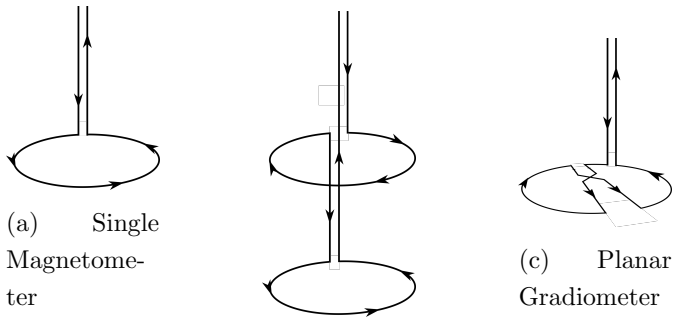


Figure 1.5: Illustration of an electric current $\vec{I}$ eliciting a magnetic field $\vec{B}$

SQUIDs *SQUIDs* make use of the Josephson effect to convert magnetic field into voltage: The supraconducting devices organized into a ring are separated by a so-called Josephson contact in two different places. The supraconducting properties over the Josephson Contact are only ensured over a critical current range. Once the critical current is exceeded, the resulting voltage over the Josephson contact drops. This fluctuation of the voltage brought about by fluctuating magnetic fields as the ones measured in *MEG* can be translated into magnetic field values. Supraconduction permits for effective and highly temporally detailed measurements of ultra-low magnetic fields (Hari & Salmelin, 2011; Wöhrle, 2020).

To conserve supraconducting qualities, *SQUIDs* require cryogenic temperatures, usually obtained by helium cooling. Magnetometers are arranged in pairs with some centimeters' distance in the form of perpendicularly arranged *(planar)* and parallelly arranged *(axial)* gradiometers. This permits a better signal-to-noise extraction, as gradiometers measure the gradient of magnetic induction between two magnetometers. Noise located further away from the gradiometers will have a more similar effect on paired magnetometers than signals of interest located more closely, which will have a less convergent effect on gradiometers arranged in pairs. Calculating the magnetic field differences in pair-related magnetometers allows for better distinction between noise and signal and better signal extraction. The gradiometers are arranged in a whole-brain helmet allowing optimal signal extraction (Ahonen et al., 1993). Typically, between 100 and 300 gradiometers are incorporated in such

a helmet. Figure 1.6 shows the spatial arrangements of magnetometer, axial, and planar gradiometer.



(a) Single Magnetometer

(b) Axial Gradiometer

(c) Planar Gradiometer

Figure 1.6: Shematic layout of magnetometers and gradiometers.

Magnetically and Electric Shielding Rooms Additionally, shielding rooms and appropriate localization of *MEG* systems contribute to lower noise levels. Magnetic and electric shielding rooms are used to suppress external electromagnetic noise. Several layers of a special metal alloy, consisting of μ-metal, are used to magnetically shield the measuring room. Electric shielding is achieved by completely surrounding the measuring environment by a metallic mesh to absorb external electromagnetic noise, making use of the *Faraday* effect (Wöhrle, 2020). Choosing an appropriate location for *MEG* measuring systems also ensures a better data quality. Typically, and also because of their very heavy apparatus, *MEG* systems are constructed in basements, preventing electromagnetic noise. Ideally, these systems would be located far away from i.e. hospitals, cities, cars and trains (Baillet, 2017).

Signal-Space-Separation* and *Independent Component Analysis Preprocessing acquired data using specific signal processing algorithms can further permit elimination of interferring noise. *Signal-space-separation* make use of the effect that externally originating noise has a specific spatio-temporal structure differing from more closely located signals of interest. Measured signals can be separated into components attributed to external and internal sources. Internal sources are located inside the sensor array, external sources are located outside of the sensor array. Both sources are linearly independent, allowing for rejection of external source signals (Taulu & Simola, 2006; MNE-Developers, 2012). This allows for differentiation of

closely located external sources, i.e. such as artefacts due to medically implanted devices in patients. *Independant component analysis (ICA)* allows for extraction of electromagnetic signals such as, i.e. saccades, eye blinks, and electrocardiographic signals. It assumes that, i.e. two sources located at different points in space both emit independent signals. Two sensors, also located at two different points in space, with one sensor more closely located to the first source, and the second located more closely to the second source will both pick up differently composed mixtures of both signals. Assuming these signals are independant, *ICA* allows for extraction of both signals. Applied to *MEG*, this allows for differentiation of electromagnetic signal such as, i.e. saccades, eye blinks, and electrocardiac signals from the signal of interest emitted by the brain.

In conclusion, the design, acquisition and subsequent processing of *MEG* data allow for signal extraction of ultra-low magnetic fields susceptible to interference in a noisy background. *MEG* signal acquisition is non-invasive, contactless, and without repercussions for measured subjects, as it only measures but does not elicit any signal that could potentially influence human physiology. This allows a wide range of applicability, i.e. fetal research, accompaniment of anxious patients, and no loud noises or (at least in comparison to *MRI* or *CT*) small spaces.

1.2.3 Spatial and Temporal range of *MEG*

Source reconstruction of *MEG* data permits spatial resolution in the range of 3-5 millimeters. Importantly, its sensitivity to neural signals and thereby spatial resolution is highly dependable of the orientation and localization of neural source: electrophysiological currents oriented axially to brain surface will emit weaker magnetic fields compared to currents vertically oriented to the brain surface (Mandal, Banerjee, Tripathi, & Sharma, 2018). Also, neural sources located deeper in the brain will require a higher signal-to-noise ratio. This implies signals located at gyral crowns will be weaker than signals located at sulcal walls (Proudfoot et al., 2014).

Regarding its temporal resolution, *MEG* data is highly precise, comprising a range of under 1 millisecond in comparison to *PET* data (ranging in tens of seconds) or *fMRI* data (ranging in seconds). This high temporal resolution is attained through the high acquisition rate of employed *SQUIDS* (Baillet, 2017; Hämäläinen, 1991) and allows measures of neural oscillation patterns.

1.2.4 Advantages and Disadvantages of *MEG*

Disadvantages of *MEG* include expensive acquisition costs, as a magnetically shielded room, liquid helium, and overall expensive materials are needed to ensure a correctly functioning *MEG*. Further, the dependency of spatial resolution on source localization and orientation constitutes a disadvantage to take into account. As stated above, meticulous subject selection and preparation are needed to ensure good quality of measured signals. Additionally, compared to neuroimaging techniques directly producing anatomical representation of measured data, raw output data in *MEG* produces a magnetic field value per channel per sampled time data point. This data structure requires knowledge in processing complex and detailed data structures, and in mathematical source reconstruction models to transform measured data into the original electrical current source and functional anatomical representation (Wöhrle, 2020). These arising techniques, requiring highly specialised knowledge, and being less beginner-friendly than more commonly known and more established neuroimaging techniques, could also constitute an advantage of *MEG*, as they enable utilizing highly detailled data, paving the path for more functionally oriented and detailled neuroimaging.

Relatively, advantages of *MEG* include that, i.e. compared to *EEG*, the measured signal is not perturbed by intermediately located brain, skull or skin tissues (Baillet, 2017). Further regarding *EEG*, no reference is needed as *MEG* signals constitute an absolute measure, and *MEG* is not dependent of subjects' head symmetry, both making the interpretation of signals easier. *MEG* can be combined with other neuroimaging techniques such as *EEG* or *MRI*, increasing its informative value (Hämäläinen, 1991). Being a non-invasive neuroimaging technique, it is deemed suitable for measurement of children and even fetal development (Preissl, Lowery, & Eswaran, 2004; Lowery, Eswaran, Murphy, & Preissl, 2007).

In comparison to other neuroimaging techniques, the magnetic field measured is a direct metric of neural activity. It is a functional and not an anatomical measuring technique, making it a whole-brain neurophysiological measuring technique (Singh, 2014). Foremost, it offers a high temporal resolution as well as an acceptable spatial resolution. In sum, this enables measurement and detailled analysis of functional connectivity patterns (Engel, Fries, & Singer, 2001; Gross, Kujala, Salmelin, & Schnitzler, 2010; Proudfoot et al., 2014).

1.2.5 Current Clinical Applications of *MEG*

Apart from being a versatile research tool, current clinical applications of *MEG* primarily consist in pre-operative brain mapping to identify functionally relevant eloquent brain areas in neurooncological and epilepsy surgery (Pang & III, 2016). As a non-invasive, functionally oriented, and temporally highly resolved measuring technique, its further application in clinical settings could be expanded to identifying aberrant neural functional connectivity patterns specifically linked to neuropsychiatric disorders at a whole-brain or local network-level. Examples of some diseases in which *MEG* has been used to characterize atypical functional connectivity patterns in research settings include epilepsy (Li Hegner et al., 2018; Van Mierlo, Höller, Focke, & Vulliemoz, 2019), autism spectrum disorder (O'Reilly, Lewis, & Elsabbagh, 2017; Roberts, Kuschner, & Edgar, 2021), anxiety, depression (Alamian, Hincapié, Combrisson, et al., 2017; Nugent et al., 2020), schizophrenia (Alamian, Hincapié, Pascarella, et al., 2017), obsessive-compulsive disorder (Zhou et al., 2022; Tan et al., 2022), Alzheimer's disease (Engels et al., 2017), and symptoms of Neuro-HIV (Becker et al., 2013; T. Wilson, Lew, Spooner, Rezich, & Wiesman, 2019).

These examples underline how *MEG* research can potentially characterize healthy functional networks in cognitive neurosciences as well as aberrant neural oscillatory dynamics in neuropsychiatric diseases. *MEG* research may contribute to establishing functional biomarkers, facilitating diagnostics and evaluation of therapeutic approaches. It also shows that presently, the capacities of *MEG* are not fully exploited for maximum yield considering the potential it holds with its rich data structure and arising knowledge in data processing (T. Wilson, Heinrichs-Graham, Proskovec, & McDermott, 2016).

1.3 Temporal Generalization Matrix

This thesis' *MEG* data analysis uses a method as presented by King & Dehaene (King & Dehaene, 2014; Dehaene & King, 2016) originating from *multivariate pattern analysis (MVPA)* in *MRI* data. The following paragraphs are meant to introduce an overall understanding of the data analysis techniques used in this thesis.

1.3.1 *MVPA* with *MRI* data

MVPA is a method that detects differences between two or more experimental conditions at a higher sensitivity rate than classical linear univariate analyses. Figure 1.7 exemplarily illustrates the principle of univariate analysis. Voxel activity measures are compared between two experimental conditions, calculating differential activity measures per voxel.

In comparison to univariate analyses, where activity rates are compared between two conditions, and the difference in activity rate is defined per voxel, *MVPA* focuses on comparing the patterns of activity over all voxels and can thus better identify differential activity patterns over all analyzed voxels (Kriegeskorte & Kievit, 2013). *MVPA* sets voxels as dimensions, expressing the activity pattern measured in condition **A** through a vector. One measurement with X voxels and their respective activity values under condition **A** thus constitutes one vector in a X-dimensional voxel space. Another measurement with the same voxels and potentially different activity values under condition **B** can be plotted in the same X-dimensional voxel space.

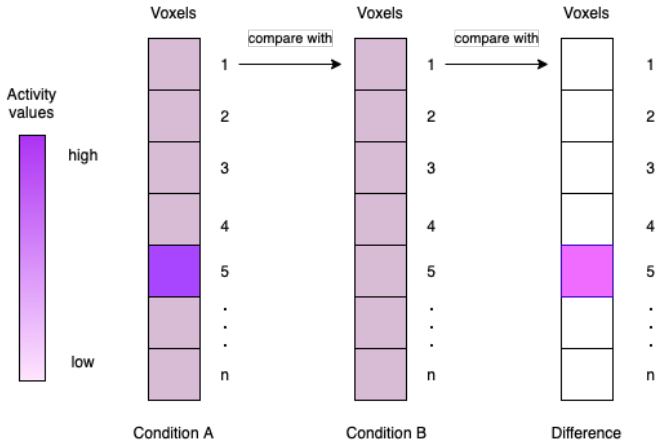
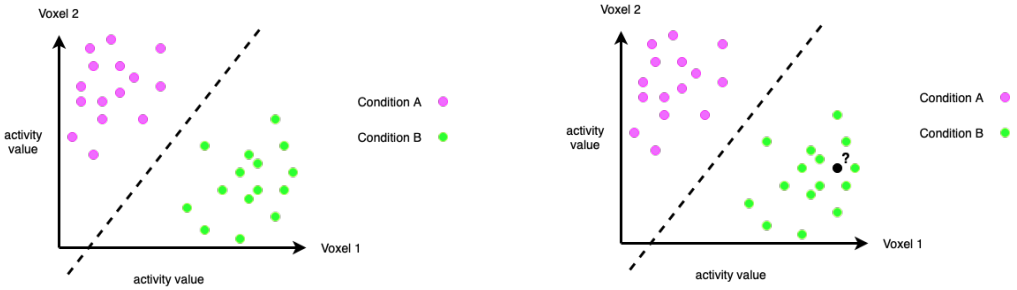


Figure 1.7: Exemplary illustration of univariate analysis: Activity parameters measured in condition **A** (left) and condition **B** (middle) are compared per voxel. The difference between conditions **A** and **B** shown right. In this exemplary depiction, voxel 5 yields different activation patterns relative to the condition.

Mathematical methods used to model the best geometric criteria to separate the conditions in X-dimensional space are i.e. *support vector machines* and *linear dis-*

criminant analysis. For example, in a 2-dimensional space, a *support vector machine* infers the optimal separation of conditions in form of an optimal line maximizing the distance between class boundaries. *Linear discriminant analysis* operates by calculating the *between-class variance* (= the distance of the means of different classes), and the *within-class variance* (= the distance between a mean and the samples of the same class); and by constructing a lower-dimensional space maximizing between-class variance and minimizing within-class variance. Both methods thus attempt to find an optimal separation of the measured conditions in the X-dimensional space and can be used as classifiers. Classifiers are trained on a given number of data points in the multidimensional voxel space and are subsequently measured on their ability to classify new data points on which they were not previously trained upon. If there is discriminative information between the activation patterns of two different conditions, a classifier trained on these two conditions should accurately classify additional data points.



(a) 30 data points comprising activity values of 2 voxels plotted as vectors in a 2-dimensional vector space. Data points emerging from condition **A** are plotted in pink, green if emerging from condition **B**. A line is fitted to best delimit both conditions, i.e. a *support vector machine* maximizing the distance between the class boundaries.

(b) The same data points as presented in subfigure 1.8a, with the dotted line marking the best fit delimiting the two conditions. The classifier (dotted line) can thus be tested on its classification accuracy regarding data points it was not trained upon, such as the black data point marked by a question mark.

Figure 1.8: Exemplary Depiction of Multivariate Pattern Analysis

This analysis offers a better discovery of differential activity patterns relative to univariate comparisons at single-voxel levels. The conditions do not have to significantly differ in the activity rate at single-voxel levels, but rather in the corresponding activity pattern (Kriegeskorte & Kievit, 2013). Figure 1.8 exemplarily depicts such

a process with two experimental conditions represented in a two-voxel activity pattern, resulting in a two-dimensional space. This technique can be completed in form of a searchlight analysis. This comprises sliding a search window over voxels in functional anatomical regions of interest and performing multivariate analysis on preselected voxels corresponding to this anatomical region of interest (Kriegeskorte, Goebel, & Bandettini, 2006). Compared to whole-brain decoding analyses allowing for detection of differential activation patterns, searchlight analyses enable localization of decoding analyses' results.

1.3.2 *MPVA* and *MEG* data

Analogously to *MVPA* classification of *MRI* data, it is possible to classify *MEG* data in an *MVPA*-style analysis. In comparison to *MRI* data consisting of *time-voxel*-pairs, MEG data consists of *time-channel*-pairs. As such, channels replace voxels as *feature* dimensions (M. Treder, Oostenveld, Schoffelen, & Janke, 2019).

The high temporal resolution in *MEG* data structure has many advantages and can be incorporated in these analyses. Data classification can be performed not over time windows of seconds such as in *fMRI*, but on all time points offered by sampling frequency (M. Treder et al., 2019), reaching up to 100 - 1000 *Hz*. This infers that per sampled time point, an individual classification analysis of channels as vectors and corresponding magnetic field values per channels can be performed (King & Dehaene, 2014; Fyshe, 2020). This permits characterization of how and whether discriminative information changes during different stages of cognitive processes such as decision-making. Further, neural patterns during cognitive processes can be identified and analyzed on their consistency of change over time.

MEG data structure offers additional possibilities; *time-frequency*-data can also be classified using this approach. Choosing channels as *feature* dimension and classifying over time identifies time points yielding the most discriminative information regarding condition classification. Inversely, an averaged time window can be chosen as feature dimension to identify channels yielding the most discriminative information (King & Dehaene, 2014; Dehaene & King, 2016).

1.3.3 Time x Time Generalization

Classifiers trained on channels and corresponding magnetic field values are tested on their ability to accurately classify data unknown to them belonging to the same

time point as well as on data of other time points the sampling rate offers. This entails classifiers trained on data belonging to time point **T** can be tested on their classification accuracy of data belonging to any time point **T'**. This can be repeated for all time points, resulting in a matrix of classification performances for all possible *train-time/test-time* pairs: a *time x time* generalization matrix (M. Treder et al., 2019; King & Dehaene, 2014). Figure 3.6a on page 49 shows, for each time point *(train time point, y-axis*) the corresponding classifier's decoding accuracy in classifying the data of another time point *(test time point, x-axis)*. The diagonal of the temporal generalization matrix gives a measure of classifiers trained and tested on the same time point (Dehaene & King, 2016).

Temporal dynamics and organization of neural codes may be inferred from the shape of the temporal generalization matrix (*tgm*) (King & Dehaene, 2014). For example, sustained activity patterns might produce a square-shaped above chance decoding performances in the *tgm* (King, Gramfort, Schurger, Naccache, & Dehaene, 2014); subsequently activated activity generators might produce a diagonal-shape of decoding performances above chance (Carlson, Tovar, Alink, & Kriegeskorte, 2013; Fyshe, 2020). This uncovers the temporal unfolding of neural dynamics and relation to subsequential neural stages, and shows if neural representation is maintained or changes over time. Decoding accross conditions, where a classifier is trained on data of condition **A** and tested on its classification accuracy on condition **B** could show if neural representation is shared across conditions in temporally distinct manners (Fyshe, 2020).

1.4 Goals of this Thesis

In this thesis, a translational rule-switching learning paradigm already shown to explain cognitive and behavioral flexibility in rats was adapted to humans while analyzing simultaneously recorded magnetoencephalographic data. This paradigm showed rats learn correct rules by sequentially testing for the correct strategy on a trial-by-trial basis, rather than learning the correct action for all possible stimuli combinations in a given environment. It proposes strategies as a possible correlate of attention-driven stimulus selection during the learning process in new environments. It also showed strategies account for sudden change points occurring during the learning process. In this thesis, the rule-learning paradigm was adapted to humans and used to verify if similar cognitive processes are identifiable. To that aim,

the behavioral data of 27 healthy participants completing the rule-switch learning paradigm was analyzed, as well as the behavioral and neural data of additional 28 healthy participants completing the paradigm during *MEG* recording. This thesis will show how:

- humans consecutively test for the correct strategy on a trial-by-trial basis to infer the correct rule in unknown, ever-changing environments;
- the use of strategies accounts for abrupt behavioral change points occuring in learning processes.
- An *MVPA*-like analysis of simultaneously recorded *MEG* data will attempt to show neural correlates of strategy-based testing for the correct rule in constantly evolving environments.
- Further, behavioral and externally validated quantitative cognitive parameters obtained with *CANTAB* will be correlated with *MEG* decoding analyses to infer if brain regions relevant to cognitive functions project to decoding accuracies of strategies in *CANTAB* test-coherent sensor spaces.

In combination with rodent results, this thesis aims at an understanding of cognitive, behavioral and neural processes underlying flexible decision-making and provides a foundation for the translational investigation of impaired cognitive flexibility in clinical populations.

Chapter 2

Method and Materials

2.1 Method Outline

2.1.1 Pilot Experiment

Participant Recruitment and Demographics A behavioral pilot with *27* healthy participants (*15* males, mean age: *26,1*, SD: *6,6* years) recruited from the local community was conducted prior to the *MEG*-experiment. The participants were screened for neurological and psychiatric disorders, as well as physical exclusion criteria. Disorders screened for were depression, suicidality, bipolar disorder, obsessive-compulsive disorder, anxiety disorders, traumatic experiences, post-traumatic stress disorder, drug and substance abuse, psychotic disorders, anorexia, bulimia, as well as sight or hearing restrictions. Only healthy subjects participated in the study. They received no financial compensation. The measurements occurred at the library of the University of Heidelberg (Germany). One participant chose to terminate the experiment prematurely, after having completed 3 of the four experiment phases.

Experiment Description

Trial Description The pilot task was a multidimensional learning paradigm comprising 600 trials. The stimuli were delivered using the software application Presentation *(Version 20.1 Build 12.04.17, NeuroBehavioral Systems, Berkeley, USA)*. Each trial started with the presentation of two stimuli. These were a white circle and a sound at a frequency of 400 *Hz*. Both appeared at the same time and could either appear on the left-hand or on the right-hand part of the screen or, respectively, on the left or right ear plug. The presentation (left versus right) of both stimuli

was pseudorandomized. The stimuli were presented for 100 *ms*, followed by a 2900 *ms* delay interval with a white central crosshair. After the stimuli presentation, the white crosshair continued to be presented for 2900 *ms*. The crosshair then turned green for 2000 *ms*, indicating to the participants to log in their response. Their response could either be a left-hand or a right-hand button press. Following the green crosshair, the feedback appeared on the screen for 1000 *ms* - either the word "richtig" for positive feedback, or a white central crosshair for negative feedback, or if participants had omitted responding on the trial. The white central crosshair was then presented for 4000 *ms* before the stimuli of the next trial were presented. Figure 2.1 shows one trials' exemplary course. Participants were instructed to press left or right during the green crosshair phase; and to use the feedback to learn a rule they could follow to yield the most positive feedback.

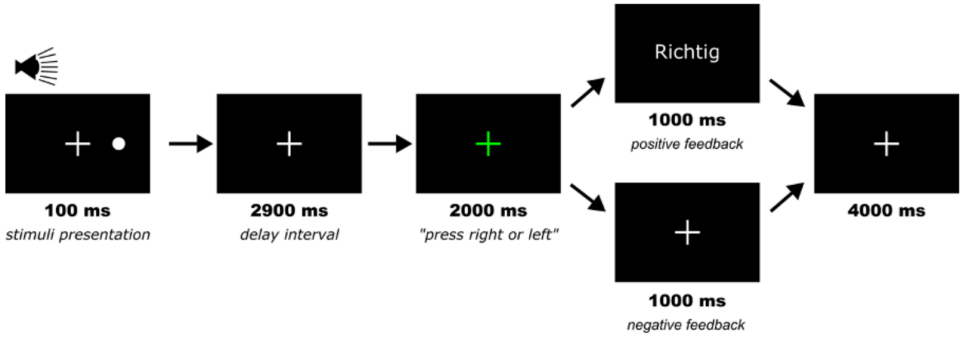


Figure 2.1: Depiction of a trial as employed in the learning paradigm, both in the pilot and in the final experiment version. Trial sequences as presented to participants are shown from left to right, starting with stimuli presentation.

Experiment Phases The experiment consisted of four different rule phases, with 150 trials each. The participants had been instructed to learn a new rule every time a new phase started. In the pilot, the rule switches were cued through a 5 second countdown. Rule switches were not cued in the main study. Feedback was set at a probabilistic rate of 80%, meaning that, if a participant had responded correctly on a trial, the corresponding correct feedback would be shown with an 80% probability,

for incorrect responses, a positive feedback was given with a 20% probability.

The four phases were, in the following order:

- *Dark* - press the opposing side of the visual stimulus *(150 trials)*
- *Right* - press the right-hand side *(150 trials)*
- *Alternate* - alternate left and right responses *(150 trials)*
- *Silent* - press the opposing side of the auditory stimulus *(150 trials)*

Debriefing Questionnaire After completing the learning paradigm, the participants were asked the following questions:

- *"What was your a priori approach?"*
- *"What did you find to be correct?"*
- *"Did you notice the rule change?"*
- *"How have you identified these rules?"*
- *"Have you found out the dark, right, alternate and silent rules?"*
- *"Did you use one or more of the following strategies (light, dark, sound, silent, right, left, alternate, win-stay-lose-shift)?"*
- *"Did you use other approaches: such as trying to infer the correct side to click per possible stimuli combination?"*
- *"What else did you try out during the experiment?"*

2.1.2 Main Experiment Combining Neural and Behavioral Data

Participant Recruitment and Demographics 32 healthy adults (22 females, mean age: *27,2*, SD: *9,1*, range: *19-55*) recruited from the local community solved a task during *MEG* recording. Participants were screened for exclusion criteria such as any current or history of mental and neurological illnesses as well as substance abuse. Participant written and informed consent was obtained following the guidelines of the Central Institute of Mental Health in Mannheim, Germany. The experiment was approved by the ethics committee of the medical faculty of Mannheim (Heidelberg

University, Germany). Participants were invited to 2 appointments, the first one consisting of the learning experiment in the *MEG*, the second one consisting of a 90-minute neuropsychological test battery (*CANTAB*). Participants received financial compensation of 50 Euros for their participation as well as reimbursement of travel expenses.

Three participants were removed from the analysis because of too strong artifacts during the measurements due to dental work. These participants are included in the demographic count. An additional 15 healthy adults (9 females, mean age: *25,5*, SD: *8,8*, range: *18-56* years) completed the final version of the learning paradigm outside of the *MEG* and without the additional *CANTAB* recording mentioned above. These 15 adults, as well as the 3 participants excluded from the *MEG* data analysis, were included in the behavioral data analysis of the final experiment version.

Experiment Description

Trial Description See section 2.1.1 and figure 2.1 on page 24 for trial structure description. Trials kept the same structure as in the pilot. The rule phases were altered in the final experiment version resulting in a total number of 650 trials and not 600 as in the pilot version. Importantly, rule switches were not cued.

Experiment Phases Participants were instructed to learn a rule throughout obtained feedback, using it to adapt and optimize their response pattern to receive the most positive feedback. Unbeknownst to them, the rule switched during the experiment.

The phases of the experiment were, in the following order:

- *Random Reward* - 50% chance of reward no matter which button was pressed *(126 trials)*
- *Dark* - always press the opposing side of the visual stimulus *(100 trials)*
- *Right* - always press the right-hand side (75 trials)
- *Alternate* - press the left and the right-hand side alternatively) *(199 trials)*
- *Silent* - always press the opposing side of the auditory stimulus *(150 trials)*

The 4 rule phases had an 80% probabilistic feedback rate, meaning correct responses received positive feedback with an 80% chance, incorrect responses with a 20% chance. The number of trials per rule phase was adapted accounting for results from the behavioral pilot. Easier rules were shortened while more difficult rules were lengthened, based on behavioral analysis of pilot data as well as participants' feedback. Because some participants inferred the first rule *dark* astonishingly fast in the pilot (sometimes in a few trials), a random reward phase with a random reward rate of 50% was added to yield behavioral data over a time series in which there was, unknown to the participants, no rule to learn. This ensured that even the subjects who would learn a rule in just a few trials' time produced a time window in which their a priori approach could be analyzed without subjects already being set to the correct rule. It also enabled examination of human behavior in experiment phases without any underlying rule. The rodent task version showed rats test strategies in rule-free environments, and research hints humans proceed in a similar structured manner even without the benefit of positive feedback (Collins, Cavanagh, & Frank, 2014; Farashahi, Rowe, Aslami, Lee, & Soltani, 2017).

The rule switch cue was eliminated, to present one continuous set of 650 trials without cued rule switches. Participants completing the non-stop learning paradigm outside of the *MEG* performed it without breaks. For participants completing the task during *MEG* acquisition, the task was split into 3 runs of 20 +/- 3 minutes. Participants were informed about the pauses before the start of the experiment.

Debriefing Questionnaire After completing the learning paradigm, the participants were asked the same questions during the debriefing as in section 2.1.1

2.1.3 *MEG* Data Aquisition

Recordings occurred in the Central Institute of Mental Health in Mannheim, Germany. *MEG* data was recorded in a magnetically shielded room *(Vakuumschmelze GmbH, Hanau, Germany)* with a 306-channel-system comprising 102 magnetometers and 204 planar gradiometers *(Elekta Neuromag Triux, MEGIN, Helsinki, Finnland)* at a sampling rate of 1000 *Hz*. A low-pass filter and high-pass-filter of 330 and 0,1 *Hz*, repectively, were applied. Internal shielding was applied. A digitization of three fiducial landmarks (nasion, left and right pre-auricular points), of 5 head position indicator coils and over 100 additional points over the scalp's surface was

performed with a Polhemus fastrak digitizer *(Polhemus, Vermont, USA)*. These additional points allowed an approximation of the participants' head shape. Inside *MEG* sensor level, head position indicator coils were magnetized each at different frequencies, thus allowing a head position estimation of the subject during the experiment. Head position indicator coils were used to measure the participant's head and shape inside the sensor helmet at the beginning of the three runs of the experiment (each 20 +/-3 minutes). Maxfilter software *(Version 2.2, Elekta Neuromag, Finland)* was used to align the participants' head shape and position according to the measured head position indicator coils and to filter out artifacts generated by head movement and ambient noise using signal space separation. Participants' vertical and horizontal electrooculograms *(EOG)* and electrocardiogramm *(ECG)* were monitored during *MEG* recordings by three bipolar electrodes, enabling an easier artifact detection and removal through independent component analysis. *EOG* electrode impedance was kept under 10 kΩ. Participants were continuously monitored by a camera and microphone following safety guidelines.

2.1.4 *CANTAB* Data Aquisition

During the second appointment, a computer-based cognitive assessment was administered using a test battery consisting of neuropsychological tests of the Cambridge neuropsychological test automated battery *(CANTAB® [Cognitive assessment software], Version 3.0.0, Cambridge Cognition (2022). All rights reserved. www.cantab.com, Cambridge, England.)*. The customized test battery focussed on measures of executive functioning and consisted of the following tests administered in the following order:

- Reaction Time *(RTI)*: This test assesses motor and response speed, as well as cognitive processing time, response accuracy and impulsivity. Performance depends on parietal and prefrontal brain networks.

- Intra- Extra- Dimensional Set Shift *(IED)*: This is a rule-switch task in which the switches are intradimensional at first, and extradimensional later. "Intradimensional" refers to the relevant stimulus dimension not changing with rule switches, "extradimensional" refers to the relevant stimulus dimension switching with rule switches. This test is sensitive to changes in the frontostriatal areas in the brain. It assesses attentional set formation maintenance, shifting, and attentional flexibility. See figure 1.3.

- Paired Associates Learning *(PAL)*: This test assesses visual memory and new learning. It is sensitive to changes in medial temporal lobe functioning.
- Spatial Span *(SSP)*: This test assesses the working memory capacity, giving a measure of the frontal lobe functioning
- One Touch Stocking of Cambridge *(OTS)*: This test assesses spatial planning and working memory capacities, giving a measure of frontal lobe functioning.
- Stop Signal Task *(SST)*: This test generates an estimate of stop signal reaction time, measuring the participants' ability to inhibit a prepotent response.
- Rapid Visual Processing *(RVP)*: This test is a sensitive measure of general performance and selective attention. It assesses functions such as sustained attention, cognitive processing, and is sensitive to parietal and prefrontal lobe dysfuntion.
- Spatial Working Memory *(SWM)*: This test measures participants' ability to retain spatial information and to manipulate remembered items in working memory. It is a sensitive measure of frontal lobe and of executive dysfunction.

The description of tests, outcome measures and informations about functionally associated brain regions are derived from the *CANTAB* website, see: `https://www.cambridgecognition.com/cantab/cognitive-tests/` (Cambridge Cognition, 2022), and from the CANTAB-Eclipse Manual version 3.0.0 (Cantab Cognition Ltd(2006), "CANTABeclipse™ Test Administration Guide", Version 3.0.0, Cambridge Cognition Ltd. (Limited, 2006) and "CANTAB Cognition handbook", 1st edition, 2nd version, 2015, Cambridge Cognition Ltd (Limited, 2015).

2.2 Analysis Outline

The analysis section is split into a behavioral analysis, a *MEG* data analysis and cluster-based correlation analyses.

2.2.1 Behavioral Data Analysis

Qualitative Questionnaire Data The following behavioral data analysis was performed for both the pilot data and the final experiment version data.

Detect Strategies The participants were asked what rules they had identified and how they solved the problem. Using the qualitative questionnaire data, it was aimed to classify the participants into one of four categories regarding their *a priori* approach to the rule-learning paradigm:

- Infer, for each possible environmental state the correct action to take;
- Use attention-modulated and stimulus-specific strategies;
- A combination of both approaches stated above;
- Alternative approaches.

Rule and Strategy Indices The experiment was administered with the Presentation software application *(Version 20.1 Build 12.04.17, NeuroBehavioral Systems, Berkeley, USA)*. During the learning paradigm data acquisition, the participants' behavior was recorded into a logfile. The logfile contained the following measures: the participants' response (either a right-hand or a left-hand click), the reaction time (time between the stimuli presentation and the participants' response), the feedback received per trial, and whether the response was correct or incorrect in accordance with the underlying rule.

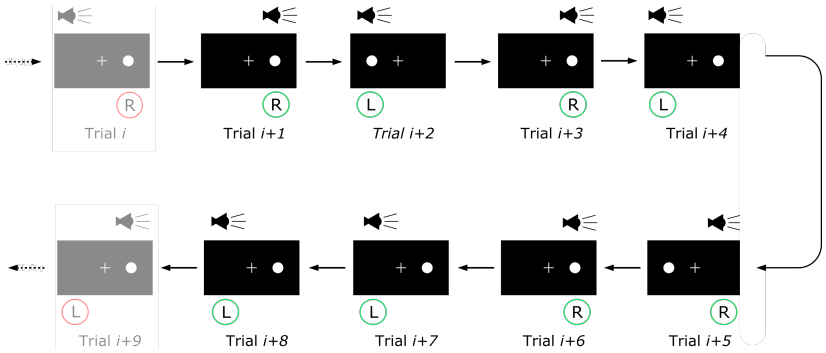
These measures allowed for analysis of the participants' response pattern. The response pattern of the participants was analyzed by a custom-made script *(Bähner F.*, written in *Matlab 2017a, Natick, USA)*. The script tested, for each subject and for each single trial, whether the according response was consistent with one of eight behavioral strategies. The aim was to find out whenever the participants had carried out any given strategy more frequently than any other strategy or than chance would predict. The strategies the algorithm was able to detect were the following:

Strategy Name	Strategy Description
Alternate	Click left and right alternately
Click	Click the side the auditory stimulus was on
Dark	Click the opposite side the visual stimulus was on
Left	Always click left
Light	Click the side the visual stimulus was on
Right	Always click right
Silent	Click the opposite side the auditory stimulus was on
Win-stay-lose-shift	Click one side (right or left) and change the side as soon as a negative feedback is given

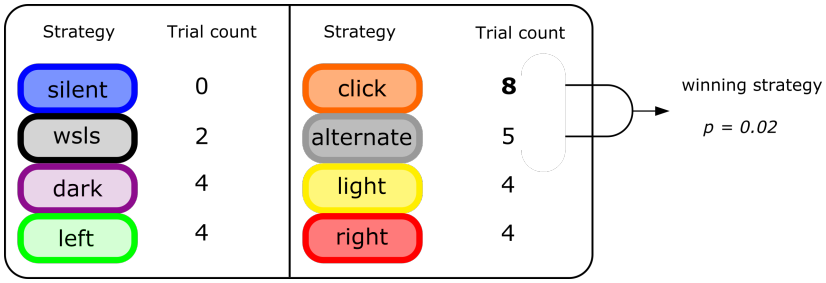
Table 2.1: Strategies detectable in the behavioral analysis and according description

For each strategy, trial sequences in which following that strategy was more probable than chance or than the other seven strategies were identified. This was done by defining time series for each of the eight detectable strategies and computing, per trial and time series, whether the participants' answer was coherent to the corresponding strategy. Figure **??** exemplarily shows such a process for a sequence of trials. Based on binomial distribution, the probability of the employment of the identified strategy per series of subsequent trials was tested against chance and against the probability that another strategy was more likely employed.

The minimal sequence length to permit significance level being *6*, it was repeatedly increased until reaching the analyzed rule learning sequence length. If sequences of different strategies identified as significant overlapped, the sequence with the lowest binomial test statistic was discarded. The explanation of the custom-made script detecting strategies as well as figures 2.2a and 2.2b are adapted from (Bähner et al., 2022).



(a) Exemplary depiction of a sequence with *click* as the underlying rule, continuously comprising trial *i+1* to trial *i+8*. Behavioral responses, *R* for right and *L* for left, are circled in green if correct, red if wrong. Stimulus presentation consists of a sound and a dot (appearing independently; either on the left or the right).



(b) Trials consistent to the respective detectable strategies are counted for the analyzed sequence as seen in figure 2.2a. Binomial test statistics show the strategy *click* was employed.

Figure 2.2: Visualisation of strategy detection algorithm for a series of subsequent trials

Rule and Strategy indices For each rule sequence and for each participant individually, the following rule and strategy indices were identified:

- the first correct trial *(first correct sequence)*;
- the length of that index sequence *(index sequence)*;
- how many additional correct sequences had been detected before the learning criterion was reached *(index sequence 2 crit)*;

- the longest sequence with the corresponding correct rule *(longest sequence length)*;
- the percentage of trials explained by a detected strategy *(% explained by strats)*;
- how many strategies were tried out in each rule phase *(i.e. strategies tried out in dark)*;
- and for each rule whether it was solved or not.

Change Point Detection The detection of abrupt behavioral changes and the computation of their statistical significance was achieved using paired adaptive regressors for cumulative sum (*PARCs*, (Toutounji & Durstewitz, 2018)). Change points represent abrupt changes in a sequence of subsequent trials (one time course represented by an individual rule phase), i.e. over the performance metric or the reaction time metric. The algorithm is able to identify more than one change point over the same time series. The details of the procedure can be found in (Toutounji & Durstewitz, 2018); the Matlab code is available at: `https://github.com/htoutounji/PARCs`. The performance metric is defined for each individual trial with a binary value coding for a correct (1) or incorrect (0) response in relation to the underlying rule. The reaction time is defined as the time needed to answer with a button-press following the stimuli presentation. *PARCS*, like other change point detection methods, first calculates the cumulative sum *(CUSUM)* of differences from the mean performance metric and the performance metric on individual trials; or of differences from the mean reaction time and the reaction time on individual trials. Typically, the change point is identified in the trial where the maximum, weighted, absolute *CUSUM* of differences to the mean is largest (i.e., the cumulative sum of differences would begin to decrease or increase subsequently to that point). Since this method tends to favor the detection of change points near the center of analyzed time series, *PARCs* applies an adaptive regression spline method by fitting pairs of nonoverlapping piecewise linear regressors to the *CUSUM* difference vector (Toutounji & Durstewitz, 2018). In the present study, trials where the overlapping splines provided the best fit to the *CUSUM* difference vector were taken as the significant change point trials for according time series. Non-parametric permutation bootstrap testing was used with the significance level set to 0.05, based on 1000 bootstrap samples.

Correlation of Rule Indices and Behavioral Variables Correlations between rule indices and behavioral variables obtained in *CANTAB* were tested for their significance using *pearson's* correlation and the function `corrcoef`. The null hypothesis that the means of the correlated variables do not differ significantly was tested with a two-sample t-test, using the Matlab function `ttest2`. *(Matlab R2017b).*

2.2.2 *MEG* Data Analysis

MEG data analysis was performed with custom-made scripts *(Popov T.)* using the fieldtrip toolbox (version: fieldtrip-20200128) in Matlab *(R2017a)* developed by (Oostenveld, Fries, Maris, & Schoffelen, 2011) as well as the mvpa-light toolbox, (version: MVPA-light VERSION) developed by (M. S. Treder, 2020).

Preprocessing *MEG* raw data was preprocessed according to spans of 2 seconds before and 4 seconds after stimuli presentation per trial. Each trial was detrended and baseline correction was applied. The data was resampled at a sampling rate of 300 *Hz* for independant component analysis and at 100 *Hz* for subsequent analyses. Eye-movement activity such as blinks and saccades monitored by the vertical and horizontal *EOG* electrodes, as well as cardiac activity monitored by the *ECG* electrodes were detected as artifacts and removed using independent component analysis.

Classification of Event-Related *MEG* Data using *MVPA-light* The method introduced in section 1.3 was applied in the following steps of *MEG* data analysis. The toolbox used was *MVPA-Light* (M. S. Treder, 2020), implemented in the fieldtrip toolbox *(fieldtrip-20200128)* (Oostenveld et al., 2011). Classifications performed included time-level and sensor-level analyses, as well as testing classifiers' ability to generalize from data of one time point to data of another to explore temporal organization of neural dynamics.

Temporal Generalization Matrix Classifiers were trained on *MEG* data labelled into conditions for each timepoint **T**, and subsequently tested on their ability to classify data into conditions from the same timepoint **T** and other time points **T'**. The resulting *time* x *time* generalization matrix is a map of decoding accuracies per *train time-test time* pair. Linear discriminant analysis *(LDA)* was used as a means of classifying. If there were more than two conditions along which the *MEG*

data was classified, multiclass linear discriminant analysis was used. `Accuracy` was chosen as an output metric. *K-fold* cross validation was used, with k set to 5, and was repeated 5 times. `Ft_timelockanalysis` was called to perform the computation. The resulting time generalization matrix with classification accuracy values per train-test time combination was plotted using `mv_plot_result`.

Classification Accuracy over Sensor Space Consequently, a searchlight-type analysis was performed to assess which channels contributed most to classification performance. The averaged time window of 0 to 1 s after stimulus presentation was used as unidimensional feature dimension. After calling `ft_prepare_neighbours` and setting the method to `triangulate` over the raw-data file of one participant, the resulting inferred topography of the channels was used to cluster neighboring channels together as search dimensions. The classification accuracy was computed for channels using a (multiclass-) LDA classifier and a 5-fold cross-validation with 5 repeats. `Accuracy` was chosen as an output metric. `Ft_timelockanalysis` was called to perform the computation. The resulting classification accuracy per channel was plotted using `ft_topoplotER` and using the layout `neuromag306mag_helmet.mat` as topography.

Classification Accuracy over Time Classification over time was performed to identify which time points in the trials contributed the most to classification performance. Channels were set as feature dimension. The classification accuracy was computed for each time point separately using a (multiclass-) LDA classifier and a 10-fold cross-validation repeated twice. `Accuracy` was chosen as an output metric; `ft_timelockanalysis` was called to perform the computation. The resulting accuracy value per time point in the trial was plotted using `mv_plot_result`. On single-participant level, the accuracy values per time points correspond to the diagonal of the *TGM*.

2.2.3 *Biochannel* Data Analysis

Preprocessing Data obtained from biochannels 2 and 3, respectively the vertical and horizontal *EOG* electrodes, was preprocessed with a windowed sinc lowpass filter at a frequency of 40 *Hz*. Each trial was detrended and baseline correction was applied. Data was resampled at a sampling frequency of 100 *Hz* and preprocessed into trials comprising 0.5 s before and 2.75 s after stimulus onset.

Classification Analyses for Biochannel Data Classification analyses on *MEG* data were repeated for the *BIO* data along the same conditions using the following methods. See section 2.2.4 for classification analyses and according classes computed. Using (multiclass-) LDA, biochannel data was classified in the following three ways:

Classification over time points Using channels as features, a classification using a (multiclass-) LDA classifier was performed by calling `ft_timelockstatistics`, with 10-fold cross-validation repeated twice, `accuracy` being the metric calculated. This resulted in accuracy values per time point.

Classification over channels Setting channels as features and generalizing over the time window of 0 s to 1 s after stimulus onset, a classification was performed by calling `t_timelockanalysis` using a (multiclass-) LDA, with 5-fold cross-validation and 5 repeats. The metric calculated was accuracy per channel.

Classification over time and channels Setting neither channels nor time points as feature dimensions, both channels and time points were used as search dimensions. This yielded information as to where in the channel space and when in the time points discriminative information could be found. The classification was performed by calling `ft_timelockanalysis` using a (multiclass-) LDA classifier, with 5-fold cross-validation repeated twice, and accuracy being the metric. The dimension order of the resulting classification accuracy was organized in a *time_channels* classification accuracy matrix, showing in which time points which of the channels carry discriminative information for classificating accurately.

2.2.4 Population *MEG* Data Analyses

Table 2.2 shows which classification analyses were performed on *MEG* data and corresponding biochannel data obtained in this experiment. For each constellation of conditions decoded, a cross-participant average was calculated using `ft_timelockgrandaverage`, excluding the first classification analysis, which was only performed on a single-participant level. The parameter was set to `accuracy`: the average was calculated for classification accuracies over time and sensor space. This returned a classification accuracy average per time point, as well as a classification accuracy average per channel. Similarly, for biochannel data the metric calculated was clas-

sification accuracy average per channel and classification accuracy per time point over averaged channels.

`Neuromag306mag_helmet.mat` was used as a layout for plotting the grand average of the searchlight analyses using `ft_topoplotER`. `Ft_singleplotER` was called to plot the average of the decoding accuracies over time.

Classifications and according conditions	***Criteria and number of participants fitting these criteria***
Alternate & *Click* & *Dark* & *Left* & *Light* & *Right* & *Silent* & ***Win-stay-lose-shift***	for *n=29* participants individually according to the detected strategies in the behavioral analysis (i.e. if 5 strategies were detected, only these 5 were used as classification conditions)
Click & *Dark* & *Silent*	for *n=15* participants with minimum 30 trials detected per according strategies in the behavioral analysis
Light-specific (*Light* & *Dark*) & Sound-specific(*Click* & *Silent*) & Place-specific (*Alternate* & *Right* & & *Left*) strategies	for *n=28* participants, excluding one participant which had not performed any strategies corresponding to the place-specific class

Table 2.2: Description of computed classifications and according classes. Other combinations of three strategies were computed (with n between 8 and 13) with similar quantitative results. The main analysis of this thesis was the third listed here.

2.2.5 Correlation Analyses of Topographical Accuracy Maps with Behavioral Variables

To test for significant correlations of classification accuracies and *CANTAB* behavioral variables per subject, cluster-based permutation statistics were applied.

Correlation analyses of topographical classification accuracies with behavioral variables Using pearson correlation coefficient, cluster-based permuta-

tion analyses were applied to identify significant correlations of topographical classification accuracies with quantitative behavioral variables. Channel-accuracy-pairs arising from classification analyses on single-subject level were used as input samples. The effect at the sample level was quantified using `ft_statfun_correlationT`, the threshold was set at an α-level of 0.05. Selected samples were clustered by spatial adjacency, with the minimum number of neighboring channels required to form a cluster set to 2. Channel structure was computed by reading in a participant's raw-data file and calling `ft_prepare_neighbours` with method set to `distance`. Distance threshold was set to 4. Reference distribution under the assumption of interchangeable samples was approximated using `monte-carlo` as a permutation-based approach, with the number of draws set to 500. This produced an estimate of the p-value under permutation distribution. α-level was set to 0.025 in a two-tailed test design. The cluster-based statistic was set to `maxsum`, meaning the maximum of the cluster-level statistics was evaluated in a two-sided-test with α set to 0.025. The maximum of the cluster-level statistic is equal to the sum of the T-statistics belonging to the samples of the same cluster.

The topography of the correlation coefficient *rho* per channel, obtained by calling `ft_topo-`
`plotER` was the output of the computation. Using `ft_topoplotER`, the output was plotted using `Neuromag306mag_helmet.mat` as a layout. Channel clusters with significant correlations between classification accuracies and *CANTAB* behavioral variables as determined by cluster-based permutation tests were highlighted in red using `cfg.highlightchannel`.

Chapter 3

Results

3.1 Behavioral Analysis

3.1.1 Qualitative Results

It was hypothesized the participants would distinctly recall using either a strategy-based approach or a state-action-pair-based approach, and that the strategy-based approach would be predominant and more apparent. A state-action-pair-based approach entails the correct course of action would have to be learned for each possible combination of environmental stimuli. Comparatively, a strategy-based approach implies sequentially testing hypothetical assumptions about the correct rule.

Using the qualitative questionnaire data as explained in section 2.1.1, it was aimed to classify the participants into on of four categories regarding their *a priori* approach to the rule-learning paradigm:

1. Infer, for each possible environmental state (2 current stimuli with 2 possible conditions per stimulus with a previous trial with 2 possible conditions for 2 stimuli = 16 states), the correct action to take; or
2. Use attention-modulated and stimulus-specific strategies.
3. A combination of both approaches stated above
4. Alternative approaches

There was no clear behavioral pattern emerging from the participants' answers, neither on a single subject basis, nor across subjects. As such, it was not possible to

classify the participants in either category. This was in part since most participants were not able to recall their approach. Noteworthy is that only when distinctly asked about the use of specific strategies, all subjects reported having used at least three distinct strategies, and these specific strategies were always detected in the behavioral analysis.

In sum, spontaneous reports during debriefing indicate that subjects were either not aware about their learning strategies or couldn't recall them.

3.1.2 Strategy Detection

It was hypothesized that if the learning process is shaped by a strategy-based approach, these strategies would be detectable in the behavioral analysis.

To that aim, the participants' behavioral response pattern was analyzed for detection of strategies.

Strategy Name	**Strategy Description**
Alternate	Click left and right alternately
Click	Click the side the auditory stimulus was on
Dark	Click the opposite side the visual stimulus was on
Left	Always click left
Light	Click the side the visual stimulus was on
Right	Always click right
Silent	Click the opposite side the auditory stimulus was on
Win-stay-lose-shift	Click one side (right or left) and change the side as soon as a negative feedback is given

Table 3.1: Strategies detectable in the behavioral analysis and according description

Exemplarily shown in figure 3.1 is the behavioral analysis of one participant over the complete time course (650 trials) with dotted lines indicating the rule switches. The method presented in section 2.2.1 detected if any and which strategy was the most probable approach for corresponding trial sequences. Identified strategy sequences were plotted per trial. The identified strategies are explained in table 3.1.

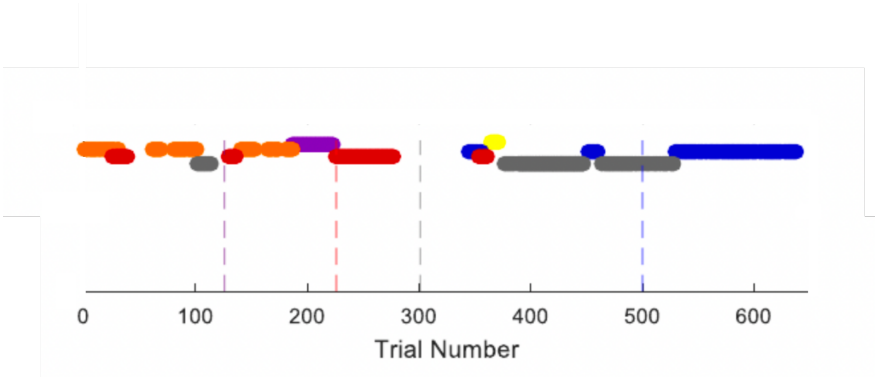


Figure 3.1: Most probable and significantly detected strategies are plotted per trial *(x-axis)*, exemplarily here for one participant. The identified strategies are color-coded: ***Alternate***, ***Click***, ***Dark***, ***Left***, ***Light***, ***Right***, ***Silent***, and ***Win-stay-lose-shift***. Dotted lines mark the beginning of the rule sequences: The first, purple line at trial 126 marks the end of the *random reward* phase and the beginning of the *dark* rule. The red line at trial 226 marks the beginning of the *right* rule, the grey line at trial 301 marks the beginning of the *alternate* rule and the blue line marks the beginning of the *silent* rule at trial 500.

The participant was able to correctly explain all rules afterwards, coinciding with findings in the behavioral analysis, showing she had effectively learned all 4 rules. The analysis shows the participant employed 7 of the possible 8 strategies the method presented in section 2.2.1 can identify. This exemplarily demonstrates that participants tested for the correct rule by sequentially testing out different hypotheses about the correct rule in the form of strategies on a trial-by-trial basis.

3.1.3 Change Points

If participants are found to sequentially test strategies on a trial-by-trial basis for the correct rule, learning could be expected to progress abruptly. A state-action-pair based approach would imply a less abrupt and more gradual learning process. It is expected that finding the tested strategy to be the correct rule would reflect in abrupt changes in behavioral parameters such as reaction time or performance metrics.

To that means, change points in the performance metric and in the reaction time measure were computed using *CUSUM*. Figure 3.2 exemplarily showcases one participant's performance during the *dark* rule sequence. Significant change points are identified over a time series by calculating the cumulative sum of differences

betwen the mean of a metric, here performance and reaction time metric, and the metric measure in individual trials. The change point is taken as the trial where the maximum, weighted, absolute cumulative sum of differences to the mean is largest, i.e., the cumulative sum of differences would begin to decrease subsequently to that point. (see section 2.2.1 for further explanation). Reaction time (peak-shaped) change points thus marked the shift in which the participants responded significantly faster relative to prior trials; performance (trough-shaped) change points marked the shift in which the participants performed significantly better relative to prior trials. Correlations between the first trial of the correct strategy as identified in the behavioral analysis and the change points in reaction time and performance metric were tested on their significance.

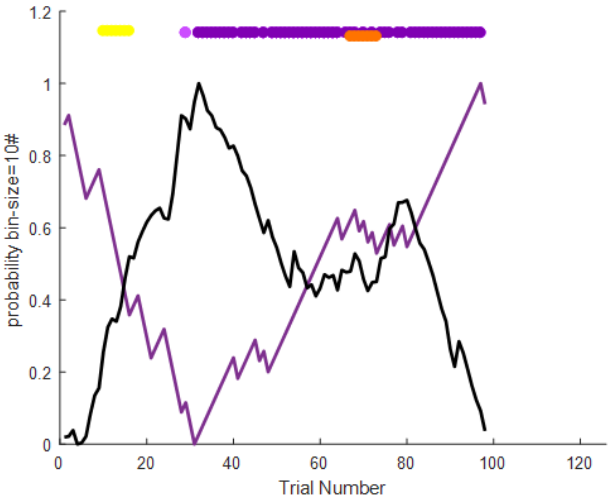
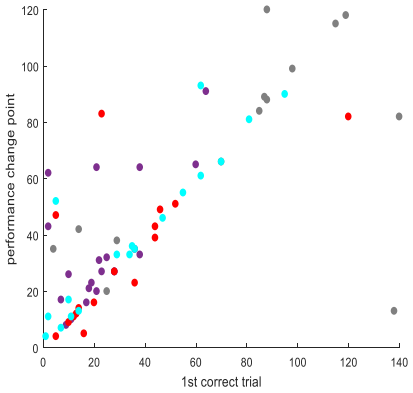


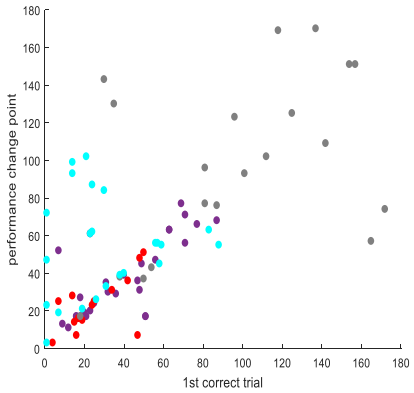
Figure 3.2: Identified strategies plotted per trial, as well as reaction time (*black*) and performance *CUSUM* plots (*purple*), here for the rule sequence *dark*. Strategies are color-coded: ***Alternate***, ***Click***, ***Dark***, ***Left***, ***Light***, ***Right***, ***Silent***, and ***Win-stay-lose-shift***. The trough-shaped performance change point marks sudden shifts of wrong responses to correct responses in relation to preceding responses' average. The peak-shaped reaction time change point marks the sudden shift in which the reaction time got significantly shorter in relation to preceding responses' average. *X-Axis*: Trial number; *Y-Axis*: strategy employment probability as identified in behavioral analysis.

For each rule sequence with a corresponding correct strategy sequence, change points in performance metric and reaction time were computed. Their correlation coefficient with each other as well as with the first trial of the first correct strategy sequence as well as the corresponding significance was calculated. After testing

correlations between change points and the strategy-based approach on their significance, the next question addressed was if the means of the measures correlated differed significantly and if one metric came significantly earlier in the observed time series than the other metric.



(a) Pilot data, *n=69*: 1st correct trials correlate significantly with performance change points *(pearson's corr. coeff.=0,7399; p=0.000)*. The means of both variables do not differ significantly *(p= 0.4973)*.



(b) Final experiment version, *n=88*: 1st correct trials correlate significantly with performance change points *(pearson's corr. coeff.=0,6563; p=0.000)*. The means of both variables do not differ significantly *(p= 0.3548)*

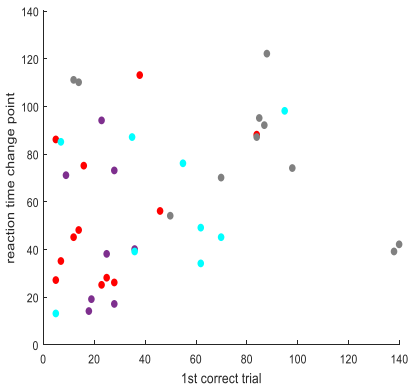
Figure 3.3: Trials corresponding to significant performance change points *(y-axis)* plotted against 1st correct trials of respective time series *(x-axis)*. Color indicates which rule sequence data points originate from (***Dark***, ***Right***, *Alternate*, and ***Silent***.)

In the final experiment version, correlation analyses showed significant correlations between reaction time change points and respective 1st correct trials, between performance change points and respective 1st correct trials, and between reaction time and performance change points. This is shown in figures 3.3, 3.4 and 3.5. On average, reaction time change points occurred 15.1 trials later than the respective 1st correct trial. The means of the other variables did not differ. This adequately replicates according results of the behavioral analysis of pilot data (see figure 3.2).

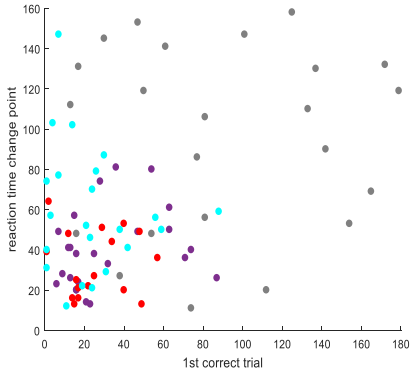
Detected change points underline learning processes are not gradual. Further, the identified correlations show the employment of strategies adequately explains behavioral change points. A decrease in reaction time indicating more confidence in the correctness of the tested strategy develops on average 15 trials *(SD: 45,7 trials*

in final experiment version, SD: 42,6 in pilot after the beginning of the correct strategy sequence. In sum, abrupt behavioral progressions in learning processes, significant correlations between behavioral change points and employment of the correct strategy, as well as the reaction time change point occurring significantly later all underline humans test strategies to infer the correct rule.

Consequently, strategies as reflected by selective attention may constitute an important mechanism enabling learning and switching rules in humans.

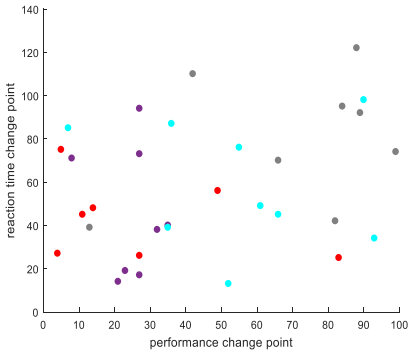


(a) Pilot data, *n=40*: 1st correct trials do not significantly correlate with reaction time change points *(pearson's corr. coeff=0,2070, p=0,2000)*. The means of both variables differ significantly *(p= 0.0326)*. On average, reaction time change points occur 14.4651 trials after the 1st correct trial.

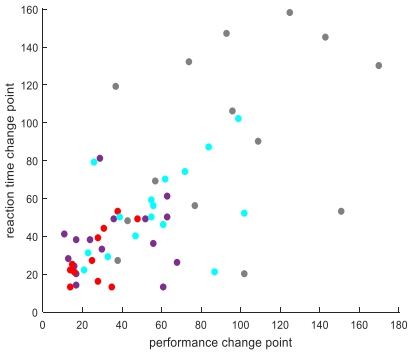


(b) Final experiment version, *n=86*: 1st correct trials correlate significantly with reaction time change points *(pearson's corr. coeff.=0,3853; p=0,0002)*. The means of both variables differ significantly *(p= 0.0224)*. On average, reaction time change points occur 15.1406 trials after 1st correct trials.

Figure 3.4: Trials corresponding to significant reaction time change points *(y-axis)* plotted against respective 1st correct trials *(x-axis)*. Color indicates which rule sequences data points originate from (***Dark***, ***Right***, *Alternate*, and ***Silent***.)



(a) Pilot data, *n=32*: Reaction time and performance change points do not correlate significantly *(pearson's corr. coeff.=0.2818; p=0.1182).*The means of both variables do not significantly differ *p=0.1155*.



(b) Final experiment version, *n=57*: Reaction time and performance change points significantly correlate *(pearson's corr. coeff.=0.6481; p=0.000).* The means of both variables do not significantly differ *p=0.9779*.

Figure 3.5: Trials corresponding to significant performance change points *(y-axis)* are plotted against reaction time change points of respective time series *(x-axis)*. Color indicates which rule sequences data points originate from (***Dark***, ***Right***, ***Alternate***, and ***Silent***.).

3.1.4 Correlation of behavioral variables with *CANTAB* performance parameters

Significant correlations between behavioral variables measured in the rule-switching paradigm and performance metrics obtained using *CANTAB* were identified using *pearson's* correlation coefficient.

Table 3.2 shows the *Rho* correlation coefficient of participants' *CANTAB* performance metrics with according "*% of trials solved correctly*", as well as corresponding p-values. 29 participants were evaluated, as 29 participants had performed both *CANTAB* tests and the multidimensional learning paradigm. "*% of trials solved correctly*" was inferred over all the trials belonging to rule-learning phases (=all phases excluding the *random-reward* phase) and showed the percentage of correctly answered trials. The metrics obtained during *CANTAB IED (EDS-errors, total errors (adjusted)*, and *total trials (adjusted))* showed a significant negative correlation with the corresponding participants' "% of trials solved correctly", with a respective correlation coefficient of -0.4204, -0.4485, and -0.4275, and respective p-values of 0.0233, 0.0147, and 0.0207.

	IED EDS	*IED*-total errors	*IED*-total trials	*OTS* mean choice	*OTS* problems solved on 1st choice	*SWM* within errors	*SWM* between errors	*SWM* double errors	*SWM* total errors	*SWM* strategies
Rho	-0.4202	-0.4485	-0.4275	-0.2154	0.2696	0.2276	-0.1385	-0.1836	-0.1499	0.0633
p-value	0.0233	0.0147	0.0207	0.2618	0.1573	0.2351	0.4736	0.3404	0.4377	0.7443

Table 3.2: Correlation coefficient and according p-value of "*% trials solved correctly*" with *CANTAB* performance metrics. Significant correlations are highlighted in yellow.

EDS-errors, the number of errors made when an extra-dimensional shift was required, *total errors (adjusted)*, the number total errors committed, a measure of the participants' efficiency at attempting the test, and the *total trials (adjusted)*, the number of trials completed on all attempted stages, are all metrics indicating poor performance when higher. The participants' performance parameters in *IED*, a validated assessment of rule learning and switching (Lowe & Rabbitt, 1998; Sahakian & Owen, 1992; Wild & Musser, 2014) correlated significantly with the participants' performance parameters in the employed multidimensional learning paradigm.

3.2 *MEG* data Analysis

The following parts aim to examine if the behaviorally detected strategies are also decodable in MEG data, at both the individual participant and cross-participant average level, using a MVPA-style approach as introduced in section 1.3.

As not all subjects employed the same number or constellation of strategies, popular strategy combinations were computed and decoded on a cross-subject level in according subjects (i.e., all subjects executing *dark*, *right* and *silent*, with a minimum of 30 trials per strategy) in order to enable cross-subject analyses. Additionally, cross-subject analyses included decoding of strategies attending the same stimulus

grouped in the same class. For example, *silent* and *click*, were grouped in the same stimulus-specific decoding class.

3.2.1 Single-participant level

Exemplarily shown and explained in figure 3.6 is the temporal generalization matrix arising from the analysis of *MEG* data of one participant. The following graphs arise from the analysis of the same participant discussed in section 3.1.2. The showcased participant had 6 identified strategies in the behavioral analysis: *light*, *dark*, *sound*, *silent*, *alternate*, and *right*. These identified strategies make up the 6 classes used in subsequent single-level classification analyses. All trials with corresponding detected strategies were included in this classification analysis. Classifiers were tested on their ability to decode the identified strategies in the neural data at a single-participant level.

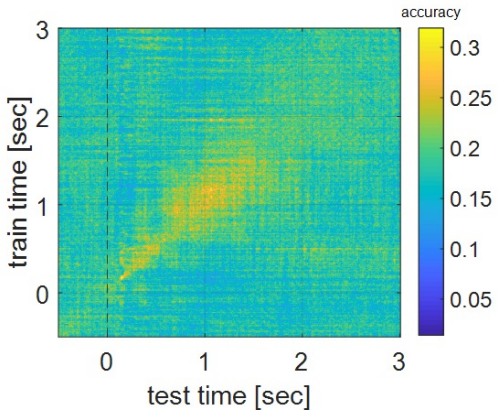
Figure 3.6a shows a temporal generalization matrix at the single-participant level for classes corresponding to single strategies. Above chance classifier performance can be inferred throughout the diagonal starting at 150 *ms* after trial onset. The diagonal is slender until 600 *ms* after stimulus onset. From 600 *ms* to 1500 *ms*, the form of the *tgm* diagonal broadens, reflecting the stages of 500 *ms* to 1500 *ms* after stimulus onset generalize over a more elongated time window than the stages decoded at 250 - 500 *ms* after stimulus onset. The form and broadness of the diagonal indicate neural activity underlying strategy completion is sustained for a longer time period in the later *ms* of the trial than at the beginning (Stokes et al., 2013). The *tgm* implies that neural encoding of strategies is firstly composed of sequentially activated, quickly unfolding neural computations only periodically present over a short window of time, then developing into a longer sustained pattern of activity representing a final decision related state.

The corresponding classification analysis over sensor space shown in figure 3.6b denotes classification accuracy to be above chance level of 16,67 % over a large portion of the sensor space. A predominance can be made out over the right parietal sensor space, as well as over the right temporal and, in smaller quantities, over the left temporoparietal and occipital sensor space. A majority of the sensor space contributes to an above chance decoding accuracy of strategies, this could reflect separately coded and decodable neural large-scale dynamics for the employed

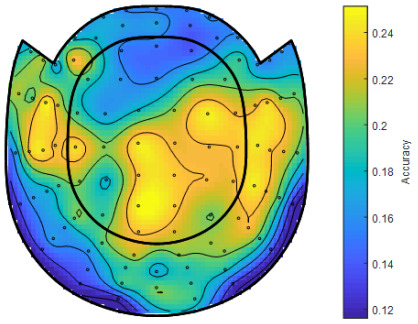
strategies.

In the temporal analysis, peak classifier performance is seen at 150 - 250 *ms* after stimulus onset as well as, less pronounced, but for a longer time period at 750 - 1500 *ms* after stimulus onset. In accordance with the diagonal of the same participant's *tgm* presented above, the peaks found in the decoding analysis per time point reflect the peaks found in *tgm* diagonal.

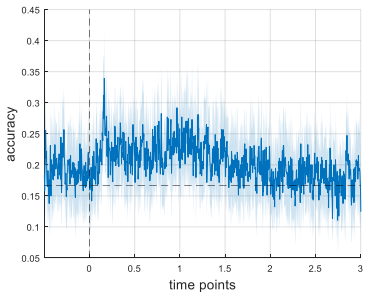
Strategies are decodable at a single-subject level, with a majority of time points between stimulus onset and 1500 *ms* after stimulus onset contributing the most to classification accuracy above chance. Strategies seem to have distinct neural representations shaped by a quick unfolding of specific dynamics gradually evolving into a more sustained activity pattern. These single-subject findings are replicated amongst subjects and reinforce the hypothesis that humans use strategies during the learning process.



(a) *Time* x *time'* generalization matrix: the classification accuracies' matrix of classifiers trained with data of a certain time point ***t*** *(y-axis)* and tested on data of another time point ***t'*** *(x-axis)*.



(b) Decoding accuracy per channel over the averaged time window of 0 - 1 *s* after stimulus onset



(c) Decoding accuracy per time point for classifiers tested and trained on the same time point over the averaged channel space, corresponding to the diagonal of the *tgm* in figure 3.6a. Chance level is indicated by the vertical dotted line.

Figure 3.6: Single participant classification analyses along 6 classes corresponding to the 6 strategies detected in the behavioral analysis, with an according chance level of 16.67% (100%/6).

3.2.2 Group-level Analyses

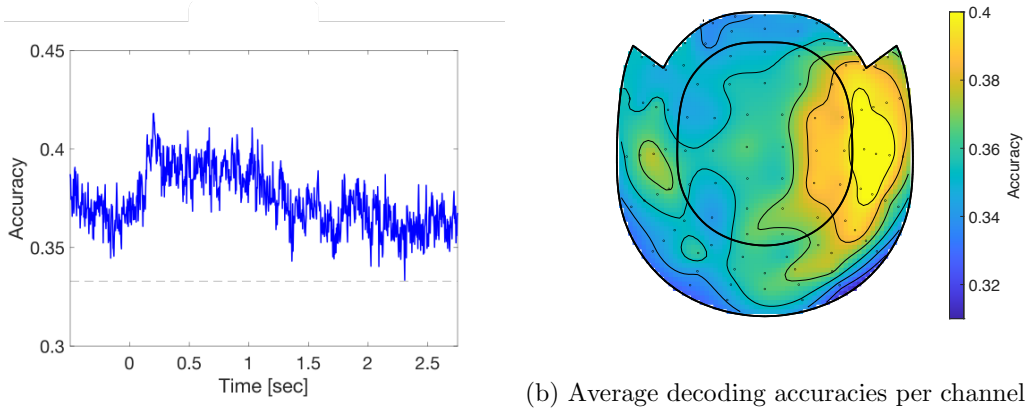
After investigating whether strategies were decodable on single-subject level, group-level analyses were conducted. The following analyses had two aims: firstly, to see if strategies were decodable on cross-subject analyses; and secondly, to see if strategies aimed at the same stimulus dimension shared neural representation while differentiating from strategies aimed at other stimulus dimensions.

Popular Strategy Combinations

Exemplarily shown and explained is one of the popular strategy combinations decoded on a cross-subject analysis. All participants with a minimum of 30 trials per selected strategy (here: *dark*, *click* and *silent*) were included. The classifiers were tested on their ability to differentiate these three strategies. Topographical and temporal decoding average maps were computed. Other popular strategy combinations with n ranging from 8 to 13 were decoded on a cross-subject level. As these yielded similar quantitative results, they are not shown here.

Average decoding accuracy for the three selected strategies is above chance at all analyzed time points, including preceding stimulus onset. Preceding stimulus onset, average decoding accuracy is above chance level at around 38 %, declining around stimulus onset. Peak decoding accuracy of 40 % is observed at approximately 250 - 1000 ms following stimulus onset, steadily declining after 1000 *ms*, levelling out at 38 %. Decoding accuracy is above chance level of 33% in the majority of the sensor space, with a broad peak concentrated over the right temporoparietal sensor space reaching an accuracy of 39,5%. Decoding accuracy is above chance level at all time points analyzed, including preceding stimulus onset. This could indicate that participants are set on which strategy to execute before stimulus onset and reflect top-down attention allocation to the stimulus to attend to in form of strategy planning and execution.

Single strategies are discernably decodable in the group-level analyses. Seeing the decoding accuracy average above chance at all time and a majority of sensor points, it is deductable distinguishable large-scale neural codes underlie single strategies.



(a) Average decoding accuracy over time, chance level of 33.33% marked by dotted line

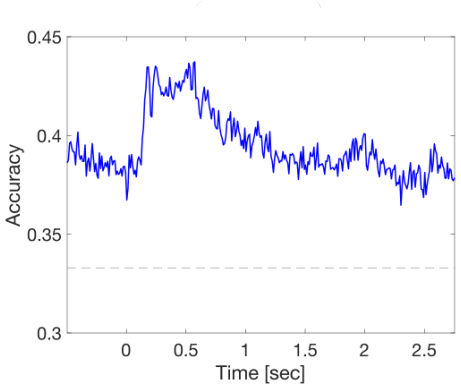
(b) Average decoding accuracies per channel for the averaged time window of 0-1 *s* after stimulus onset

Figure 3.7: Average (*n=15*) decoding accuracies over time *(a)* and sensor space *(b)* for *dark*, *click* and *silent*. Chance level is 33.33 %.

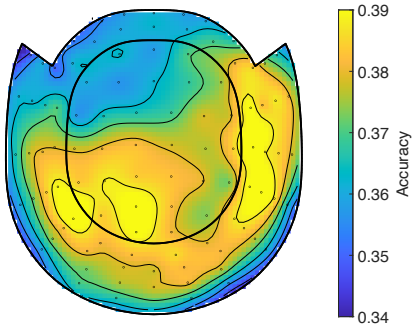
Attention-specific Strategy Classes

Attention-specific strategy classes were defined, aiming to see if common neural codes are shared over strategies abiding to the same attention class and differentiating them from strategies oriented at another stimulus dimension. It investigated whether participants' use of strategies was attention-driven; and if attention to stimuli impacted the selection and execution of strategies and the approach to identify the underlying rule. To that end, behaviorally identified strategies were divided into light-specific (*light* and *dark*), sound-specific (*click* and *silent*), and place-specific (*right*, *left*, and *alternate*) classes. One participant was excluded from this analysis, as she had only completed strategies belonging to two different attention classes.

Average decoding accuracy is above chance at all time points, including before stimulus onset. Decoding average rises from 40 % before stimulus onset to an elongated peak from 250 to 500 *ms* after stimulus onset around 44 %. Following this peak, average accuracy gradually levels out at around 40 %. The complete sensor space shows an average decoding accuracy above chance level, with the left and right parietal, right temporal, and left and right occipital sensor space contributing most to high decoding accuracy for attention-specific strategy classes.



(a) Average decoding accuracy over time, chance level of 33.33% marked by dotted line



(b) Average decoding accuracies per channel for the averaged time window of 0-1 *s* after stimulus onset

Figure 3.8: Average (*n=28*) decoding accuracies over time *(a)* and sensor space *(b)* for attention-specific strategy classes: light specific strategies (*light* and *dark*); sound-specific strategies (*click* and *silent*); and place-specific strategies (*right*, *left*, and *alternate*). Chance level is 33.33 %.

Decoding accuracy above chance in all sensors could indicate distinctly different large-scale networks in respect to the dimension the subjects are attending to. Above chance decoding accuracy before stimulus onset indicates the participants might be set on the stimulus dimension they will attend to for strategy selection before stimulus onset, reflecting the use of top-down attention. The slight peak after stimulus presentation might be related to the consolidation of attention-specifc strategy representation, indicating determination to apply the strategy already planned for once confronted with a specific environmental set of stimuli.

3.2.3 Conclusion - Decoding Analyses

Neural decoding on single-subject level shows strategies are made up of quickly unfolding neural dynamics, initially generalizing over brief periods of time, then evolving into activity patterns sustained over longer periods of time. Single strategies were decodable in time and sensor space on a single-subject-level. Decoding accuracies' average over multiple subjects for popular strategy combinations replicated this result on a cross-subject level. Attention-specific strategy classes attending the same dimension were shown to share neural representation over the complete time course

and sensor space. Both cross-subject analyses showed decoding accuracy above chance over time and the complete sensor space, implicating separate large-scale neural dynamics underlying single strategies as well as common attention-specific neural large-scale representations underlying attention-specific strategy classes. Above average classifier performance preceding stimulus onset for both cross-subject analyses suggests participants are set on 1) which strategy to execute and 2) which stimulus to attend to before stimulus presentation. This underlines the role of top-down attention in modulating strategies and the learning process in humans.

3.3 Correlation of Behavioral Variables with *MEG* Data

`Ft_statfuncorrelation_T` and cluster-based permutation test statistics were used to identify significant correlations between participants' behavioral variables and decoding accuracy topography of attentional strategy classes.

The following behavioral variables obtained from the behavioral paradigm (*StratsRR*, *% of correct trials* and *Number of rules solved*) or from the *CANTAB* assessment were tested on their correlations with topographical decoding accuracies (the explanations of the variables of the tests are taken from *CANTAB* Cognition Handbook, *1st Edition* (Limited, 2015) and the *CANTAB*eclipse Test Administration guide (Limited, 2006)):

Behavioral Variable	*n*	*Explanation of Behavioral Variable*	*Significant cluster identified?*
StratsRR	28	Number of strategies tried out in the random reward phase	no
% of correct trials	28	Percentage of correctly answered trials	no
Number of rules solved	28	Number of rule sequences also featuring the corresponding correct strategy sequence in the behavioral analysis	no

Continued on next page

Table 3.3 – continued from previous page

Behavioral Variable	*n*	*Explanation of Behavioral Variable*	*Significant cluster identified?*
***RTI** mean simple reaction time*	28	Median duration between stimulus onset and button release for correct trials with one possible location for stimulus appearance	**yes**
***RTI** mean 5-choice reaction time*	28	Median duration between stimulus onset and button release for correct trials with 5 possible locations for stimulus appearance	**yes**
***RTI** mean simple movement time*	28	Median time between button release and touching the stimulus for correct trials with one possible location for stimulus appearance	no
***RTI** mean 5-choice movement time*	28	Median time between button release and touching the stimulus for correct trials with 5 possible locations for stimulus appearance	no
***RVP** mean latency*	28	Mean time taken to respond correctly, time windows greater than 1800 *ms* not included	no
***RVP** total hits*	28	Number of correct responses to target sequences	**yes**
***RVP** total misses*	28	Number of target sequences participants failed to respond to	**yes**
***RVP** probability of hits*	28	Probability of a correct response to target sequences (calculated by *(hits)/(hits+misses))*	**yes**
***RVP** probability of false alarms*	28	Probability of a false alarm to non-target sequences (calculated by *(false alarms)/(false alarms + correct rejections))*	no

Continued on next page

Table 3.3 – continued from previous page

Behavioral Variable	*n*	*Explanation of Behavioral Variable*	*Significant cluster identified?*
RVP *A' prime*	28	Uses the probability of hit and probability of misses to give a measure of how good a participant can detect target sequences	**yes**
SWM *mean time to first response*	28	Mean time between problem presentation and first screen touch	no
SWM *mean token-search preparation time*	28	Mean of the time between problem presentation to first screen touch for the first touch; and, for subsequent touches, the times between token retrieval and the next box touch	no
SWM *mean time to last response*	28	Mean time for participants' last response to a problem	no
SWM *between-errors*	28	Number of times a participant revisited a box in which a token had previously been found	**yes**
SWM *within-errors*	28	Number of times a participant revisited a box already found empty during the same search	**yes**
SWM *double-errors*	28	Errors categorized both as between-errors and within-errors	**yes**
SWM *total-errors*	28	Number of times a box is certain to not have contained a blue token was visited by the participant	**yes**
OTS *mean choices to correct*	28	Mean number of unique box choices made on each problem to make the correct choice	no
OTS *problems solved on first choice*	28	Number of problems solved on the first choice	no

Continued on next page

Table 3.3 – continued from previous page

Behavioral Variable	*n*	*Explanation of Behavioral Variable*	*Significant cluster identified?*
***OTS** mean latency to first choice*	28	Mean latency from stimuli appearance until box touch	**yes**
***OTS** mean latency to correct*	28	Mean latency from stimuli appearance until correct box touch	**yes**
***IED** EDS*	28	Number of errors made during an extra-dimensional set shift	no
***IED** total errors (adjusted)*	28	Total number of errors (adjusted *+25* for not completed stages)	no
***IED** total trials (adjusted)*	28	Total number of trials completed on all attempted stages (adjusted *+50* per stage not reached)	no
***IED** completed stage trials*	28	Number of trials employed on all successfully completed stages	**yes**
***PAL** total errors (adjusted)*	28	Total number of errors (adjusted for each stage not completed)	no
***PAL** mean errors to success*	28	Sum of all errors divided by the number of completed stages	no
***PAL** total trials*	28	Total trials required to correctly complete a stage	no
***PAL** total trials (adjusted)*	28	Total trials required to correctly complete a stage (adjusted +10 per uncompleted stage)	no
***PAL** mean trials to success*	28	total trials needed to complete all stages divided per number of completed stages	no
***SSP** span length*	28	Longest sequence successfully reproduced	no
***SSP** total errors*	28	Total number times an incorrect box was selected	no

Continued on next page

Table 3.3 – continued from previous page

Behavioral Variable	*n*	*Explanation of Behavioral Variable*	*Significant cluster identified?*
***SSP** number of attempts*	28	Total number of attempts	no
***SSP** mean time to first response*	28	Mean time to first response, from end of stimulus presentation to screen touch	no
***SST** direction errors on **stop** trials*	28	Number of incorrect responses on stop trials	no
***SST** direction errors on **go** trials*	28	Number of incorrect responses on textbfgo trials	no
***SST** median correct reaction time on **go** trials*	28	Median reaction time in **go** trials with correct responses	no
***SST** stop signal delay*	28	Mean of stop signal delay at which the assessed subject could stop at a 50 % rate, giving a measure of inhibition	no
***SST** stop signal reaction time*	28	gives an estimate of time length between *go* and *stop* stimuli during which participants successfully inhibit responses at a 50% rate (Limited, 2006)	no

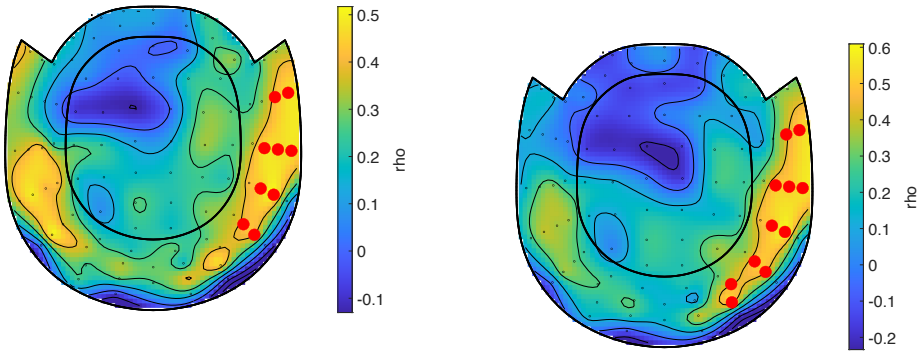
Table 3.3: Correlations of behavioral parameters with topographical decoding accuracies

3.3.1 Reaction Time *RTI*

RTI assesses motor reaction time and cognitive processing speed separately (Limited, 2006). Participants were asked to keep a button pressed; and, when presented with a goal stimulus, to touch this stimulus on screen as quick as possible. Participants

were presented with an increasing number of possible goal stimuli, increasing task demand of attention (Limited, 2015). Outcome measures include reaction time measured from stimulus presentation and button release; as well as movement time from button release to screen touch. These measures are taken from correct trials either in simple trials, with one possible stimulus location, or 5-choice trials with 5 possible stimulus location.

Poor performance as measured by *mean simple reaction time* and *mean 5-choice reaction time* correlated significantly with higher decoding accuracies for attention-specific strategy classes over the right temporal sensor space (figure 3.9).



(a) Correlation of attention-specific decoding topography with *RTI mean simple reaction time*

(b) Correlation of attention-specific decoding topography with *RTI 5 choice reaction time*

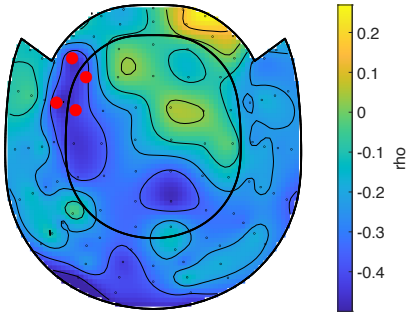
Figure 3.9: *RTI* variables correlated with topographical decoding accuracy of attention-specific strategy classes for *n=28*. The color-bar indicates the corresponding correlation coefficient *RHO*. Significant clusters identified through cluster-based permutation analyses are highlighted in red.

3.3.2 Rapid Visual Processing *RVP*

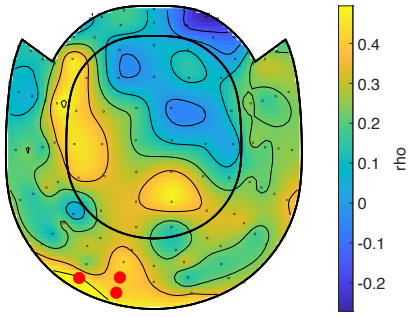
Similarly to *RTI*, *RVP* assesses cognitive processing time but places heavier demands on selective and sustained attention (Limited, 2015). It is a measure of parietal and frontal lobes (Coull & Nobre, 1998). It has been shown to activate the fronto-parieto-occipital attention network (Limited, 2015). Participants are presented with single digits appearing in pseudo-randomized order at a rate of 100 digits/minute. They are instructed to press a button as soon as they detect target

sequences such as i.e. *4-6-8*. Outcome measures include *latency, hits, misses, false alarms*, and *A' (A prime)*. *A prime* provides an index of how good a participant can detect a relevant stimulus amid distracting irrelevant stimuli, representing a measure of selective attention.

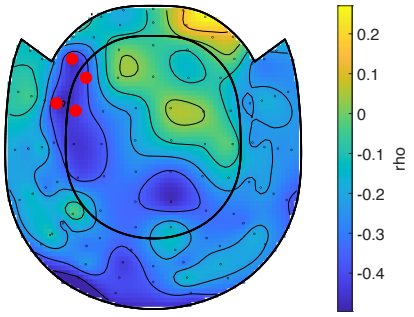
Significant correlation was identified between participants' *RVP* performance and topographical decoding accuracies of attentional strategy classes in left frontotemporal and left occipital sensor space as shown in figure 3.10. Poor performance, as indicated by *total hits* and *probability of hits*, correlated significantly with higher decoding accuracy in the left frontotemporal sensor space. Aditionally, poor performance, as indicated by *total misses* and *A prime*, correlated significantly with higher decoding accuracy of attention-specific strategy classes in the left occipital sensor space.



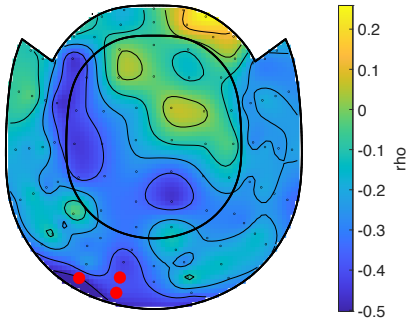
(a) Correlation of attention-specific decoding topography with *RVP total hits*



(b) Correlation of attention-specific decoding topography with *RVP total misses*



(c) Correlation of attention-specific decoding topography with *RVP probability of hits*



(d) Correlation of attention-specific decoding topography with *RVP A'*

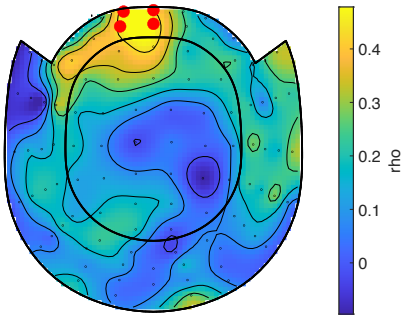
Figure 3.10: *RVP* variables correlated with topographical decoding accuracy of attention-specific strategy classes for *n=28*. The color-bar indicates the corresponding correlation coefficient *RHO*. Significant clusters identified through cluster-based permutation analyses are highlighted in red.

3.3.3 Spatial Working Memory *SWM*

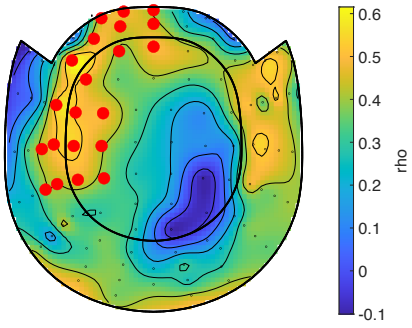
SWM assesses working memory facets such as information storage and incorporation of information into action planning and execution (Limited, 2006). It measures frontal lobe functioning (Duncan & Owen, 2000; A. M. Owen, Evans, & Petrides, 1996). Participants are presented with colored boxes and must find the concealed token in one of the boxes by process of elimination. A new token is then concealed in one of the previously not used boxes. This is repeated until all boxes have been

used once. The number of boxes gradually increases. Error measures include revisiting boxes that had already been found empty or revisiting boxes that had already held a token. Participants are also measured on how well they use a highly efficient strategy as proposed by Owen et al. (A. M. Owen, Downes, Sahakian, Polkey, & Robbins, 1990) to solve the *SWM* task.

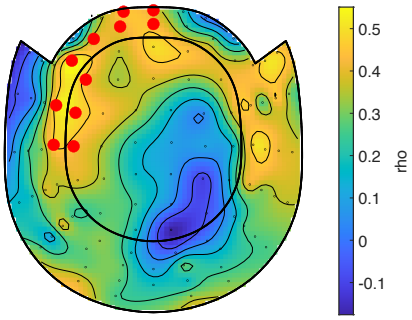
Significant correlation clusters between poor *SWM* performance and topographical decoding accuracies of attentional strategy classes were found in frontal and left temporal sensor spaces as shown in figure 3.11. Poor *SWM* performance as measured by *within-errors* and *double-errors* correlated with higher decoding of attentional strategies in the left frontotemporal sensor space; aditionally, poor performance scores in *between-errors*, *total-errors*, and the *strategy* metric correlated significantly with higher decoding accuracies for attentional strategies in the frontal sensor space.



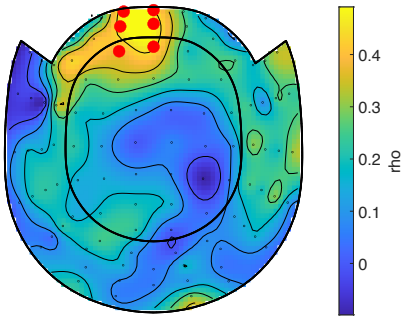
(a) Correlation of attention-specific decoding topography with *SWM between-errors*



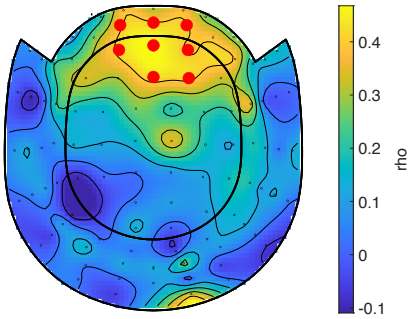
(b) Correlation of attention-specific decoding topography with *SWM within-errors*



(c) Correlation of attention-specific decoding topography with *SWM double-errors*



(d) Correlation of attention-specific decoding topography with *SWM total-errors*



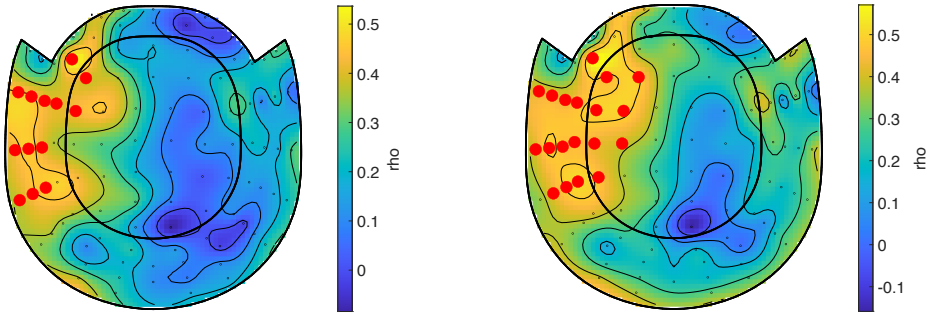
(e) Correlation of attention-specific decoding topography with *SWM strategy*

Figure 3.11: *SWM* variables correlated with topographical decoding accuracy of attention-specific strategy classes for *n=28*. The color-bar indicates the corresponding correlation coefficient *RHO*. Significant clusters identified through cluster-based permutation analyses are highlighted in red.

3.3.4 One Touch Stockings of Cambridge *OTS*

OTS assesses executive functions such as working memory and spatial planning. It is a measure of frontal lobe functioning (Limited, 2006). Participants are presented with two displays of colored balls. Their task is to infer the minimum number of moves required to rearange the lower display into the upper display correctly. Outcome measures include speed and accuracy, the number of problems solved on first choice and latency to first and to correct choice.

Significant correlation were identified, showing poor *OTS* performance as measured through *mean latency to correct* and *mean latency to first choice* correlated with topographical decoding of attention-specific strategies in the left frontotemporal sensor space (figure 3.12).



(a) Correlation of attention-specific decoding topography with *OTS latency to first choice*

(b) Correlation of attention-specific decoding topography with *OTS latency to correct*

Figure 3.12: *OTS* variables correlated with topographical decoding accuracy of attention-specific strategy classes for *n=28*. The color-bar indicates the corresponding correlation coefficient *RHO*. Significant clusters identified through cluster-based permutation analyses are highlighted in red.

3.3.5 Intra- Extra- Dimensional Set Shift *IED*

IED assesses rule learning and set-switching. It assessed visual discrimination and attentional set formation, as well as attention maintenance and flexibility (Limited, 2006). This test has been shown to be sensitive to changes to the fronto-striatal brain areas (Chamberlain et al., 2010). Participants were presented with simple stimuli such as colored shapes or white lines, or compound stimuli comprising both

later on. Obtained response feedback is used to infer the rule-relevant shape. When the learning criterion of six consecutive correct responses is reached, a rule switch occurs, which can either be intra-dimensional (i.e. white lines remain the correct stimulus) or extra-dimensional (i.e. colored shapes are now the correct stimulus). Outcome metrics include the number of trials and stages completed, the number of errors made and latency. See section 1.1.3 and figure 1.3 for more details.

Significant correlations were only found between *completed stage trials* and topographical decoding accuracies of attentional strategy classes as shown in figure 3.13. Poor *IED* performance correlated with a higher decoding accuracy of attention-specific strategy classes in the occipital sensor space.

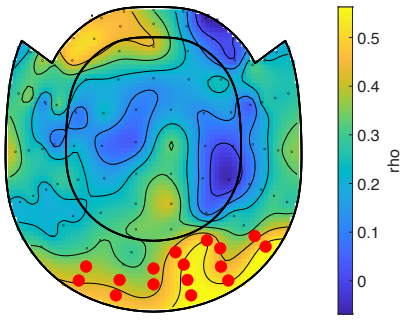


Figure 3.13: *IED completed strage trials* correlated with topographical decoding accuracy of attention-specific strategy classes for *n=28*. The color-bar indicates the corresponding correlation coefficient *RHO*. Significant clusters identified through cluster-based permutation analyses are highlighted in red.

Chapter 4

Discussion

The main goal of this thesis was to investigate whether humans use strategies to infer the correct task rule in multidimensional learning environments. To that aim, a novel rule-learning paradigm implemented in rodent research was adapted and investigated on human behavior using simultaneously recorded *MEG* data. The following questions were investigated in this thesis; ensuing answers this thesis offers are discussed in the next paragraphs:

4.1 Do humans use strategies?

This thesis aimed to investigate if humans use similar behavioral strategies to learn rules as previously identified in rodent research. Results suggest humans use strategies. These can be identified in the behavioral analysis and are decodable on both single-subject and cross-subject analysis of *MEG* data. The temporal generalization method shows strategies are comprised of quickly altering neural dynamics only shortly generalizing before evolving into a more general neural code sustained for a longer time period.

4.1.1 Qualitative Behavioral Analysis

The aim of the qualitative questionnaire was to see if participants could distinctly recall and report their approach to learning the correct rule. The qualitative questionnaire data was not conclusive, as many participants did not recall using a certain approach and often could not articulate what approaches they had used to infer the correct rule. Rule learning and underlying mechanisms are not always clearly obvious, rather, conscious cognitive computations are sometime more implicit (Borghi

et al., 2017; Borghi, Barca, Binkofski, & Tummolini, 2018; Kiefer & Pulvermüller, 2011).

When explicitly asked about the employment of specific strategies (i.e. *"did you use the strategy dark?"*), all participants were able to distinctly recall they had used some of the strategies. While the qualitative results did not permit conclusions regarding the approach, it showed participants had at one or another point employed strategies in the experiment.

4.1.2 Quantitative Results

Behavioral Findings Quantitative behavioral analysis identified strategy use at the behavioral level. This replicates findings in the rodent version of the employed paradigm, as well as behavioral results of the pilot study.

Neural Analysis of Single Strategies Subject and cross-subject analysis of *MEG* data revealed neural decoding above chance of behaviorally identified strategies on both single-subject and cross-subject level. The dynamics uncovered in the single partcipants' *TGM* permit the assumption for the underlying neural code to be sequentially evolving activated neural processes, present only for a short period of time. This entails classifiers only accurately classifying data arising from their individual training time and not able to generalize over longer time periods (King & Dehaene, 2014; Dehaene & King, 2016; Stokes et al., 2013). Single strategies were decodable throughout time and sensor space on participant level. Popular strategy combinations decoding as presented in sections 2.2 and 3.2.2 further support the dinstinctly different neural dynamics coding for single strategies. This hints at strategy-based learning as a mechanism of finding structure and reducing task dimensionality in order to efficiently infer the correct task rule in unknown environments.

These results are concordant with insight from research showing rodents (Powell & Redish, 2016; Bähner et al., 2022) and humans (Donoso et al., 2014; Liu, Braunlich, Wehe, & Seger, 2015; Kachergis, 2018; Song, Baah, Cai, & Niv, 2023) test for assumptions and hypotheses about the environment for accuracy and adapted according to feedback. Further, pilot behavioral results were replicated in the main experiment version, supporting the role of strategy use in flexible learning.

4.2 Are learning processes shaped by abrupt transitions?

Results show learning processes in humans are shaped by change points in reaction time and performance, possibly corresponding to moments of sudden insight. Change points are shown to correlate significantly with testing for the correct strategy rule on a trial-by-trial basis, underlining the relevance of strategy-based testing during learning. Results are consistent with this thesis' pilot experiment.

4.2.1 Abrupt Behavioral Shifts

Averaging responses over numerous sequential trials and multiple subjects leads to viewing learning as gradually evolving processes (Gallistel, Fairhurst, & Balsam, 2004). This does not allow trial-by-trial evaluation, possible notice or investigation of occuring change points. Evidence arising from mice (Rosenberg et al., 2021), rat (Durstewitz et al., 2010), primate (Bartolo & Averbeck, 2020), and human (Leber, Turk-Browne, & Chun, 2008; Aziz-Zadeh et al., 2009; Kachergis, 2018) research suggests learning processes are not gradual. Rather, learning experiences are indicated to be shaped by abrupt behavioral change points, which can be detected in behavioral and neural analyses.

In this thesis, *Paired Adaptive Regressors for Cumulative Sum* (Toutounji & Durstewitz, 2018) were used to identify significant change points from the current value relative to the mean of the preceding values over sequential trials. Both change point metrics investigated in this thesis were found to correlate significantly with each other. This supports the hypothesis learning processes are not gradual but shaped by abrupt transitions. Additionally, these results align themselves with findings of change point shaped learning dynamics in animal and human research corresponding to moments of sudden insight (Bowden, Beeman, Fleck, & Kounios, 2005). Change points corresponding to conscious moments of insight, mirrored in *"Eureka!"* or *"Aha!-Effects"* provide an emotional positively connotated feeling of success in the task at hand and further strengthens memory (Danek, Fraps, V. Müller, Grothe, & Öllinger, 2012; Ludmer, Dudai, & Rubin, 2011).

Further, beginnings of correct strategy sequences correlated significantly with both reaction time and performance change points. The reaction time change point

occurred on average 15 trials later, replicating pilot results: A decrease in reaction time indicating more confidence in the tested strategy being correct developed only after the testing of a strategy; some trials' time may be needed to grasp the strategy tested is correct.

This further underlines the use of strategies as a means of sequentially testing for hypotheses about the correct action in multidimensional environments. This is reflected in abrupt behavioral change points as well as in the significant correlations between change points and the beginning of correct strategy sequences. This supports findings in human (Berens, Horst, & Bird, 2018; Kachergis, 2018; Song et al., 2023) research.

4.3 How is learning guided by attention using strategies?

Behavioral and neural results of this thesis further support the hypothesis of humans using selective attention to prioritize the most probably relevant stimuli in testing for task-relevant dimension features.

4.3.1 Behavioral Results

Research proposes selective attention in humans as a means of reducing task dimensionality and permitting resource-efficient computations inferring the correct course of action when confronted with highly-dimensional ever-changing environments (Niv et al., 2015; Leong et al., 2017; Radulescu, Shin, & Niv, 2021). Strategies can be seen as a means of selective attention modulating the learning process and can be summarized as attending to a specific stimulus dimension and learning the correct response for that stimulus instead of learning the correct stimulus response per state, where all possible variations of stimulus dimensions comprise one state.

Additionally, structure learning hypothesizes humans find structure in their perception and environment based on past knowledge about similar situations (Leong et al., 2017). This allows for generalizing across possible environmental states and across tasks: Preconceptions about the current task and assumptions of the relevant stimulus dimension and hypotheses to test can thus be formulated using information about previous similar situations (Gershman & Niv, 2010). One example of

structure learning is proposed as modulating attention allocation when searching the relevant stimulus dimension.

4.3.2 Neural decoding accross time

Additionally, cross-subject decoding of popular strategy combinations and attention-specific strategy classes showed a distinct temporal course.

Cross-subject decoding of attention-specific strategy classes over time points revealed strategies sharing neural representations of attended stimulus dimensions with different strategies attending the same stimulus dimension, while decoding of popular strategies revealed distinctly different strategy representations for single strategies. Both decoding analyses showed classifier performance above average preceding stimulus onset. Subjects were set on which stimulus to attend to and which strategy to employ before stimulus onset, showing strategy employment was planned and not stimulus-driven. Above chance decoding of attention-driven strategy representation preceding stimulus onset reflects the use of top-down attention in the learning process, a function essential to flexible cognitive processes (Ress, Backus, & Heeger, 2000; Womelsdorf & Everling, 2015; Diamond, 2020) and further enforces the role of attention modulating learning. Single strategy decoding above average before stimulus onset reinforces the concept of structure learning (Gershman & Niv, 2010), showing subjects test for preconceived assumptions and hypotheses about the correct rule.

The role of top-down attention Further, these results reject the idea of the brain as a "passive stimulus-driven device" (Engel et al., 2001) and underline the importance of computations taking place before stimulus onset. This aligns itself with findings emphasizing the role of top-down attention modulating perception to enhance relevant stimuli (Menon & Uddin, 2010; Katsuki & Constantinidis, 2013), while rejecting the view of neuronal activity merely being a reflection of the outside world. Rather than viewing perception as a passive, externally-driven sequential processing of external information building an internal neural reflection to most accurately depict the outside world, research suggests viewing cognition as action and goal-oriented (Vossel, Geng, & Friston, 2014). This entails cognition being subjected to goal-oriented modulation during perception itself. Top-down attention is a means of efficiently allocating limited neural resources to achieving this end.

Decoding accuracy above chance preceding stimulus presentation could mean subjects are set on which strategy they will execute and which stimulus attention will be allocated to. The decision is already deterministically locked in before stimulus onset. This could thus be a neural correlate of an abstract rule not heavily influenced by the stimuli themselves.

4.3.3 Sensor-level analysis of attention-specific strategy classes

Networks such as the dorsal attention network, enforcing top-down attention and comprised of the frontal eye field and the intraparietal sulcus, as well as the ventral attention network, enforcing bottom-up attention and comprised of the ventrolateral and dorsolateral prefrontal cortices, the right temporo-parietal junction, the anterior cingulate cortex, and the premotor cortex have been implicated in attention allocation to stimuli in goal-directed behavior (Katsuki & Constantinidis, 2013; Vossel et al., 2013; Uddin, 2021). Research further suggests attentional control arising from multiple distributed prefrontal, parietal, and temporo-parieto-occipital cortices enhancing and modulating neuronal synchrony in brain areas representing the attended dimension (Engel et al., 2001; Siegel, Donner, & Engel, 2012; Womelsdorf & Everling, 2015; Vossel, Weidner, Moos, & Fink, 2016).

These thesis' results show decoding accuracy of attention-specific strategy classes above chance level in all channels; the left and right parietal, left temporal, and occipital sensor space contributing most to high decoding accuracy of attention-specific strategies. This indicates stimulus dimension-specific large-scale brain dynamics involved in attention allocation by modulating attention-specific areas and could lead to decoding accuracy above average in all channels, peak accuracy in left and right parietal, left temporal, and occipital sensor space could correspond to specific brain regions representing attentional control and modulated by aforementioned large networks.

Although it seems plausible the activation of regions more heavily implicated in attentional control such as i.e. the frontal eye field, intraparietal sulcus, right temporo-parietal junction and the temporo-parieto-occipital cortex could correspond to peak decoding accuracy over the parietal, temporal and occipital sensor space, source reconstruction analyses would be needed to make a statement confirming or denying that claim.

While results hereby remain inconclusive regarding source localization, decoding accuracies above chance for stimulus-specific strategy classes over the complete sensor space show strategies related to the same task feature share common neural large-scale brain representations delimiting them from strategy classes allocating attention to another stimulus. This further supports the claim that strategies are a means of attention guiding learning and aligns itself with the claim of large-scale brain dynamics modulating attentional control (Engel et al., 2001; Womelsdorf & Everling, 2015).

Conclusion Consequently, the detection of strategies reflects the use of selective attention and structure-based learning. Strategies constitute an important mechanism enabling flexible learning in humans.

4.4 Does decoding accuracy correlate with validated neuropsychological assessment in coherent sensor spaces?

CANTAB is a standardized neuropsychological assessment tool initially developed to assess cognitive functions in aging subjects with dementia in the 1980s (Sahakian et al., 1988). It has since been developed to assess multiple cognitive subdomains such as executive funtions, visual and verbal memory, attention, decision making and response control (Sandberg, 2011). It has been externally validated as adequately providing measures of the assessed cognitive subdomains (Lowe & Rabbitt, 1998) both in healthy subjects as well as in psychiatric and neurological patients (Fray, Robbins, & Sahakian, 1996). It assesses functions ranging from the age cognitive functions arise in development to the age in which they decline (Luciana, 2003). Participants' performance on selected *CANTAB* tasks (see table 3.3) was correlated with topographical decoding accuracy of attention-specific strategy classes using cluster-based permutation statistics. The ensuing resulting correlation topographies are discussed per test individually, while section 4.4.6 will aim to draw overall conclusions.

4.4.1 Reaction Time *(RTI)*

RTI assesses motor reaction time and cognitive processing speed (Limited, 2006). Poor performance as measured with *mean simple reaction time* and *mean 5-choice*

reaction time correlated significantly with a higher decoding accuracy for attention-specific strategy classes in the right temporal sensor space. Importantly, *mean simple movement time* and *mean 5-choice movement time* showed no correlations with decoding accuracies. *reaction time* quantifies *RTI* cognitive demands while *movement time* allows for separation of motoric components by assessing motoric slowing. While *simple* trials measure cognitive processing speed, attentional demand is increased in *5-choice* trials (Limited, 2006, 2015).

The lateral prefrontal and right inferior parietal cortices have been linked to cognitive processing speed in human reaction time tasks (Limited, 2015; Ohata, Ogawa, & Imamizu, 2016). Primate research has further identified the intraparietal cortex, the frontal eye field and the middle temporal area as key regions (Schall, 2003). It is conceivable *RTI* performance relying on the lateral prefrontal and especially the right inferior parietal cortices could correlate with neural discriminatory information regarding attentional strategy classes. While this relationship could plausibly project to the right temporal sensor space, further source reconstruction would be needed to verify that claim.

4.4.2 Rapid Visual Processing *(RVP)*

Similarly to *RTI*, *RVP* assesses cognitive processing time but places heavier demands on selective and sustained attention (Limited, 2015). Poor *RVP* performance as measured by *total hits* and *probability of hits* correlated with higher decoding accuracies for attentional strategy classes in the left fronto-parietal sensor space. Poor *RVP* performance measured by *A'* and *total misses* correlated with higher decoding accuracies in the left occipital space. Importantly, these measures and their variance correlated with each other, leading the ensuing correlation maps to directly correlate as well.

Among various regions distributed along frontal, parietal and occipital cortices, the left ventral posterior visual cortex, the left motor and premotor, as well as the left inferior parietal cortex have been most heavily linked to sustained attention during human rapid visual processing tasks (Limited, 2015; Coull & Nobre, 1998). Sustained attention relying on left motor, premotor, parietal and occipital cortices as assessed in *RVP* could conceivably project to correlating with decoding accuracies of attentional strategy classes in identified left frontoparietal and left occipital sensor

spaces.

4.4.3 Spatial Working Memory *(SWM)*

SWM assesses working memory functions such as information storage and incorporation of information into action planning and execution. By measuring how well subjects adhere to the most ideal strategy to solve the task, it provides a measure of executive functioning (Limited, 2006). An intact prefrontal cortex is traditionally associated with good performance in the *SWM* task (A. M. Owen et al., 1990; Manes et al., 2002). Further, the dorsolateral and mid-ventrolateral prefrontal cortices have been specifically indentified as functionally relevant in solving the *SWM* task (Mehta, Owen, Mavaddat, Pickard, & Robbins, 2000).

Poor *SWM* performance correlated with higher decoding accuracy of attentional strategy classes in frontal and left temporal sensor space. This suggests frontal sources required when using working memory playing a role in attention allocation during the examined rule-learning task. While it seems plausible prefrontal regions required to performing well in *SWM* (A. M. Owen et al., 1990; Mehta et al., 2000) would project to higher topographical decoding accuracies for attentional strategies in frontal sensor space, further source reconstruction would be needed to confirm or deny that claim.

Neural analysis of a *SWM*-like task in macaques showed poor performance to correlate with higher recorded activity for representation of search strategies in the lateral prefrontal cortex (Chiang & Wallis, 2018). Poor *SWM* performers would require stronger strategy representation and need to make a larger effort to solve a task. This supports human findings, as it indicates activity differences leading to better differentiation are related to performance parameters.

4.4.4 One Touch Stockings of Cambridge *(OTS)*

OTS assesses working memory and spatial planning (Limited, 2006). *OTS* and *SOC* (Stockings of Cambridge) are derived from the towers of London and Hanoi tasks. While these tests, assessing spatial working memory and planning are dependent on frontal lobe functioning, *SOC* places heavier demands on motor sequencing, which is more strongly linked to left frontoparietal regions, while *OTS* is designed to place a heavier load on mental imagery, more strongly linked to right frontoparietal

regions (A. M. Owen et al., 1996; Limited, 2006; Wild & Musser, 2014). Further, the right parietal cortex has multiply been linked to planning, working memory and ensuing attention allocation (A. M. Owen et al., 1996; Berryhill & Olson, 2008; Langel, Hakun, Zhu, & Ravizza, 2014).

Poor *OTS* performance correlated with higher decoding accuracy of attentional strategy classes over the left frontotemporal sensor space. While this contradicts feasible projections of right frontoparietal activity relevant to working memory and spatial planning to coherent sensor spaces, it is important to realize the only *OTS* measures correlating were *mean latency to correct* and *mean latency to first choice*, both measures of reaction time. As noted above, both *OTS* and *SOC* rely on frontoparietal regions. While *SOC* places a heavier load on motoric sequencing, more reliant on left regions, *OTS mean latency* measures do, too. Poor *mean latency* performances reliant on left frontoparietal regions could plausibly project on the left frontotemporal space regarding correlations with decoding accuracy of attentional classes.

4.4.5 Intra- Extra- Dimensional Set Shift *(IED)*

IED, a computerized variant of the Wisconsin Card Sorting Test (see figures 1.2 and 1.3 for detailled depiction and explanation) assesses attentional set-shifting (Wild & Musser, 2014). It is traditionally associated with frontostriatal functioning (A. Owen et al., 1993; Limited, 2006). Prefrontal and parietal areas have further been identified as functionally relevant in humans during attentional set-shifting (Periáñez, Maestú, Barceló, Amo, & Alonso, 2004; Oh, Vidal, Taylor, & Pang, 2014).

Cluster-based statistics showed no significant clusters between *IED* performance metrics and the decoding topography of attention-specific strategy classes, except for *completed stage trials* correlating with topographical decoding accuracy over the occipital sensor space. It is assumed prefrontal activation dependent of *IED* performance would not project to occipital sensor clusters. As explained in part 4.4.6, the statistical method employed here represents a compromise in picking a threshold adequately sensitive to statistically significant effects while maintaining a low false alarm rate. Considering no other significant clusters were identified between *IED* performance and decoding topography of attentional strategy classes

except for *completed stage trials* in a non-coherent, occipital sensor space, this sensor space correlation is concluded to represent a false positive outlier and discarded as such.

4.4.6 Conclusions to draw from the correlation of topographical decoding accuracies and *CANTAB* performance

Cluster-based statistical methods are employed here as a means of reducing *type I errors* (mistakenly rejecting the null hypothesis) in multiple comparisons by defining a threshold for statistical tests (Oostenveld, Spaak, Delorme, & Arana, 2018). The multiple comparison problem arises when comparing a large set of i.e. *channel-accuracy* pairs: Given an imense number of statistical comparisons (i.e. one per *channel-accuracy*-pair) are performed to evaluate the effect of interest (i.e. statistical dependence on decoding accuracy per channel on *CANTAB* performance parameters to identify relevant sensor space), the false alarm rate arising from the sum of *type I errors* among a group of tests (Tukey, 1973) is not controllable. To control for that error, statistical methods are put to use to control the false alarm rate at α level of i.e. *0.05* (see section 2.2.5) (Maris & Oostenveld, 2007). While employing cluster-based permutation statistics reduce the false alarm rate, it represents compromising for a treshold adequately sensitive to statistically significant effects while maintaining a low false alarm rate.

As discussed above, regions specifically implicated in cognitive functions as assessed by *CANTAB* projected to coherent sensor spaces for decoding of attentional strategy classes. There are two major limitations regarding the employed methods to take into consideration: Firstly, cluster-based statistics show a significant effect, but not where exactly or how strong that effect is (Sassenhagen & Draschkow, 2019). Secondly, as calculations were executed on sensor level and not on source level, source reconstructions would be needed to make more detailled assumptions about involved brain regions. Since source localization of identified effects is already limited by employing cluster-level statistics, source reconstruction analysis would not have led to a completely straightforward spatial effect demonstration. Rather, the interpretation and embedding of identified effects are key factors when drawing conclusions (Oostenveld, Gonaz, Sassenhagen, & Schoffelen, 2018).

Plausability of identified effects is suggested by the relationship between *CANTAB* performance and topographical decoding of attentional strategies: Identified projections of correlations between *CANTAB* parameters and decoding of attentional strategies corresponded to *CANTAB*-task specific brain regions (see previous sections) relevant in executive functioning. Plausability is further suggested by correlation of multiple *CANTAB* test-specific parameters and attentional decoding projecting onto similar sensor space.

For example, *SWM* parameters' significant correlations with topographical decoding all projected upon the test-coherent frontal sensor space. Cognitive functions examined by the discussed *CANTAB* tests are cognitive processing speed (*RTI*), selective and sustained attention (*RVP*), working memory (*SWM* and *OTS*), incorporation of information into action planning (*SWM*) and spatial planning (*OTS*). Is it conceivable these functions as quantified by the respective *CANTAB* performance parameters may be closely related and play a role in the employment of attention-guided strategies in a learning task. This could lead to correlations between decoding of attention-specific strategies and respective *CANTAB* performance parameters. This further underlines the plausability of the significant relationship between various externally validated cognitive variables and the decoding accuracy of attention-specific strategies.

Importantly, plausability is further suggested by both *OTS* and *SWM* variables and their correlations to attentional decoding projecting to the left frontotemporal sensor space. Both tests have been shown to place demands on working memory and planning (A. M. Owen et al., 1990). It has been shown that *SWM* and *OTS* are sensible to frontal cortex lesions, whereas *SSP (spatial span)* marked no sensitivity to such lesions (A. M. Owen et al., 1990). More specifically, activity in the dorsolateral frontal cortex has been shown to rise during tasks with higher working memory demands (Bor, Duncan, Wiseman, & Owen, 2003). Distinctively, none of the *SSP* variables were shown to correlate with attentional decoding, as this test focusses more on short term spatiovisual memory and less on working memory or action planning.

The latency variables in *OTS* correlating with attentional decoding are both associated with action planning. The strategy measure in *SWM* is presented as an ideal strategy most able to simplify task demands in *SWM*. Analogous to this, the

strategies presented and analyzed in this thesis are a means of simplifying task demands and reducing working memory load and present a valid expression of action planning and executive functioning. This claim is supported by the identified correlations between attentional strategy decoding and the multiple *SWM* and *OTS* variables projecting onto the very plausible left frontotemporal sensor space, possibly corresponding to the dorsolateral prefrontal cortex.

Despite aforementioned limitations, results of this thesis were able to show poor performance on multiple *CANTAB* tests correlated with decoding of attentional strategy classes in test-specific coherent sensor spaces. An important notion of decoding analyses is that high classification accuracies do not reflect if a subject performed well. Rather, it means discriminative information regarding decoded classes was highest as measured by sensors with the highest decoding accuracies. As such, correlations of *CANTAB* performance with decoding topographies of attentional strategy classes do not reflect if a subject performed well in the multidimensional rule-learning paradigm employed. Rather, they showed poor *CANTAB* performers displayed more discriminative information for classification of attentional strategy classes in identified sensor spaces.

Further adressing the previous point, poor *CANTAB* performers could require stronger strategy representation, thus higher neural and cognitive efforts to master equivalent cognitive prowesses as good *CANTAB* performers in the employed paradigm. Good *CANTAB* performers with better attention and working memory abilities could rely more on efficient and correct functioning of employed brain regions, and thus show less strategy representation and activation of the task-relevant regions. Showing brain regions essential to executive functioning were linked to discriminative information regarding decoding of attentional strategy classes confirms the employed multidimensional learning task's validity. In combination with the correlation of behavioral variables in the employed paradigm and *CANTAB* behavioral variables, as well as replicating rodent results and pilot results, the employed multidimensional rule-learning paradigm constitutes a valid means of quantifying and uncovering learning mechanisms in flexible behavior.

4.5 Translation of rodent results and outlook

Rats have been shown to use strategies in the corresponding learning task as introduced in section 1.1.3. This has been demonstrated on a behavioral as well as on a neural level (Hermann, 2022; Bähner et al., 2022). Abrupt transitions in behavior have been identified during the learning process in rats. Significant correlations were identified between change points and strategy use in rats (Hermann, 2022). These results translated to human findings in this thesis. Further, using reinforcement learning models and an eye-tracking camera to assess rat head position, body position and movement as an attention marker enabled showing the close relationship between selective attention and learning in rats (Bähner et al., 2022). These results further translate to this thesis' results, where a strong relationship was found between decoding of attentional strategies and independent and validated markers of executive functioning in humans. In combination with rodent results, this thesis aims at an understanding of cognitive, behavioral and neural processes underlying flexible decision-making and provides a foundation for investigating impaired cognitive flexibility in clinical populations. A possible outlook may be to deepen this translational link using both eye-tracking and reinforcement learning models in human research. Extending this work to clinical populations suffering from cognitive deficits such as, i.e., patients with schizophrenia or Parkinson's disease (Uddin, 2021) may permit a better understanding of the associated neurocognitive deficits.

Chapter 5

Summary

Cognitive flexibility, enabling behavioral and cognitive adjustments to ever-changing environments is essential to survival. It is often examined using rule-learning and rule-switching paradigms. Current learning models often operate accurately in environments with few dimensions but fail to perform accurately in multidimensional settings such as our outer world. They also fail to account for moments of sudden insight shaping learning processes.

The employed multidimensional learning paradigm offers a solution as to how humans achieve a reduction of the multidimensional outer world by attention-modulated trial-by-trial testing for strategies as well as offering the employment of strategies as an explanation to moments of sudden insight occuring in learning. It further achieves to demonstrate that sequentially activated neural dynamics constitute behavioral strategies on a single subject-level, as well as showing strategies are differentially neurally encoded both on single- and cross-subject level; while displaying strategies attending to the same stimulus share common whole-brain representation differentiating them from strategies attending to other stimuli. It shows participants were set on which strategy they would execute, and which stimulus dimension they would attend to preceding stimulus onset. This reflects the use of top-down attention in attention-modulated strategy use during learning.

The paradigm is validated both at the behavioral and neural level through the correlation with behavioral variables obtained during independent and validated neuropsychological testing of executive functioning. At the behavioral level, participants' performance in the rule-learning and switching task *IED* transferred to per-

formance metrics in the empoyed paradigm. At the neural level, decoding accuracy of attentional strategies correlated with OTS and SWM performance parameters in the test-specific frontal sensor space. This notably underlines the employement of strategies as an efficient means of executive planning and task organization.

These findings validate the employed multidimensional rule-learning paradigm as an effective assessment of executive functions in learning. Together with rodent results, this paradigm is suited as a translational paradigm. The implication and importance of these questions stem from the number of neuropsychiatric disorders affected by cognitive inflexibility, with hope of a better understanding giving way to better prevention, diagnose and treatment for affected patients.

Figures

1.1 Core regions of networks involved in cognitive processes 6
1.2 Wisconsin Card Sorting Test 8
1.3 Intra- Extra- Dimensional Set Shift Test 9
1.4 Rat Paradigm 10
1.5 Magnetic effect of electric currents 13
1.6 Magnetometer, axial, and planar gradiometers 14
1.7 Univariate Analysis 18
1.8 Multivariate Pattern Analysis 19

2.1 Trial Depiction 24
2.2 Visualisation of the strategy detection algorithm 32

3.1 Strategies behaviorally identified 41
3.2 Strategies and *CUSUM* plots 42
3.3 Correlation analysis for performance change points and 1st correct trials 43
3.4 Correlation analysis for reaction time change points and 1st correct trials 44
3.5 Correlation analysis for reaction time and performance change points 45
3.6 Single Participant Classification Analyses 49
3.7 Decoding Analyses for *dark, click* and *silent* 51
3.8 Decoding Analyses for attention-specific strategy classes 52
3.9 Significant Clusters *RTI* 58
3.10 Significant Clusters *RVP* 60
3.11 Significant Clusters *SWM* 62
3.12 Significant Clusters *OTS* 63
3.13 Significant Clusters *IED* 64

Tables

2.1 Description of detectable strategies 31
2.2 Description of computed classifications and according classes 37

3.1 Description of detectable strategies 40
3.2 Correlations of behavioral parameters with *CANTAB* performance metrics 46
3.3 Correlations of behavioral parameters with topographical decoding accuracies 57

Bibliography

Ahonen, A., Hämäläinen, M., Ilmoniemi, R., Kajola, M., Knuutila, J., Simola, J., & Vilkman, V. (1993, 10). Sampling theory for neuromagnetic detector arrays. *IEEE transactions on bio-medical engineering*, *40*, 859-69. doi: 10.1109/10.245606

Alamian, G., Hincapié, A., Combrisson, E., Thiery, T., Martel, V., Altukhov, D., & Jerbi, K. (2017, 03). Alterations of intrinsic brain connectivity patterns in depression and bipolar disorders: A critical assessment of magnetoencephalography-based evidence. *Frontiers in Psychiatry*, *8*. doi: 10.3389/fpsyt.2017.00041

Alamian, G., Hincapié, A., Pascarella, A., Thiery, T., Combrisson, E., Saive, A.-L., ... Jerbi, K. (2017, 07). Measuring alterations in oscillatory brain networks in schizophrenia with resting-state meg: State-of-the-art and methodological challenges. *Clinical Neurophysiology*, *128*. doi: 10.1016/j.clinph.2017.06.246

Alvarez, J., & Emory, E. (2006, 04). Executive function and the frontal lobes: A meta-analytic review. *Neuropsychology review*, *16*, 17-42. doi: 10.1007/s11065-006-9002-x

Aron, A., Robbins, T., & Poldrack, R. (2014, 04). Inhibition and the right inferior frontal cortex: One decade on. *Trends in Cognitive Sciences*, *18*, 177-185. doi: 10.1016/j.tics.2013.12.003

Aziz-Zadeh, L., Kaplan, J., & Iacoboni, M. (2009, 03). Aha!: The neural correlates of verbal insight solutions. *Human brain mapping*, *30*, 908-16. doi: 10.1002/hbm.20554

Bailey, C. (2007, 12). Cognitive accuracy and intelligent executive function in the brain and in business. *Annals of the New York Academy of Sciences*, *1118*, 122-41. doi: 10.1196/annals.1412.011

Baillet, S. (2017, 02). Magnetoencephalography for brain electrophysiology and imaging. *Nature Neuroscience*, *20*, 327-339. doi: 10.1038/nn.4504

Baler, R., & Volkow, N. (2007, 01). Drug addiction: The neurobiology of disrupted self-control. *Trends in molecular medicine*, *12*, 559-66. doi: 10.1016/j.molmed.2006.10.005

Bartolo, R., & Averbeck, B. (2020, 04). Prefrontal cortex predicts state switches during reversal learning. *Neuron*, *106*. doi: 10.1016/j.neuron.2020.03.024

Becker, K., Heinrichs-Graham, E., Fox, H., Robertson, K., Sandkovsky, U., O'Neill, J., ... Wilson, T. (2013, 12). Decreased meg beta oscillations in hiv-infected older adults during the resting state. *Journal of neurovirology*, *19*. doi: 10.1007/s13365-013-0220

-8

Beddington, J., Cooper, C., Field, J., Goswami, U., Huppert, F., Jenkins, R., ... Thomas, S. (2008, 11). The mental wealth of nations. *Nature*, *455*, 1057-60. doi: 10.1038/4551057a

Ben-Zion, Z., Fine, N., Keynan, J., Admon, R., Green, N., Halevi, M., ... Shalev, A. (2018, 10). Cognitive flexibility predicts ptsd symptoms: Observational and interventional studies. *Frontiers in Psychiatry*, *9*, 477. doi: 10.3389/fpsyt.2018.00477

Berens, S., Horst, J., & Bird, C. (2018, 03). Cross-situational learning is supported by propose-but-verify hypothesis testing. *Current Biology*, *28*. doi: 10.1016/j.cub.2018.02.042

Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of General Psychology*, *39*(1), 15-22. Retrieved from `https://doi.org/10.1080/00221309.1948.9918159` (PMID: 18889466) doi: 10.1080/00221309.1948.9918159

Berryhill, M., & Olson, I. (2008, 06). Is the posterior parietal lobe involved in working memory retrieval? evidence from patients with bilateral parietal lobe damage. *Neuropsychologia*, *46*, 1775-1786. doi: 10.1016/j.neuropsychologia.2008.03.005

Blakeman, S., & Mareschal, D. (2022, 03). Selective particle attention: Rapidly and flexibly selecting features for deep reinforcement learning. *Neural Networks*, *150*. doi: 10.1016/j.neunet.2022.03.015

Bock, A. M., Gallaway, K. C., & Hund, A. M. (2015). Specifying links between executive functioning and theory of mind during middle childhood: Cognitive flexibility predicts social understanding. *Journal of Cognition and Development*, *16*(3), 509-521. Retrieved from `https://doi.org/10.1080/15248372.2014.888350` doi: 10.1080/15248372.2014.888350

Bor, D., Duncan, J., Wiseman, R., & Owen, A. (2003, 02). Encoding strategies dissociate prefrontal activity from working memory demand. *Neuron*, *37*, 361-7. doi: 10.1016/S0896-6273(02)01171-6

Borghi, A., Barca, L., Binkofski, F., & Tummolini, L. (2018, 08). Varieties of abstract concepts: Development, use and representation in the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*, 20170121. doi: 10.1098/rstb.2017.0121

Borghi, A., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., & Tummolini, L. (2017, 01). The challenge of abstract concepts. *Psychological Bulletin*, *143*. doi: 10.1037/bul0000089

Borzyszkowska, A., & Basińska, M. (2020, 06). Two types of flexibility: in coping and cognitive, and their relationship with stress among firefighters. *Medycyna Pracy*, *71*. doi: 10.13075/mp.5893.00928

Bowden, E., Beeman, M., Fleck, J., & Kounios, J. (2005, 08). New approaches to demystifying insight. *Trends in cognitive sciences*, *9*, 322-8. doi: 10.1016/j.tics.2005.05.012

Brockmeyer, T., Febry, H., Leiteritz-Rausch, A., Wünsch-Leiteritz, W., Leiteritz, A., & Friederich, H.-C. (2022, 02). Cognitive flexibility, central coherence, and quality of life in anorexia nervosa. *Journal of Eating Disorders*, *10*. doi: 10.1186/s40337-022-00547-4

Brown, T. E., & Landgraf, J. M. (2010). Improvements in executive function correlate with enhanced performance and functioning and health-related quality of life: Evidence from 2 large, double-blind, randomized, placebo-controlled trials in adhd. *Postgraduate Medicine*, *122*(5), 42-51. Retrieved from `https://doi.org/10.3810/pgm.2010.09.2200` (PMID: 20861587) doi: 10.3810/pgm.2010.09.2200

Bähner, F., Popov, T., Hermann, S., Böhme, N., Merten, T., Zingone, H., ... Durstewitz, D. (2022, 11). *Species-conserved mechanisms of cognitive flexibility in complex environments.* doi: 10.1101/2022.11.14.516439

Cambridge Cognition, C. (2022). *Cognitive tests.* Retrieved from `https://www.cambridgecognition.com/cantab/cognitive-tests/`

Carlson, T., Tovar, D., Alink, A., & Kriegeskorte, N. (2013, 08). Representational dynamics of object vision: The first 1000 ms. *Journal of vision*, *13*. doi: 10.1167/13.10.1

Chamberlain, S., Robbins, T., Winder-Rhodes, S., Müller-Sedgwick, U., Sahakian, B., Blackwell, A., & Barnett, J. (2010, 11). Translational approaches to frontostriatal dysfunction in attention-deficit/hyperactivity disorder using a computerized neuropsychological battery. *Biological psychiatry*, *69*, 1192-203. doi: 10.1016/j.biopsych.2010.08.019

Chamberlain, S., Solly, J., Hook, R., Vaghi, M., & Robbins, T. (2021, 02). Cognitive inflexibility in ocd and related disorders. *Current topics in behavioral neurosciences*, *49*. doi: 10.1007/7854_2020_198

Chiang, F.-K., & Wallis, J. (2018, 04). Spatiotemporal encoding of search strategies by prefrontal neurons. *Proceedings of the National Academy of Sciences*, *115*, 201805044. doi: 10.1073/pnas.1805044115

Collins, A., Cavanagh, J., & Frank, M. (2014, 03). Human eeg uncovers latent generalizable rule structure during learning. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *34*, 4677-85. doi: 10.1523/JNEUROSCI.3900-13.2014

Corbetta, M., & Shulman, G. (2002, 04). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews. Neuroscience*, *3*, 201-15. doi: 10.1038/nrn755

Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both pet and fmri. *Journal of Neuroscience*, *18*(18), 7426–7435. Retrieved from `https://www.jneurosci.org/content/18/18/7426` doi: 10.1523/JNEUROSCI.18-18-07426.1998

Dajani, D., & Uddin, L. (2015, 09). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends in Neurosciences*, *38*, 571-578. doi: 10.1016/j.tins.2015.07.003

Danek, A., Fraps, T., V. Müller, A., Grothe, B., & Öllinger, M. (2012, 09). Aha! experiences leave a mark: Facilitated recall of insight solutions. *Psychological Research*, *77*. doi: 10.1007/s00426-012-0454-8

Davidson, M., Amso, D., Anderson, L., & Diamond, A. (2006, 02). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*, 2037-78. doi: 10.1016/j.neuropsychologia.2006.02.006

Davis, J., Marra, C., Najafzadeh, M., & Liu-Ambrose, T. (2010, 04). The independent contribution of executive functions to health related quality of life in older women. *BMC geriatrics*, *10*, 16. doi: 10.1186/1471-2318-10-16

D'Cruz, A.-M., Ragozzino, M., Mosconi, M., Shrestha, S., Cook, E., & Sweeney, J. (2013, 03). Reduced behavioral flexibility in autism spectrum disorders. *Neuropsychology*, *27*, 152-60. doi: 10.1037/a0031721

Dehaene, S., & King, J.-R. (2016, 05). Decoding the dynamics of conscious perception: The temporal generalization method. In (p. 85-97). doi: 10.1007/978-3-319-28802-4_7

de Santana, A. N., Roazzi, A., & Nobre, A. P. M. C. (2022). The relationship between cognitive flexibility and mathematical performance in children: A meta-analysis. *Trends in Neuroscience and Education*, *28*, 100179. Retrieved from `https://www.sciencedirect.com/science/article/pii/S2211949322000096` doi: https://doi.org/10.1016/j.tine.2022.100179

Diamond, A. (2005, 02). Attention-deficit disorder (attention-deficit/hyperactivity disorder without hyperactivity): A neurobiologically and behaviorally distinct disorder from attention-deficit/hyperactivity disorder (with hyperactivity). *Development and psychopathology*, *17*, 807-25. doi: 10.1017/S0954579405050388

Diamond, A. (2020, 01). Executive functions. *Handbook of clinical neurology*, *173*, 225-240. doi: 10.1016/B978-0-444-64150-2.00020-4

Dirnberger, G., & Jahanshahi, M. (2013, 09). Executive dysfunction in parkinson's disease: A review. *Journal of neuropsychology*, *7*, 193-224. doi: 10.1111/jnp.12028

Donoso, M., Collins, A., & Koechlin, E. (2014, 05). Human cognition. foundations of human reasoning in the prefrontal cortex. *Science (New York, N.Y.)*, *344*. doi: 10.1126/science.1252254

Downes, J., Roberts, A., Sahakian, B., Evenden, J., Morris, R., & Robbins, T. (1989, 02). Impaired extra-dimensional shift performance in medicated and unmedicated parkinson's disease: Evidence for a speciwc attentional dysfunction. *Neuropsychologia*, *27*, 1329-43. doi: 10.1016/0028-3932(89)90128-0

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences*, *23*(10), 475-483. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0166223600016337` doi: https://doi.org/10.1016/S0166-2236(00)01633-7

Durstewitz, D., Vittoz, N., Floresco, S., & Seamans, J. (2010, 05). Abrupt transitions

between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron*, *66*, 438-48. doi: 10.1016/j.neuron.2010.03.029

Eakin, L., Minde, K., Hechtman, L., Ochs, E., Krane, E., Bouffard, R., ... Looper, K. (2004). The marital and family functioning of adults with adhd and their spouses. *Journal of Attention Disorders*, *8*(1), 1-10. Retrieved from `https://doi.org/10.1177/108705470400800101` (PMID: 15669597) doi: 10.1177/108705470400800101

Engel, A., Fries, P., & Singer, W. (2001, 11). Engel, a.k., fries, p. & singer, w. dynamic predictions: oscillations and synchrony in top-down processing. nat. rev. neurosci. 2, 704-716. *Nature reviews. Neuroscience*, *2*. doi: 10.1038/35094565

Engel de Abreu, P., Abreu, N., Nikaedo, C., Puglisi, M., Tourinho, C., Miranda, M., ... Engel, P. (2014, 06). Executive functioning and reading achievement in school: A study of brazilian children assessed by their teachers as poor readers. *Frontiers in Psychology*, *5*. doi: 10.3389/fpsyg.2014.00550

Engels, M., Flier, W., Stam, C., Hillebrand, A., Scheltens, P., & Straaten, E. (2017, 05). Alzheimer's disease: The state of the art in resting-state magnetoencephalography. *Clinical Neurophysiology*, *128*. doi: 10.1016/j.clinph.2017.05.012

Evans, J., Olm, C., McCluskey, L., Elman, L., Boller, A., Moran, E., ... Grossman, M. (2015, 03). Impaired cognitive flexibility in amyotrophic lateral sclerosis. *Cognitive and behavioral neurology : official journal of the Society for Behavioral and Cognitive Neurology*, *28*, 17-26. doi: 10.1097/WNN.0000000000000049

Farashahi, S., Rowe, K., Aslami, Z., Lee, D., & Soltani, A. (2017, 11). Feature-based learning improves adaptability without compromising precision. *Nature Communications*, *8*. doi: 10.1038/s41467-017-01874-w

Fray, P., Robbins, T., & Sahakian, B. (1996, 04). Neuropsychiatric applications of cantab. *International Journal of Geriatric Psychiatry*, *11*, 329-336. doi: 10.1002/(SICI)1099-1166(199604)11:4⟨329::AID-GPS453⟩3.0.CO;2-6

Fyshe, A. (2020, 02). Studying language in context using the temporal generalization method. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*, 20180531. doi: 10.1098/rstb.2018.0531

Gallistel, C., Fairhurst, S., & Balsam, P. (2004, 10). The learning curve: Implications of a quantitative analysis. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 13124-31. doi: 10.1073/pnas.0404965101

Gershman, S., & Niv, Y. (2010, 03). Learning latent structure: Carving nature at its joints. *Current opinion in neurobiology*, *20*, 251-6. doi: 10.1016/j.conb.2010.02.008

Grant, D., & Berg, E. (1948, 09). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in weigl-type card-sorting problem. *Journal of experimental psychology*, *38*, 404-11. doi: 10.1037/h0059831

Gross, J., Kujala, J., Salmelin, R., & Schnitzler, A. (2010, 06). Noninvasive functional tomographic connectivity analysis with magnetoencephalography. *MEG: An Introduction to Methods*, 1-32. doi: 10.1093/acprof:oso/9780195307238.003.0009

Gruner, P., & Pittenger, C. (2016, 08). Cognitive inflexibility in obsessive-compulsive disorder. *Neuroscience*, *345*. doi: 10.1016/j.neuroscience.2016.07.030

Guarino, A., Favieri, F., Boncompagni, I., Agostini, F., Cantone, M., & Casagrande, M. (2019, 01). Executive functions in alzheimer disease: A systematic review. *Frontiers in Aging Neuroscience*, *10*. doi: 10.3389/fnagi.2018.00437

Hari, R., & Salmelin, R. (2011, 12). Magnetoencephalography: From squids to neuroscience neuroimage 20th anniversary special edition. *NeuroImage*, *61*, 386-96. doi: 10.1016/j.neuroimage.2011.11.074

Hermann, S. (2022). *Wie soll ich das lernen? - aufmerksamkeit steuert entscheidungs- und lernprozesse bei ratten* (Dissertation). Retrieved from `http://dx.doi.org/10.17169/refubium-36161`

Hirt, E., Devers, E., & McCrea, S. (2008, 03). I want to be creative: Exploring the role of hedonic contingency theory in the positive mood-cognitive flexibility link. *Journal of personality and social psychology*, *94*, 214-30. doi: 10.1037/0022-3514.94.2.94.2.214

Hsu, N., & Jaeggi, S. (2013, 08). The emergence of cognitive control abilities in childhood. *Current topics in behavioral neurosciences*, *16*. doi: 10.1007/7854_2013_241

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis* (Vol. 26). doi: 10.1002/0471221317

Hämäläinen, M. (1991, 02). Anatomical correlates for magnetoencephalography: Integration with magnetic resonance images. *Clinical physics and physiological measurement : an official journal of the Hospital Physicists' Association, Deutsche Gesellschaft für Medizinische Physik and the European Federation of Organisations for Medical Physics*, *12 Suppl A*, 29-32. doi: 10.1088/0143-0815/12/A/006

JC, M., AW, T., Evans, A., PT, F., JL, L., Zilles, K., ... Mazoyer, B. (2001, 09). A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (icbm). *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *356*, 1293-322. doi: 10.1098/rstb.2001.0915

Jones, D., & Graff-Radford, J. (2021, 12). Executive dysfunction and the prefrontal cortex. *CONTINUUM: Lifelong Learning in Neurology*, *27*, 1586-1601. doi: 10.1212/CON.0000000000001009

Kachergis, G. (2018, 05). Word learning: Associations or hypothesis testing? *Current Biology*, *28*, R555-R557. doi: 10.1016/j.cub.2018.02.077

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of artificial intelligence research*, *4*, 237–285.

Katsuki, F., & Constantinidis, C. (2013, 12). Bottom-up and top-down attention. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*, *20*. doi: 10.1177/1073858413514136

Kiefer, M., & Pulvermüller, F. (2011, 04). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex; a journal devoted to the study of the nervous system and behavior*, *48*, 805-25. doi: 10.1016/

j.cortex.2011.04.006

King, J.-R., & Dehaene, S. (2014, 03). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in cognitive sciences*. doi: 10.1016/j.tics.2014.01.002

King, J.-R., Gramfort, A., Schurger, A., Naccache, L., & Dehaene, S. (2014, 01). Two distinct dynamic modes subtend the detection of unexpected sounds. *PloS one*, *9*, e85791. doi: 10.1371/journal.pone.0085791

Kirova, A.-M., Bays, R., & Lagalwar, S. (2015, 10). Working memory and executive function decline across normal aging, mild cognitive impairment, and alzheimer's disease. *BioMed Research International*, *2015*, 1-9. doi: 10.1155/2015/748212

Kopp, B., Lange, F., & Steinke, A. (2019, 08). The reliability of the wisconsin card sorting test in clinical practice. *Assessment*, *28*. doi: 10.1177/1073191119866257

Kounios, J., & Beeman, M. (2014, 01). The cognitive neuroscience of insight. *Annual review of psychology*, *65*, 71-93. doi: 10.1146/annurev-psych-010213-115154

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006, 04). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 3863-8. doi: 10.1073/pnas.0600244103

Kriegeskorte, N., & Kievit, R. (2013, 07). Representational geometry: Integrating cognition, computation, and the brain. *Trends in cognitive sciences*, *17*. doi: 10.1016/j.tics.2013.06.007

Kruczek, A., Basińska, M., & Janicka, M. (2020, 06). Cognitive flexibility and flexibility in coping in nurses - the moderating role of age, seniority and the sense of stress. *International Journal of Occupational Medicine and Environmental Health*, *33*. doi: 10.13075/ijomeh.1896.01567

Langel, J., Hakun, J., Zhu, D., & Ravizza, S. (2014, 06). Functional specialization of the left ventral parietal cortex in working memory. *Frontiers in human neuroscience*, *8*, 440. doi: 10.3389/fnhum.2014.00440

Leber, A., Turk-Browne, N., & Chun, M. (2008, 10). Neural predictors of moment-to-moment fluctuations in cognitive flexibility. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 13592-7. doi: 10.1073/pnas.0805423105

Leong, Y., Radulescu, A., Daniel, R., DeWoskin, V., & Niv, Y. (2017, 01). Dynamic interaction between reinforcement learning and attention in multidimensional environments. *Neuron*, *93*, 451-463. doi: 10.1016/j.neuron.2016.12.040

Li Hegner, Y., Marquetand, J., Elshahabi, A., Klamer, S., Lerche, H., Braun, C., & Focke, N. (2018, 09). Increased functional meg connectivity as a hallmark of mri-negative focal and generalized epilepsy. *Brain Topography*, *31*. doi: 10.1007/s10548-018-0649-4

Limited, C. C. (2006). *Cantabeclipsetm test administration guide, manual version 3.0.0.* Author.

Limited, C. C. (2015). *Cantab cognition handbook, 1st edition.* CambridgeCognitionLimited.

Liu, Z., Braunlich, K., Wehe, H., & Seger, C. (2015, 07). Neural networks supporting switching, hypothesis testing, and rule application. *Neuropsychologia*, *77*. doi: 10 .1016/j.neuropsychologia.2015.07.019

Lowe, C., & Rabbitt, P. (1998, 10). Test/re-test reliability of the cantab and ispocd neuropsychological batteries: Theoretical and practical issues. cambridge neuropsychological test automated battery. international study of post-operative cognitive dysfunction. *Neuropsychologia*, *36*, 915-23. doi: 10.1016/S0028-3932(98)00036-0

Lowery, C., Eswaran, H., Murphy, P., & Preissl, H. (2007, 01). Fetal magnetoencephalography. *Seminars in fetal & neonatal medicine*, *11*, 430-6. doi: 10.1016/ j.siny.2006.09.002

Luciana, M. (2003, 08). Practitioner review: Computerized assessment of neuropsychological function in children: clinical and research applications of the cambridge neuropsychological testing automated battery (cantab). *Journal of child psychology and psychiatry, and allied disciplines*, *44*, 649-63. doi: 10.1111/1469-7610.00152

Ludmer, R., Dudai, Y., & Rubin, N. (2011, 03). Uncovering camouflage: Amygdala activation predicts long-term memory of induced perceptual insight. *Neuron*, *69*, 1002-14. doi: 10.1016/j.neuron.2011.02.013

Magnusson, K., & Brim, B. (2014, 12). The aging brain. *Reference Module in Biomedical Sciences*. doi: 10.1016/B978-0-12-801238-3.00158-6

Majerus, S., Peters, F., Bouffier, M., Cowan, N., & Phillips, C. (2017, 10). The dorsal attention network reflects both encoding load and top–down control during working memory. *Journal of Cognitive Neuroscience*, *30*, 1-16. doi: 10.1162/jocn_a_01195

Mandal, P., Banerjee, A., Tripathi, M., & Sharma, A. (2018, 07). A comprehensive report on magnetoencephalography (meg) studies for brain functionality in healthy aging and alzheimer's disease. *Frontiers in Computational Neuroscience*, *12*. doi: 10.3389/ fncom.2018.00060

Manes, F., Sahakian, B., Clark, L., Rogers, R., Antoun, N., Aitken, M., & Robbins, T. (2002, 04). Decision-making processes following damage to the prefrontal cortex. *Brain : a journal of neurology*, *125*, 624-39. doi: 10.1093/brain/awf049

Manoach, D. (2017, 12). Cognitive deficits in schizophrenia.. doi: 10.1016/B978-0-12-809324 -5.01921-0

Maris, E., & Oostenveld, R. (2007, 09). Nonparametric statistical testing of eeg- and meg-data. *Journal of neuroscience methods*, *164*, 177-90. doi: 10.1016/j.jneumeth.2007 .03.024

Mehta, M., Owen, A., Mavaddat, N., Pickard, J., & Robbins, T. (2000, 03). Methylphenidate enhances working memory by modulating discrete frontal and parietal lobe regions in the human brain. *J Neurosci*, *20*, 1-6. doi: 10.1523/JNEUROSCI.20-06-j0004.2000

Menon, V., & Uddin, L. (2010, 06). Saliency, switching, attention and control: A network model of insula function. *Brain structure & function*, *214*, 655-67. doi: 10.1007/ s00429-010-0262-0

Milner, B. (1963, 01). Effects of different brain lesions on card sorting. *Archives of Neurology*, *9*, 101-110.

Miyake, A., Friedman, N., Emerson, M., Witzki, A., Howerter, A., & Wager, T. (2000, 09). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive psychology*, *41*, 49-100. doi: 10.1006/cogp.1999.0734

MNE-Developers. (2012, Jan). *Signal-space separation (sss) and maxwell filtering#.* MNE developer. Retrieved from `https://mne.tools/stable/auto_tutorials/preprocessing/60_maxwell_filtering_sss.html#footcite-taulukajola2005`

Murakami, S., & Okada, Y. (2006, 10). Contributions of principal neocortical neurons to magnetoencephalography and electroencephalography signals. *The Journal of physiology*, *575*, 925-36. doi: 10.1113/jphysiol.2006.105379

Niendam, T., Laird, A., Ray, K., Dean, Y., Glahn, D., & Carter, C. (2012, 01). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, affective & behavioral neuroscience*, *12*, 241-68. doi: 10.3758/s13415-011-0083-5

Niv, Y. (2019, 10). Learning task-state representations. *Nature Neuroscience*, *22*, 1544-1553. doi: 10.1038/s41593-019-0470-8

Niv, Y., Daniel, R., Geana, A., Gershman, S., Leong, Y., Radulescu, A., & Wilson, R. (2015, 05). Reinforcement learning in multidimensional environments relies on attention mechanisms. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *35*, 8145-57. doi: 10.1523/JNEUROSCI.2978-14.2015

Nugent, A., Ballard, E., Gilbert, J., Tewarie, P., Brookes, M., & Jr, C. (2020, 08). Multilayer meg functional connectivity as a potential marker for suicidal thoughts in major depressive disorder. *NeuroImage: Clinical*, *28*, 102378. doi: 10.1016/j.nicl.2020.102378

Oh, A., Vidal, J., Taylor, M., & Pang, E. (2014, 03). Neuromagnetic correlates of intra- and extra-dimensional set-shifting. *Brain and cognition*, *86C*, 90-97. doi: 10.1016/j.bandc.2014.02.006

Ohata, R., Ogawa, K., & Imamizu, H. (2016, 06). Single-trial prediction of reaction time variability from meg brain activity. *Scientific Reports*, *6*, 27416. doi: 10.1038/srep27416

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011, 01). Fieldtrip: Open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Computational intelligence and neuroscience*, *2011*, 156869. doi: 10.1155/2011/156869

Oostenveld, R., Gonaz, G., Sassenhagen, J., & Schoffelen, J.-M. (2018, Oct). *How not to interpret results from a cluster-based permutation test.* Retrieved from `https://www.fieldtriptoolbox.org/faq/how_not_to_interpret_results_from_a_cluster-based_permutation_test/`

Oostenveld, R., Spaak, E., Delorme, A., & Arana, S. (2018, October). *Cluster-based permu-*

tation tests on time-frequency data. Retrieved from https://www.fieldtriptoolbox.org/tutorial/cluster_permutation_freq/#introduction

Owen, A., James, M., Leigh, P., Summers, B., Marsden, C., Quinn, N., ... Robbins, T. (1993, 01). Fronto-striatal cognitive deficits at different stage of parkinson's disease. *Brain : a journal of neurology*, *115 (Pt 6)*, 1727-51. doi: 10.1093/brain/115.6.1727

Owen, A. M., Downes, J. J., Sahakian, B. J., Polkey, C. E., & Robbins, T. W. (1990). Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia*, *28*(10), 1021-1034. Retrieved from https://www.sciencedirect.com/science/article/pii/002839329090137D doi: https://doi.org/10.1016/0028-3932(90)90137-D

Owen, A. M., Evans, A. C., & Petrides, m. (1996, 01). Evidence for a Two-Stage Model of Spatial Working Memory Processing within the Lateral Frontal Cortex: A Positron Emission Tomography Study. *Cerebral Cortex*, *6*(1), 31-38. Retrieved from https://doi.org/10.1093/cercor/6.1.31 doi: 10.1093/cercor/6.1.31

O'Reilly, C., Lewis, J. D., & Elsabbagh, M. (2017, 05). Is functional brain connectivity atypical in autism? a systematic review of eeg and meg studies. *PLOS ONE*, *12*(5), 1-28. Retrieved from https://doi.org/10.1371/journal.pone.0175870 doi: 10.1371/journal.pone.0175870

Pang, E., & III, O. (2016, 06). From structure to circuits: The contribution of meg connectivity studies to functional neurosurgery. *Frontiers in Neuroanatomy*, *10*. doi: 10.3389/fnana.2016.00067

Park, J., & Moghaddam, B. (2016, 06). Impact of anxiety on prefrontal cortex encoding of cognitive flexibility. *Neuroscience*, *345*. doi: 10.1016/j.neuroscience.2016.06.013

Parr, T., & Friston, K. J. (2019). Attention or salience? *Current Opinion in Psychology*, *29*, 1-5. Retrieved from https://www.sciencedirect.com/science/article/pii/S2352250X18301593 (Attention & Perception) doi: https://doi.org/10.1016/j.copsyc.2018.10.006

Periáñez, J., Maestú, F., Barceló, F., Amo, C., & Alonso, T. (2004, 03). Spatiotemporal brain dynamics during preparatory set shifting: Meg evidence. *NeuroImage*, *21*, 687-95. doi: 10.1016/j.neuroimage.2003.10.008

Perpiñá, C., Segura, M., & Sanchez-Reales, S. (2017, 09). Cognitive flexibility and decision-making in eating disorders and obesity. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity*, *22*. doi: 10.1007/s40519-016-0331-3

Powell, N., & Redish, A. (2016, 09). Representational changes of latent strategies in rat medial prefrontal cortex precede changes in behaviour. *Nature communications*, *7*, 12830. doi: 10.1038/ncomms12830

Preiss, D. (2022, 09). Metacognition, mind wandering, and cognitive flexibility: Understanding creativity. *Journal of Intelligence*, *10*, 69. doi: 10.3390/jintelligence10030069

Preissl, H., Lowery, C., & Eswaran, H. (2004, 12). Fetal magnetoencephalography: Current progress and trends. *Experimental neurology*, *190 Suppl 1*, S28-36. doi: 10.1016/

j.expneurol.2004.06.016

Proudfoot, M., Woolrich, M., Nobre, A., & Turner, M. (2014, 03). Magnetoencephalography. *Practical neurology*, *14*. doi: 10.1136/practneurol-2013-000768

Pruessner, L., Barnow, S., Holt, D., Joormann, J., & Schulze, K. (2020, 03). A cognitive control framework for understanding emotion regulation flexibility. *Emotion*, *20*, 21–29. doi: 10.1037/emo0000658

Radulescu, A., Shin, Y., & Niv, Y. (2021, 07). Human representation learning. *Annual Review of Neuroscience*, *44*. doi: 10.1146/annurev-neuro-092920-120559

Ram, D., Chandran, S., Sadar, A., & Gowdappa, B. (2019, 07). Correlation of cognitive resilience, cognitive flexibility and impulsivity in attempted suicide. *Indian Journal of Psychological Medicine*, *41*. doi: 10.4103/IJPSYM.IJPSYM_189_18

Ravizza, S., Moua, K., Long, D., & Carter, C. (2009, 10). The impact of context processing deficits on task-switching performance in schizophrenia. *Schizophrenia research*, *116*, 274-9. doi: 10.1016/j.schres.2009.08.010

Ress, D., Backus, B., & Heeger, D. (2000, 10). Activity in primary visual cortex predicts performance in a visual detection task. *Nature neuroscience*, *3*, 940-5. doi: 10.1038/78856

Roberts, T., Kuschner, E., & Edgar, J. C. (2021, 12). Biomarkers for autism spectrum disorder: opportunities for magnetoencephalography (meg). *Journal of Neurodevelopmental Disorders*, *13*. doi: 10.1186/s11689-021-09385-y

Rolls, E., Huang, C.-C., Feng, J., & Marc, J. (2019, 09). Automated anatomical labelling atlas 3. *NeuroImage*, *206*, 116189. doi: 10.1016/j.neuroimage.2019.116189

Rolls, E., Marc, J., & Tzourio-Mazoyer, N. (2015, 08). Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *NeuroImage*, *122*. doi: 10.1016/j.neuroimage.2015.07.075

Rosenberg, M., Zhang, T., Perona, P., & Meister, M. (2021, jul). Mice in a labyrinth show rapid learning, sudden insight, and efficient exploration. *eLife*, *10*, e66175. Retrieved from `https://doi.org/10.7554/eLife.66175` doi: 10.7554/eLife.66175

Sahakian, B., Morris, R., Evenden, J., Heald, A., Levy, R., Philpot, M., & Robbins, T. (1988, 07). A comparative study of visuospatial memory and learning in alzheimer-type dementia and parkinson's disease. *Brain : a journal of neurology*, *111 (Pt 3)*, 695-718. doi: 10.1093/brain/111.3.695

Sahakian, B., & Owen, A. (1992, 08). Computerized assessment in neuropsychiatry using cantab: Discussion paper. *Journal of the Royal Society of Medicine*, *85*, 399-402.

Sandberg, M. (2011, 01). Cambridge neuropsychological testing automated battery. In (p. 480-482). doi: 10.1007/978-0-387-79948-3_169

Sassenhagen, J., & Draschkow, D. (2019, 01). Cluster-based permutation tests of meg/eeg data do not establish significance of effect latency or location. *Psychophysiology*, *56*. doi: 10.1111/psyp.13335

Schall, J. (2003, 05). Schall, j.d. neural correlates of decision processes: neural and mental

chronometry. curr. opin. neurobiol. 13, 182-186. *Current opinion in neurobiology*, *13*, 182-6. doi: 10.1016/S0959-4388(03)00039-4

Seeley, W., Menon, V., Schatzberg, A., Keller, J., Glover, G., Kenna, H., ... Greicius, M. (2007, 03). Dissociable intrinsic connectivity networks for salience processing and executive control. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *27*, 2349-56. doi: 10.1523/JNEUROSCI.5587-06.2007

Siegel, M., Donner, T., & Engel, A. (2012, 02). Spectral fingerprints of large scale neuronal interactions. *Nature reviews. Neuroscience*, *13*, 121-34. doi: 10.1038/nrn3137

Singh, S. (2014, 03). Magnetoencephalography: Basic principles. *Annals of Indian Academy of Neurology*, *17*, S107-S112. doi: 10.4103/0972-2327.128676

Song, M., Baah, P., Cai, M., & Niv, Y. (2023, 01). Humans combine value learning and hypothesis testing strategically in multi-dimensional probabilistic reward learning. *preprint*. doi: 10.31234/osf.io/ux5pr

Stange, J., Alloy, L., & Fresco, D. (2017, 06). Inflexibility as a vulnerability to depression: A systematic qualitative review. *Clinical Psychology Science and Practice*, *24*. doi: 10.1111/cpsp.12201

Stokes, M., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2013, 04). Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, *78*, 1-12. doi: 10.1016/j.neuron.2013.01.039

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning - an introduction*. MIT Press. Retrieved from `https://www.worldcat.org/oclc/37293240`

Tan, V., Dockstader, C., Moxon-Emre, I., Mendlowitz, S., Schacter, R., Colasanto, M., ... Ameis, S. (2022, 12). Preliminary observations of resting-state magnetoencephalography in nonmedicated children with obsessive-compulsive disorder. *Journal of child and adolescent psychopharmacology*, *32*, 522-532. doi: 10.1089/cap.2022.0036

Taulu, S., & Kajola, M. (2005, 07). Presentation of electromagnetic multichannel data: The signal space separation method. *Journal of Applied Physics*, *97*, 124905 - 124905. doi: 10.1063/1.1935742

Taulu, S., & Simola, J. (2006, 05). Spatiotemporal signal space separation method for rejecting nearby interference in meg measurements. *Physics in medicine and biology*, *51*, 1759-68. doi: 10.1088/0031-9155/51/7/008

Thomason, M., Race, E., Burrows, B., Whitfield-Gabrieli, S., Glover, G., & Gabrieli, J. (2008, 06). Development of spatial and verbal working memory capacity in the human brain. *Journal of cognitive neuroscience*, *21*, 316-32. doi: 10.1162/jocn.2008.21028

Toutounji, H., & Durstewitz, D. (2018). Detecting multiple change points using adaptive regression splines with application to neural recordings. *Frontiers in Neuroinformatics*, *12*. Retrieved from `https://www.frontiersin.org/articles/10.3389/fninf.2018.00067` doi: 10.3389/fninf.2018.00067

Treder, M., Oostenveld, R., Schoffelen, J.-M., & Janke, A. (2019, Mar). *Classification of event-related meg data using mvpa-light.* Retrieved from `https://www`

.fieldtriptoolbox.org/tutorial/mvpa_light/

Treder, M. S. (2020, 06). Mvpa-light: A classification and regression toolbox for multi-dimensional data. *Frontiers in Neuroscience*, *14*, 289. doi: 10.3389/fnins.2020.00289

Tukey, J. (1973, 01). A problem of multiple comparisons. *Biometrics*, *9*.

Uddin, L. (2014, 11). Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience*. doi: 10.1038/nrn3857

Uddin, L. (2021, 02). Cognitive and behavioural flexibility: neural mechanisms and clinical considerations. *Nature Reviews Neuroscience*, *22*. doi: 10.1038/s41583-021-00428-w

Uddin, L., Yeo, B. T., & Spreng, R. N. (2019, 11). Towards a universal taxonomy of macro-scale functional human brain networks. *Brain Topography*, *32*. doi: 10.1007/s10548-019-00744-6

Van Mierlo, P., Höller, Y., Focke, N., & Vulliemoz, S. (2019, 07). Network perspectives on epilepsy using eeg/meg source connectivity. *Frontiers in Neurology*, *10*. doi: 10.3389/fneur.2019.00721

Verhaeghen, P., & Cerella, J. (2002, 12). Aging, executive control, and attention: A review of meta-analyses. *Neuroscience and biobehavioral reviews*, *26*, 849-57. doi: 10.1016/S0149-7634(02)00071-4

Vossel, S., Geng, J., & Fink, G. (2013, 07). Dorsal and ventral attention systems: Distinct neural circuits but collaborative roles. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*, *20*. doi: 10.1177/1073858413494269

Vossel, S., Geng, J., & Friston, K. (2014, 07). Attention, predictions and expectations, and their violation: attentional control in the human brain. *Frontiers in human neuroscience*, *8*, 490. doi: 10.3389/fnhum.2014.00490

Vossel, S., Weidner, R., Moos, K., & Fink, G. (2016, 01). Individual attentional selection capacities are reflected in interhemispheric connectivity of the parietal cortex. *NeuroImage*, *129*. doi: 10.1016/j.neuroimage.2016.01.054

Wild, K., & Musser, E. (2014, 04). The cambridge neuropsychological test automated battery in the assessment of executive functioning. In (p. 171-190). doi: 10.1007/978-1-4614-8106-5_11

Williamson, S., & Kaufman, L. (1989, 09). Advances in neuromagnetic instrumentation and studies of spontaneous brain activity. *Brain Topography*, *2*, 129-139. doi: 10.1007/BF01128850

Wilson, C., Nusbaum, A., Whitney, P., & Hinson, J. (2018, 09). Trait anxiety impairs cognitive flexibility when overcoming a task acquired response and a preexisting bias. *PLoS ONE*, *13*, e0204694. doi: 10.1371/journal.pone.0204694

Wilson, R., & Niv, Y. (2011, 01). Inferring relevance in a changing world. *Frontiers in human neuroscience*, *5*, 189. doi: 10.3389/fnhum.2011.00189

Wilson, T., Heinrichs-Graham, E., Proskovec, A., & McDermott, T. (2016, 01). Neuroimaging with magnetoencephalography: A dynamic view of brain pathophysiology. *Translational Research*, *175*. doi: 10.1016/j.trsl.2016.01.007

Wilson, T., Lew, B., Spooner, R., Rezich, M., & Wiesman, A. (2019, 01). Aberrant brain dynamics in neurohiv: Evidence from magnetoencephalographic (meg) imaging. In (Vol. 165). doi: 10.1016/bs.pmbts.2019.04.008

Womelsdorf, T., & Everling, S. (2015, 11). Long-range attention networks: Circuit motifs underlying endogenously controlled stimulus selection. *Trends in neurosciences*, *38*, 682-700. doi: 10.1016/j.tins.2015.08.009

Wöhrle, J. (2020, 09). Magnetoenzephalografie (meg). In (p. 167-167). doi: 10.1007/978-3-662-60676-6_10

Xia, M., Wang, J., & He, Y. (2013, 07). Brainnet viewer: A network visualization tool for human brain connectomics. *PLoS ONE*, *8*, e68910. doi: 10.1371/journal.pone.0068910

Zhou, Z., Li, B., Jiang, J., Li, H., Cao, L., Zhang, S., ... Gong, Q. (2022, 11). Abnormal resting-state functional connectivity of the insula in medication-free patients with obsessive-compulsive disorder. *BMC Psychiatry*, *22*. doi: 10.1186/s12888-022-04341-z

Curriculum Vitae

Personal Information

Name and Surname:	Hélène Laura Valentina Zingone
Year of Birth:	1996
Place of birth:	Courbevoie

Academic Employment

2022	Third and final state examination
2021	Second state examination
2019	Medical graduate student, research group "Behavioral physiology in Psychiatry", Central Institute for Mental Health, Mannheim
2018 - 2019	Research assistant, section "Health equity studies & migration", Heidelberg University, field work at the State Office for Initial Refugee Registration "Patrick Henry Village", Heidelberg
2018	Semester abroad in Florence, Italy
2016	First state examination
2014 - 2022	Human Medicine Studies, Universität Heidelberg, Medizinische Fakultät Heidelberg

Education

2014	High School Graduation
2007 - 2014	Internationale Deutsche Schule Brüssel, Brussels
2005 - 2007	Ecole Française Internationale de Philadelphie, Philadelphia
2002 - 2005	Ecole Maternelle Primaire du Luxembourg, Luxemburg

Acknowledgements

This thesis was planned, performed and completed at the Central Institute for Mental Health in Mannheim, Germany. My sincere gratitude goes to the following people for their help and support:

To Prof. Dr. Andreas Meyer−Lindenberg, for accepting my thesis and providing support and ressources within the Central Institute of Mental Health.

To my supervisor and advisor, Dr. Florian Bähner during the last years. Without his assistance and dedicated involvement in every step, this project would not have been possible. I would like to thank you very much for your support and understanding over these past four years.

To my advisor Dr. Tzvetan Popov, for guidance, supervision and introduction on *MEG*-related questions.

To Dr. Konstantina Atanasova and Dr. med. vet. Selina Hermann, for introducing and teaching me methods employed in the project.

To my friends and family for providing mental support and advice.

To all the participants, who took time and energy and tried their best to perform well in the employed task.

And lastly, to the large community of developers and users of multiple and immense online ressources, enabling me to easily work with programs such as Matlab, Toolboxes such as BrainNet Viewer and Fieldtrip, Latex, and Inkscape.