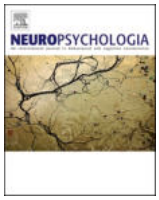
Neuropsychologia 188 (2023) 108615

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia



Cognitive processing stages in mental rotation – How can cognitive modelling inform HsMM-EEG models?

Linda Heimisch [a,*,1,2], Kai Preuss [a,1], Nele Russwinkel [b]

[a] Technische Universität Berlin, Department of Psychology and Ergonomics, Marchstraße 23, 10587, Berlin, Germany
[b] Universität zu Lübeck, Institut für Informationssysteme, Ratzeburger Allee 160, 23562, Lübeck, Germany

ARTICLE INFO

Keywords:
Electroencephalography
Cognitive modelling
HsMM
ACT-R
Mental rotation

ABSTRACT

The aspiration for insight into human cognitive processing has traditionally driven research in cognitive science. With methods such as the Hidden semi-Markov Model-Electroencephalography (HsMM-EEG) method, new approaches have been developed that help to understand the temporal structure of cognition by identifying temporally discrete processing stages. However, it remains challenging to assign concrete functional contributions by specific processing stages to the overall cognitive process. In this paper, we address this challenge by linking HsMM-EEG[3] with cognitive modelling, with the aim of further validating the HsMM-EEG[3] method and demonstrating the potential of cognitive models to facilitate functional interpretation of processing stages. For this purpose, we applied HsMM-EEG[3] to data from a mental rotation task and developed an ACT-R cognitive model that is able to closely replicate human performance in this task. Applying HsMM-EEG[3] to the mental rotation experiment data revealed a strong likelihood for 6 distinct stages of cognitive processing during trials, with an additional stage for non-rotated conditions. The cognitive model predicted intra-trial mental activity patterns that project well onto the processing stages, while explaining the additional stage as a marker of non-spatial shortcut use. Thereby, this combined methodology provided substantially more information than either method by itself and suggests conclusions for cognitive processing in general.

1. Introduction

Gaining insight into human information processing is one of the prime endeavours of cognitive science. To this end, electroencephalography (EEG) is a notably central and fruitful neuropsychological method of data acquisition for modern research on cognition. Especially in complex cognition research (Funke, 2010), EEG monitoring has turned out to be an indispensable tool. Despite this, there are limits to its explanatory power: in most cases, experimental effects are derived from heavily filtering EEG signals by location, frequency spectra, event relation, or combinations of these factors. Interpretations of entire activity patterns, i.e. the cortical distribution of all active signals during a given time span, are seldom considered.

Cognitive modelling has been similarly useful in helping to gain a deeper understanding of how the interplay of cognitive processes results in particular patterns of behaviour. Simulating task demands and human task solving within cognitive constraints, cognitive models allow for the verification of theories about cognitive processing. They enable understanding of intra-trial processes, predicting distinct cognitive activity during specific time spans. Cognitive architectures like ACT-R offer a proven, reliable framework for the generation of valid models by differentiating cognitive processes into separate modules. Cognitive models can be evaluated by their predictive power of reaction times (RTs), learning curves or individual behaviour (e.g. strategies, errors etc.) during human solving of a task.

Yet, a direct comparison of the proposed cognitive mechanisms with neural activity remains difficult, even more so on a fine-grained temporal scale: model output and EEG signals are not directly comparable to each other, as high EEG activity rarely equals a heightened effort in a clearly defined cognitive process. This in turn can be alleviated by the application of Markov models: by further post-processing of EEG signals, they can be put in a context of time series which are generated by

* Corresponding author.
E-mail addresses: heimisch@tu-berlin.de (L. Heimisch), preuss@tu-berlin.de (K. Preuss), nele.russwinkel@tu-berlin.de (N. Russwinkel).
[1] Authors contributed equally.
[2] Postal address: Office MAR 3–2, Marchstraße 23, 10587 Berlin.
[3] Hidden semi-Markov Model-Electroencephalography.

https://doi.org/10.1016/j.neuropsychologia.2023.108615
Received 28 November 2022; Received in revised form 2 June 2023; Accepted 6 June 2023
Available online 7 July 2023

patterns of common neuronal activity in a particular sequence relative to cognitive events. These contexts thus represent manifestations of cognitive processing stages. While such newly created processing stages offer a way to aggregate activity patterns over time, they elude easy interpretation regarding their semantic content, that is, their contribution to the task solving process as a whole.

Associating cognitive process assumptions with Markov models provides intriguing possibilities to further understand the inner workings of higher order cognition. An additional level of interpretation thus becomes available, providing additional information not apparent from each respective method applied by itself. This paper demonstrates the advantages of this methodology by way of a mental rotation study.

In the remainder of this introduction we will first introduce the concept of processing stages in the context of cognition research, followed by the related issue of interpretability. Then, we will give an introduction to the method of cognitive modelling. After that, we will outline mental rotation as the exemplary subject that we will use for our demonstration. Finally, the section closes with the hypotheses underlying this paper.

1.1. Cognitive processing stages

Processing stages are defined as successive processes between stimulus and response that together make up one trial, where one processing stage begins only when the preceding one has ended (Sternberg, 1969). The research interest in processing stages in cognitive tasks can be dated back to the 19th century where Donders (1868) argued that subtracting the RTs of two tasks that hypothetically share all but one processing stage would yield the duration of that processing stage. Building on this RT-based approach, Sternberg (1969) proposed the so-called additive factor method that assumes overall RTs to vary depending on experimental manipulations. Separate processing stages can, as per this reasoning, be identified based on additive effects of certain manipulations on overall RT.

Behavioural measures are however insufficient for the identification of the temporal structure in a cognitive process. The RT itself cannot be decomposed into stages, that is, the absolute durations of the discovered stages cannot be determined and their order remains unclear (Sternberg, 1969). To alleviate these limitations, stage-identifying methods based on neuropsychological data are needed. As an answer, the MVPA-HMM method, which is based on functional magnetic resonance imaging (fMRI), was introduced (Anderson and Fincham, 2014a; Anderson et al., 2012). This method combines hidden Markov models (HMMs) with multivariate pattern analysis (MVPA) to identify the number and duration of processing stages in a mathematical problem solving task based on fMRI data. A HMM simulates a system that is at any given time in one of a set of distinct states, between which it transitions at certain times. 'Hidden' refers to the fact that the stochastic process underlying the Markov model is not observable (Rabiner, 1990).

However, fMRI-based methods are limited by their poor temporal resolution that only allows for the identification of multi-second stages in tasks lasting on the order of 10+ seconds (Anderson et al., 2016). EEG, on the other hand, has a millisecond resolution. Therefore, EEG-based stage-identifying methods allow for the identification of much briefer stages, and hence for an investigation of elementary cognitive processes (Anderson et al., 2016; Borst and Anderson, 2015). Borst et al. (2013) proposed to identify processing stages in trial-wise EEG data by means of machine-learning techniques that were initially outlined by Sudre et al. (2012). They applied this method to an associative recognition memory task and identified three main processing stages which they labelled as perceptual encoding stage, retrieval stage, and decision stage, respectively. Anderson et al. (2016) proposed an EEG-based HsMM-MVPA method that combines hidden semi-Markov models (HsMMs), where "semi" refers to a type of HMM characterised by variable state durations (Rabiner, 1990), with MVPA to identify processing stages based on brief sinusoidal peaks, so-called "bumps", that are added to the ongoing EEG signal and mark the onsets of distinct processing stages. Just like Borst et al. (2013), they applied this method to the study of associative recognition and found evidence for six processing stages in this task. Additionally, they applied their method to a data set on the Sternberg working memory task and identified five processing stages. The EEG-based HsMM-MVPA method has since been applied to a variety of fields in neuropsychology, e.g., associative recognition (Anderson et al., 2018; Portoles et al., 2018; Zhang et al., 2017), visual working memory (Zhang et al., 2018a), arithmetic retrieval (Anderson et al., 2018), a combined memory and arithmetic task (Zhang et al., 2018b), or stimulus categorisation (Berberyan et al., 2020; Walsh et al., 2017).

With HsMM-EEG[3], Borst and Anderson (2015) proposed another stage-identifying method. In their approach, just like in the HsMM-MVPA method by Anderson et al. (2016), a processing stage corresponds to a state of the HsMM. In contrast to HsMM-MVPA, it is not defined by an initial "bump", but by a period of constant neuronal activity, called neuronal signature. Borst and Anderson (2015) define the HsMM-EEG method in four basic steps: **(1)** HsMMs with different numbers of states are fitted to a set of trial-wise EEG data, where only HsMMs with linear structures are considered (i.e, state 1 always transitions to state 2, state 2 to state 3, etc.). **(2)** The number of HsMM states, representing the processing stages that best describe the task solving process is identified by comparing the likelihoods of the fitted HsMMs using leave-one-out-cross-validation (LOOCV). **(3)** The resulting HsMM is inspected for number, order, durations, and neuronal signature of the stages. **(4)** The functions of the identified processing stages are deduced. The authors note that while being similar to microstate analysis, HsMM-EEG also differs from this established approach in that neuronal signatures are assumed to be fixed, numbers of states are to be determined, and all EEG data is taken into account during the parameter estimation procedure, with the latter difference being advantageous for determining the optimal number of states (Borst and Anderson, 2015).

Anderson et al. (2016) applied their HsMM-MVPA method to the same EEG data as Borst and Anderson (2015) and confirmed the results regarding number and duration of processing stages, as well as influence of experimental conditions on processing stage duration. This suggests that HsMM-EEG provides a robust approach to processing stage identification from EEG data (Anderson et al., 2016).

Through the identification of clearly differentiated stages, this method allows to temporally structure EEG data segments spanning complete experimental trials. This structure then opens the possibility for examining cognitive processes in an informative way: How many processing stages does a cognitive process consist of, how is the total duration of a cognitive process distributed among these processing stages, which processing stages vary in duration between different experimental conditions, and to what extent? Other than in ERP analyses, which are either stimulus-locked or response-locked and only look at a fixed period within an EEG data segment, the HsMM-EEG method allows to use the entire EEG signal from the beginning of an experimental trial until a subject's response (Borst and Anderson, 2015), allowing for substantial insights into information processing during a task solving process without presupposing strong hypotheses about what mental step is happening at which point in time. Further, as opposed to established methods of EEG data analysis, HsMM-EEG enables the identification of all stages that compose a concrete cognitive process and thereby offers to examine qualitatively different stages of information processing within one analysis. These possibilities make the HsMM-EEG method a promising tool for various research questions in the fields of neuropsychology and cognition research that aim at the timing and sequence of human information processing.

Thinking of cognitive processes as sequences of clear-cut stages raises important conceptual questions, however. Processing stages are by definition distinct, while in human cognition many processes happen in parallel. Still, we argue that through the application of likelihood-based methods, distinct phases of neuronal activity patterns can be detected that mark clearly definable stages within a cognitive process. These

processing stages are in the first place a statistical phenomenon. Nevertheless, by combining the HsMM-EEG method with cognitive modelling data we aim at finding real world manifestations of these stages.

1.2. Interpreting cognitive processing stages

A challenge of stage-identifying methods lies in identifying each stage's functional contribution to the cognitive process, since the ground truth is inaccessible, that is, the cognitive process cannot be directly observed (Anderson and Fincham, 2014a; Berberyan et al., 2020). Semantic interpretation of cognitive processing stage content can be approached by examining the results of a stage-identifying analysis for differences in stage durations between different experimental conditions. If the duration of a certain stage varies with conditions, that stage is likely to contain information processing that differs between the relevant conditions (Berberyan et al., 2020; Borst and Anderson, 2015; Zhang et al., 2018a). Also, functional interpretation of stage content can be guided by pre-existing knowledge about the task-solving process under investigation (Borst and Anderson, 2015; Zhang et al., 2018b).

These approaches are, however, limited for several reasons. When the results of a stage-identifying analysis are compared for different experimental conditions, the most significant semantic interpretations will be found for those processing stages that effectively differ in duration between conditions. Moreover, this calls for specific requirements of an experimental design for it to be meaningfully analysable with a stage-identifying method, since it presupposes that conditions can be outlined beforehand in terms of the cognitive demands they make on subjects to solve the respective task. Likewise, relying on previous findings about the cognitive process under investigation will produce the most meaningful results for tasks already backed by substantial knowledge of the underlying cognitive processes. We thus argue that the applicability of stage-identifying methods is as of yet limited. Further sources of information about semantic stage content are needed that enrich these approaches with easily interpretable information.

As a solution, we propose to use cognitive modelling. As it allows for the simulation of mental activity during task solving by subjects, it offers a unique way to formalise and test theories on intra-trial human reasoning processes. We developed an ACT-R cognitive model that simulates the same task that subjects performed in the underlying EEG study, and that implements a well-validated theory about how specific cognitive processes evolve over time. Hence, the ACT-R model constitutes a theory-driven source of semantic information about the cognitive process under investigation. This semantic information shall then allow inference of the content of processing stages that were identified by the data-driven HsMM-EEG method.

1.3. Cognitive modelling

Cognitive architectures are frameworks that enable simulations of task behaviour based on psychological models and assumptions about cognitive processes. They consist of sets of rules and constraints that are derived from the current state of research in cognitive science and allow researchers to implement psychological or process models into a standardised form (Newell, 1990). The resulting cognitive models simulate human behaviour during task solving and generate predictions about behavioural markers such as RTs, error rates or strategy decisions. When creating these models for a specific task, this task is simulated within the cognitive architecture as well. A set of adjustable parameters allows for a better fit to data from human experiments in regards to timing or task errors, while still adhering to cognitive constraints. Especially useful for exploring human biases, common errors or complicated tasks, cognitive architectures allow for insights beyond psychological or mathematical models. In contrast to a psychological model, cognitive models can be used to generate predictions for a myriad of different task variants and contexts, while simultaneously offering explainability and transparency not possible with mathematical models. This in turn makes them uniquely useful for testing hypotheses on human task solving, e.g. effects of experimental conditions, strategy choices or individual strengths and weaknesses.

The most common demands and constraints of cognitive modelling have been unified and formalised as Common Model of Cognition (CMC), originally called Standard Model of the Mind (Laird et al., 2017). The CMC aims to set a universally valid formalisation of cognition, i.e. a standardised abstraction for its implementation in an arbitrary computational architecture. It proposes guidelines on (processing) structure, memory and information content, learning as well as somatosensory interfaces. Currently, the most complete implementation of requirements posed of a unified standard model is the modular cognitive architecture ACT-R. It encodes cognitive models in the form of a declarative knowledge base on one hand, and procedural production rules on the other (Anderson and Lebiere, 2014). Information in ACT-R is symbolic, meaning objects, concepts etc. are referred to by symbols (as opposed to visual or mathematical references). Knowledge is represented in the form of *chunks*, basic units of information that are known (world knowledge) or learned (task-specific knowledge). Production rules are processing steps that manage the exchange of information between ACT-R's modules, structures that implement specific forms of cognitive processing, for instance visual, motor or imaginal cognition. Thus, these modules instantiate cognition on a conceptual level between psychological and neurobiological measures, akin to the algorithmic level in Marr's three Levels of Analysis (computational, algorithmic, implementational; Marr, 1982). As these modules can run independently of and in parallel to each other, the activity of each module can be used as an additional predictive measure that can be generated by a cognitive model. While modules are not exact analogues to brain areas, they are close enough both functionally and structurally to exhibit common features, and as such efforts have been made to compare module activity to brain activity in a meaningful manner. The following literature section lists a selection of papers about using ACT-R to predict cognitive processes.

ACT-R is able to generate activity predictions for all of its modules during a task, facilitating comparisons between this module activity and other, external measures. For instance, activity produced by the visual module can be related to human data on visual perception, e.g. from eye tracking experiments (Byrne et al., 1999). ACT-R was used to simulate cognitive workload and effects of multitasking of subjects during an air traffic controller task (Lebiere et al., 2001), measuring workload by the proportion of time the model spent on critical tasks on one hand and subject self-reports on the other. The authors reported a high accuracy of the cognitive model regarding human performance, producing a good fit in terms of RTs and errors. Jo et al. (2012) generated a workload measure from the aggregated activity of selected modules during memorisation, visuomanual and menu selection tasks. The predicted workload was then compared to NASA-TLX self-reports from subjects in corresponding studies, resulting in a high correlation between both. In a military unit dispatch task, Stevens et al. (2019) introduce Cognitive Metrics Profiling, a custom set of workload predictions based on the performance of an ACT-R model, and suggest it could be used to predict and eventually replace physiological measurements like heart rate, EEG, and eye movements.

Further research outlined methods to relate ACT-R output to brain imaging data, particularly fMRI. fMRI provides high spatial resolution of activity for the whole brain, including medial structures such as basal ganglia or hippocampi, and thus lends itself easily to comparisons with cognitive models. Anderson et al. (2003) showed that module activity produced by a model solving an algebraic task was able to predict BOLD[4] responses present in fMRI data of subjects solving the same task. The

[4] Blood Oxygen Level Dependency; a slow but reliable measure for locally increased oxygen consumption and therefore activity in the brain.

same method was applied by Stocco and Anderson (2008) to differentiate between functionally different structures, showing domain-specific processing of algebraic tasks in part explained by a cognitive model. In 2008, Anderson and colleagues associated activity generated by the four ACT-R modules for declarative, imaginal, goal, and procedural processing during a complex, high-order task with fMRI activity in pre-selected areas[5] (Anderson et al., 2008b). Subsequently, they were able to differentiate active brain areas in time by the task progression predicted by the model. Anderson et al. (2008a) similarly showed up methodological hurdles that need to be overcome for valid model-brain comparisons on BOLD response in complex cognition tasks, like temporal variability between trials or model-fit assessment. Anderson and Fincham (2014b) already link HMM states – as identified by their fMRI-based multi-voxel pattern analysis – and corresponding processing stages to ACT-R module activity, arguing that states may reflect time periods when the current goal demands a rather constant pattern of resource deployment. They suggest HMM states to serve as outlines for developing more detailed information processing models. Also, they suggest to use ACT-R models as a source of information on how consistent activation patterns within states should be expected to be. Nonetheless, the higher spatial resolution of fMRI compared to EEG already allows for easier interpretation of semantic stage content. Tenison and Anderson (2016) examined mathematical fluency and proposed designing ACT-R models of skill acquisition based on stages derived from HMMs of fMRI data.

In the vein of the works above, module activity can also be compared to EEG data. These offer the advantage of higher temporal resolution of cortical neuronal activity sampling, with the downside of lower spatial resolution and lack of access to inner brain regions. Nonetheless, EEG's fine-grained temporal information facilitates research into activity sequences as predicted by ACT-R. The following review only includes works that directly relate ACT-R output, i.e. time course or module activity, to EEG data – studies that model behavioural data and only inform EEG data by proxy are omitted. With an attentional blink task, van Vugt (2012) related module activity to EEG, finding correlates not only between modules and specific topographies, but specific frequency bands as well. A second study by the researcher proposed markers in EEG oscillation for information transfer between ACT-R modules (van Vugt, 2013), while another follow-up study further established how modules can predict cortical coherence (van Vugt, 2014).

EEG data can be post-processed into so called epochs, sections of EEG with a fixed length. These epochs are often created based on experimental trials which enables a standardised view of EEG potentials relating to each trial. The resulting event-related potentials (ERPs) can then be compared to ACT-R output. Griffiths et al. (2011) designed cognitive models able to reproduce the timing and firing patterns of ERPs in children with a simple selective attention task and were able to generate a very high fit of the model prediction to the recorded ERPs. Cassenti et al. (2011) improved ACT-R models by applying temporal constraints on the models' time course inferred from ERP components. In 2016, Prezenski and Russwinkel outlined a method for both validating ACT-R models directly with EEG instead of behavioural data and verifying additional ACT-R modules, based on matching modules to clusters generated by an Independent Component Analysis (Prezenski and Russwinkel, 2016). This method was further developed in a paper by Klemm et al. (2021) and applied to an associative recognition task.

Klaproth et al. (2020) introduced an ACT-R model based on EEG recordings and interactions of aeroplane pilots that was able to "trace", i.e. predict expert behaviour and produce crucial alerts on deviations from the expected behaviour, using a brain-computer interface to train a classifier on ERP markers and subsequently adapt the model to the classifier. Borst et al. (2013) used the information on processing stages gained from their own EEG-based classifier method in an associative recognition task to evaluate an ACT-R model. Similarly, Borst and Anderson (2015) used the results from their HsMM-EEG study on associative recognition to assess three different theories of associative recognition, where one theory is implemented in an ACT-R model. Anderson et al. (2016) used an ACT-R model of associative recognition to constrain an HsMM model.

As complex approaches are best applied to well-known, proven experiments, we chose the popular mental rotation paradigm for a combined HsMM-EEG and cognitive model analysis. Mental rotation serves as a good exemplary paradigm since it is highly likely to contain qualitatively different stages involving at least visual, spatial and decision-making components. Also, it can be studied in straightforward neuropsychological and behavioural experimental settings consisting of a series of experimental trials with clear-cut temporal markers (trial onset, stimulus onsets, and response).

1.4. Mental rotation

Mental rotation is a mainstay in experimental psychology, particularly in spatial cognition research. As one of the most popular paradigms for testing mental spatial transformation abilities, its simplicity has lent itself to the investigation of distinctions in cognitive processing, leading to a rich history of landmark studies. Shepard and Metzler (1971) introduced mental rotation as an experimental paradigm, in which two abstract objects consisting of linear arrangements of cubes are presented side-by-side. The objects are either exact copies or mirrored forms of one another and, depending on the experiment condition, are rotated 0–180° on the picture plane and/or in depth. Subjects are tasked with deciding if the two objects match or not, and need to mentally rotate either of the objects to reach an informed conclusion. The frequently replicated results showed a linear effect of rotation disparity with no difference between axes, implying mental transformations happen at a constant rate of rotation.

Analysing ocular fixations during a mental rotation study, Just and Carpenter (1976) first outlined a proposal for generating cognitive stages deduced from eye movements of their subjects. Scan paths and fixation lengths were considered respective markers for individual stages of processing during a trial. They proposed three sequential task solving steps, distinguished by processes underlying eye movement: a search stage, a transformation and comparison stage, and a confirmation stage.

Yuille and Steiger (1982) further evaluated processes involved in mental rotation and possible task solving strategies by using stimuli varying in complexity. On the basis that object complexity directly influences RT, they differentiated between two main strategies selected by familiarity with an object: a piecemeal strategy separates stimuli into their respective features and successively transforms them separately to reach a conclusion, while a holistic strategy skips individual features and rotates an object as a whole. Effectively, the first strategy is based on procedural processing and the latter strategy on declarative processing. The holistic strategy thus allows for much faster task processing than the piecemeal approach, but relies on the object being sufficiently familiar to the subject, i.e. by being processed repeatedly throughout the experiment. This distinction in holistic (or wholesale) and piecemeal strategies proved to be a popular explanation for both the influence of object complexity and a learning effect over the course of the experiment (Bethell-Fox and Shepard, 1988; Yuille and Steiger, 1982). Moreover, these strategies can be linked to the more general declarative and procedural strategies respectively, commonly found in psychology or cognitive science and a popular explanation for task learning effects (e.g. Smith, 1994 or Ashby and Crossley, 2010).

Cognitive models of mental rotation have been presented before, for example by Peebles (2019a) or Peebles (2019b). Slight changes to the visual module of ACT-R enabled rudimentary spatial processing features. Using a simplified mental rotation in 2D space, both wholesale

[5] namely lateral inferior prefrontal, posterior parietal, and anterior cingulate cortices, as well as caudate nuclei, respectively.

and piecemeal strategies were included and reached a close fit to human task data.

A review paper by Harris et al. (2013) found indications for the generalizability of mental rotation findings to processes of mental spatial transformation in general. While differences between paradigms were found, spatial transformation paradigms seem to differ mostly in whether they are rigid or non-rigid transformations, i.e. transforming the perspective or the object itself, with mental rotation being rigid and mental folding non-rigid.

Work by Khooshabeh et al. found individual differences in the influence of object appearance on mental rotation difficulty, suggesting that subjects with poor spatial ability incorporate additional features like colour or distinctiveness, while high ability solvers mostly relied on abstract spatial representations (Khooshabeh et al., 2013). They also reaffirmed the finding that the piecemeal strategy is preferably used for complex objects or by subjects with lower solving ability.

Mental rotation serves as one of the most popular and well-defined paradigms in spatial cognition research. A deeper understanding of the cognitive mechanisms guiding mental rotation task solving would therefore undoubtedly help grasp other spatial phenomena, both abstract and applied. For instance, the number and extent of the HsMM-extracted stages could correlate with differences in strategy use or spatial ability on the subject side, or with stimulus complexity on the experiment side. As such, it seemed ideal for the application of this rather exploratory method.

1.5. Research goals

We claim that **(1)** the HsMM-EEG method is a powerful tool to extract the number, sequence, and duration, i.e. the structure of processing stages during a cognitive process from EEG data, and that **(2)** ACT-R models provide a useful source of information about cognitive processes that can guide the semantic interpretation of these processing stages. We therefore argue that **(2a)** linking the HsMM-EEG method with ACT-R models, thereby combining a data- and a theory-driven approach, extends the interpretability of HsMM-EEG results. We additionally explore if **(2b)** further insights into cognitive processing stages in mental rotation are possible by bringing together the HsMM-EEG method with an ACT-R model of mental rotation. By extension, we analyse if **(2c)** common features are apparent visually or statistically in both data sets, and if our conclusions are applicable to spatial cognitive processing in general. To the best of our knowledge, this work is the first to apply a stage-identifying method to EEG data from a mental rotation study, and from a spatial task in general.

2. Methods

Our analyses are based on experimental data from two neuropsychological mental rotation studies. HsMMs were fit to EEG recordings. A cognitive model for the task was designed and fit to behavioural data, incorporating an additional ACT-R module for spatial processing and able to predict behavioural task outcomes and cortical activity. The HsMMs and the predictions by the cognitive model are then compared and set in relation to each other.

2.1. Experiment

The data used for both cognitive model and HsMM analysis was collected as part of a master's thesis (Schischigin, 2014), as well as a bachelor's thesis (Raddatz, 2014). The studies were based on the experimental paradigm used by Shepard and Metzler (1971) and originally conducted to explore the influence of navigation strategy and sex differences in mental rotation, respectively.

Altogether, 42 subjects were acquired in both studies (22 subjects, 11 female from Schischigin, 2014; 20 subjects, 10 female from Raddatz, 2014). For the HsMM-EEG analysis, 19 subjects were excluded. Exclusion criteria were erroneous EEG data files (three subjects), missing behavioural data files (one subject), extremely long RTs, when a subject's mean RT in more than 35% of conditions was larger than the mean RT plus two standard deviations (SDs) (three subjects), electrode impedance above 5 kΩ (seven subjects), and an extreme occurrence of movement artefacts (five subjects). Since the HsMM-EEG method is based on single trials and therefore noise is not reduced by averaging over trials as in ERP analysis, exclusion due to extreme occurrence of artefacts was handled liberally. Exclusion due to electrode impedance and artefacts was blind to behavioural data. In the final sample (N = 23, 15 male) the mean age was 30.52 years (SD = 8.42) with a range from 21 to 52 years. All subjects were right-handed and had normal or corrected-to-normal vision.

Written informed consent was obtained from all subjects. They were compensated with either 8€ or course credit. On average, subjects spent 60 min on an experiment run. Trials were presented on a monitor in a darkened room. Each trial started with the presentation of a fixation cross for 1 s. Afterwards, the first stimulus (*reference stimulus*) was presented as-is on either the left or the right side of the screen. After 1 more second, the second stimulus (*target stimulus*) was presented on the other side rotated by either 0°, 50°, 100°, or 150° on the picture plane, depending on the trial condition (see Fig. 1 for an example).

Each stimuli pair was selected from a set of 16 figures, taken from Peters and Battista (2008). Figures consisted of a set of 10 white cubes with black outlines, attached together in 3–4 straight lines each. Presentation side and figure type were balanced for every subject. The task consisted in discerning if the two shown stimuli depicted the same figure, or if the target stimulus showed a mirrored version of the figure. Subjects were instructed to mentally rotate one of the objects clockwise until an informed decision could be reached. Responses were given per two vertically aligned buttons with the right hand (left hand in Schischigin, 2014), with the top button for a "same" answer and the bottom button for a "mirrored" answer. Each possible trial combination was presented during an experiment block with 6 blocks overall, resulting in 768 trials in sum.[6]

Behavioural as well as EEG data were collected in each study. EEG measurements were recorded by 64 electrodes per the 10–20 system and a sample rate of 500 Hz. The FCz electrode was used as online reference. Note that with 64 electrodes, the EEG data used in the present work was recorded with twice as many electrodes as the EEG data recorded by Borst and Anderson (2015) who used 32 electrodes. A 250 Hz low-pass filter was applied to avoid aliasing artefacts. The signal was amplified via BrainAmps (Brain Products GmbH, 2014).

2.2. HsMM-EEG method

2.2.1. EEG preprocessing

Subject-wise EEG data were first re-referenced offline to the Cz channel and a 0.016 Hz high-pass filter was applied. In addition, a 0.5–30 Hz band-pass filter was applied to attenuate high-frequency noise. Power line noise was removed with a 50 Hz DFT filter. The subject-wise continuous EEG data was epoched from reference stimulus onset until subject response. The data was then downsampled to 100 Hz. Within each subject data set, the first 30 trials (20 training trials plus the first 10 trials of the first experimental block) were removed since "warming-up" effects can be assumed for this early experimental phase, in that subjects are typically converting from a slow, declarative to a smooth, procedural execution of the task (Anderson et al., 2004). Likewise, the last 30 trials were removed due to possible fatigue effects. Also, trials with RTs longer than two SDs from mean RT per subject, as well as trials containing muscle or jump artefacts as identified by visual inspection, were excluded. For each subject, only trials that were solved

[6] 16 figures * 4 rotation conditions * 2 sameness conditions * 6 blocks = 768 trials.

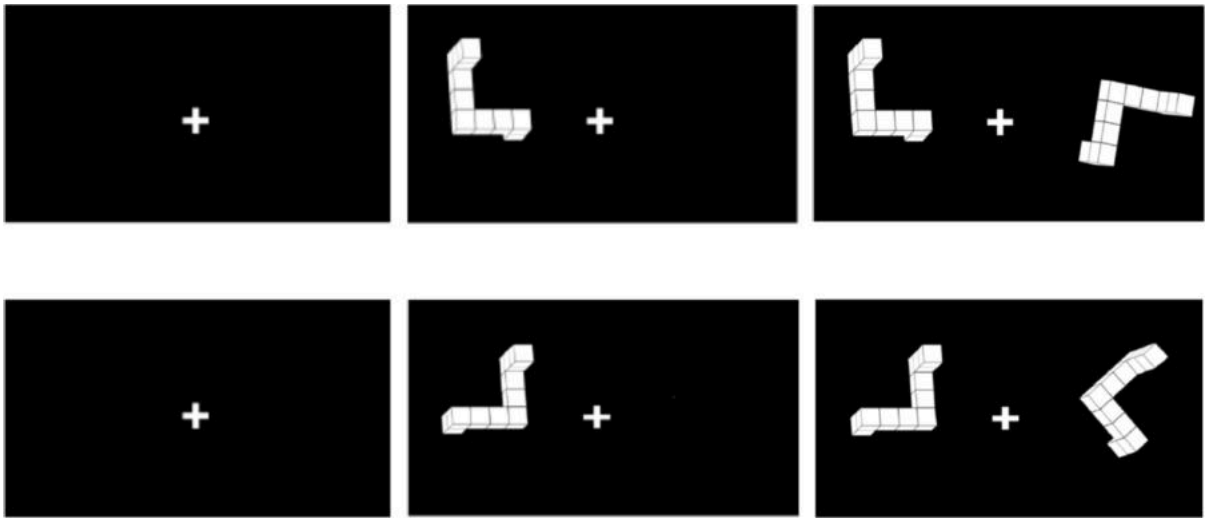


Fig. 1. Two examples of a mental rotation trial as presented to subjects. After a fixation cross is displayed for 1s (*first column*), the reference stimulus is displayed (*second column*) left or right of the fixation cross. Following another 1s, the target stimulus appears (*third column*) opposite of the reference. Both figures remain on screen until a response is given. *Top row:* same stimuli, *bottom row:* mirrored stimuli.

correctly were considered for further analyses. Overall, 9,84% of trials were rejected due to RT outliers, artefacts or incorrect subject responses.

The subject-wise EEG recording was decomposed into independent components by means of the EEGLAB infomax algorithm (Bell and Sejnowski, 1995). Components reflecting eye blinks and saccades or heartbeat were visually identified and projected out of the data. Subsequently, the VEOG and the Cz electrode (as the offline reference) were excluded from the EEG data set, leaving 62 channels. The trial-wise EEG epochs were baseline-corrected to remove random signal drift within trials, with the baseline window spanning the whole trial length. In ERP analyses such drift would disappear due to the averaging procedure. Since the HsMM estimation uses single-trial data as input, removal of random variation between trials is especially important (Borst and Anderson, 2015).

Note that to date no objective criteria exist for tailoring EEG data preprocessing to the HsMM-EEG analysis. Therefore, to ensure methodological similarity, our preprocessing followed the procedure applied by Borst and Anderson (2015) concerning the steps of band-pass filtering including the exact filter span, removal of incorrect trials, and the necessity of baseline correction. To ensure consideration of task specifics, the steps of re-referencing and high-pass filtering including the filter size followed the procedure of Raddatz (2014).

2.2.2. *HsMM-EEG analysis*

Our application of the HsMM-EEG method is based on the steps conducted by Borst and Anderson (2015) (see Fig. 2 for an overview of the procedure). We point the interested reader to the original paper for a detailed description of the underlying stochastic concepts.

In the HsMM-EEG analysis, two main parameters are estimated for each state i: (1) A brain signature M_i that reflects the average EEG activation pattern during this processing stage, and (2) the gamma distributions G_{in} that represent the state's durations (and their variability) for the conditions n of the experiment. Borst and Anderson (2015) note that since the HsMM-EEG method assumes that the brain signatures are the same for each condition and conditions only differ in their duration estimates, conditions should be analysed jointly only if they can be hypothesised to consist of the same duration and order of processing stages; where this is not the case, conditions should be committed to separate HsMM-EEG analyses. For our analysis of the mental rotation task we decided to analyse the experimental conditions with a 0° rotation disparity between reference and target stimulus (in the following: *non-rotated conditions*) separately from those conditions with a 50°, 100°, or 150° rotation disparity (in the following: *rotated conditions*). We hypothesise that the rotated experimental conditions involve different processing stages than the non-rotated conditions, which are special conditions in that subjects do not have to execute mental rotation at all to solve the task.

Prior to the actual HsMM-EEG analyses, the preprocessed EEG data had to pass through a preparation procedure that is tailored to the HsMM-EEG method. We based this preparation on the concepts and suggestions by Borst and Anderson (2015). First, so-called "snapshots", each spanning 80 ms, were created by restructuring the data. From eight 10 ms samples in 62 real channels we created one 80 ms snapshot with 496 virtual channels (consequently, each real channel appeared 8 times in a snapshot, representing the different time points within the 80 ms period; see Borst and Anderson (2015) for a discussion of 80 ms as snapshot length). Creating snapshots allows for presenting the temporal profile of the data to the HsMM estimation procedure, which is important because crucial information is contained in the temporal profile of the EEG data (Borst et al., 2013). Also, it reduces the amount of data points in a trial for the HsMM-estimation process. After snapshot creation, each data point contained information not only about the mean voltage in each channel, but also about whether this voltage increased or decreased over the 80 ms interval of each channel (Borst and Anderson, 2015). After snapshot creation, each trial was composed of a certain number of snapshots (with snapshots that crossed into the post-response period being excluded). The second step in the data preparation procedure was to normalise each virtual channel to a mean of 0 and a standard deviation of 1, and to apply a spatial principal component analysis (PCA) to the 492 virtual channels. Afterwards, the results of the PCA were again normalised. This was performed across all subjects, resulting in a single set of PCA coefficients and PCA components matching across subjects. In accordance with Borst and Anderson (2015), we decided to use the first 100 PCA components for the HsMM-EEG analysis.

For the HsMM-EEG analysis, we adapted the HsMM algorithms provided by Borst and Anderson (2015). The PCA matrix and four additional vectors indicating subject, trial, snapshot-in-trial, and experimental condition for each snapshot were entered into the algorithms. Thereby, those brain signatures M_i and gamma distributions G_{in} that maximise the likelihood of the data, given a particular number of states, were calculated. Since the 100 PCA factors are basically distributed as independent normals, each brain signature M_i is described by a vector of 100 means μ_{ik} that represent independent normal distributions

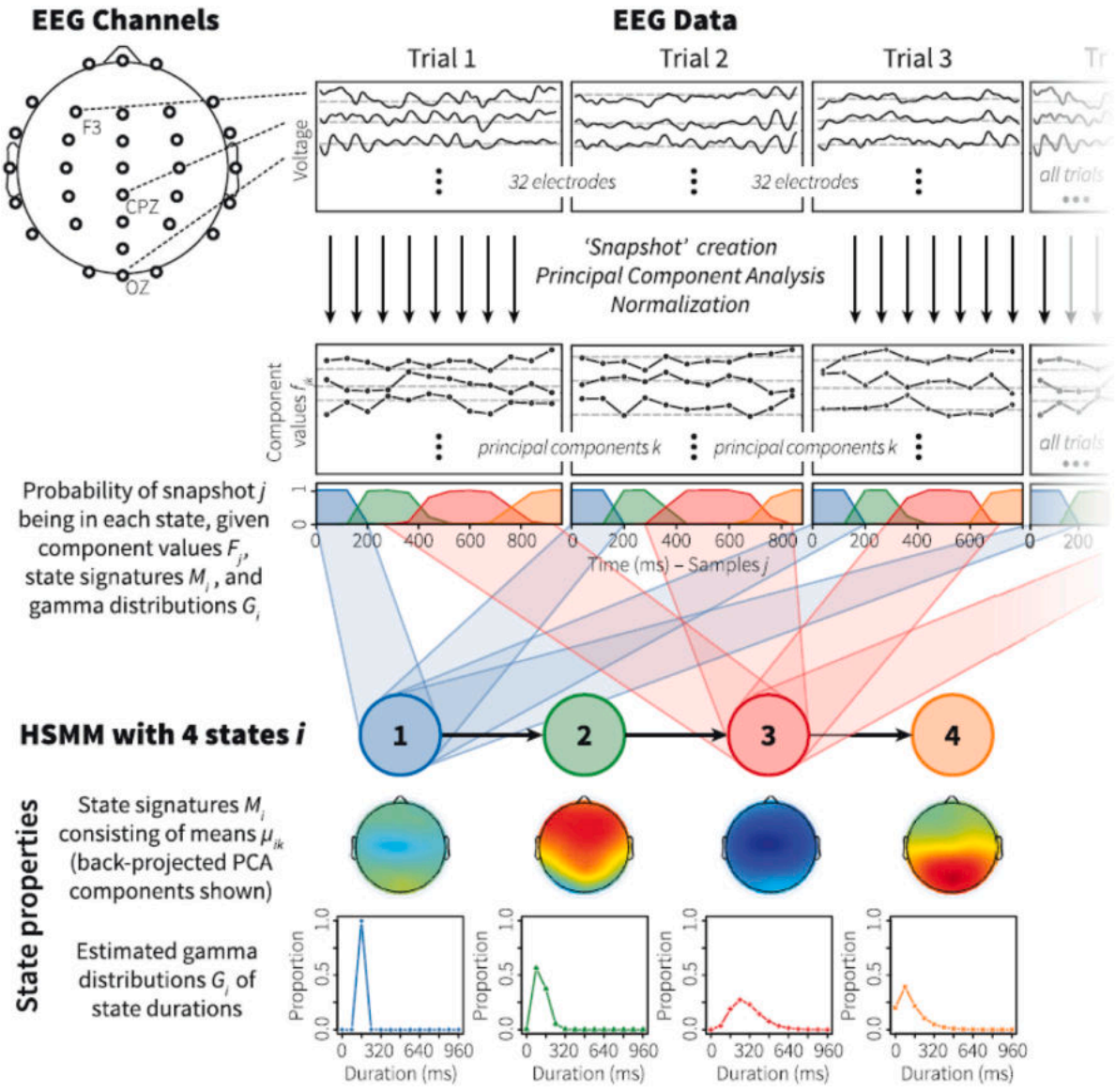


Fig. 2. Overview of the HsMM-EEG method. *Reprinted from "The discovery of processing stages: Analyzing EEG data with hidden semi-Markov models", by* Borst, J.P., & Anderson, J.R., 2015*, NeuroImage, p. 63.*

with unit variance (Borst and Anderson, 2015).

Following Borst and Anderson (2015), we applied a LOOCV with a subsequent sign test to check if the extra parameters explain sufficient extra variance to be justified, since HsMMs with more states have more parameters to fit the data and will thus typically result in a better fit. The LOOCV method estimated maximum-likelihood parameters for an *r*-state HsMM using the data of all but one subject. The data of the remaining subject was then split in half, and maximum-likelihood gamma distributions were estimated for the first half of the remaining subject's data given the brain signatures of the other subjects. Using the gamma distributions and the state signatures of the other subjects, we calculated the log-likelihood of the second half of the remaining subject's data. This was repeated for the other half of the remaining subject's data. Then, the two log-likelihoods were averaged. We repeated this process for all subjects and for HsMMs with different numbers of states.

Although the likelihood maximisation algorithm is in itself stochastic, repeated runs confirmed that both the likelihood maximisation algorithm as well as the LOOCV output are deterministic and always yield the same results when the same data is used as input.

We used a sign test to select the HsMM with the optimal number of states, and chose the highest *k*-state model that fitted the data of a significant number of subjects better than all ($l < k$)-state models. Note that a ($k + 1$)-state model will fit the data of $n - 1$ subjects better in the estimation phase because it has more parameters, even if the true number of states is *k*; however, it is at least as likely to fit the *n*th subject worse (Anderson and Fincham, 2014a). With 23 subjects, the increase becomes significant when at least 16 subjects improve ($p < 0.05$). Note that a high likelihood for a solution with a certain number of states does not mean that solutions with fewer states are wrong, rather, they may just be at a different level of aggregation (Borst and Anderson, 2015).

After we had determined the optimal number of states, we estimated a single HsMM based on the data of all subjects. The state signatures of this model were then used to estimate separate gamma distributions for each subject and condition. Based on these gamma distributions, the average state duration for each subject and condition could be calculated. Using these subject-specific models, we calculated the state durations per condition, averaging the state durations over all subjects. Further, the subject-specific models served as basis for calculating the probability for each snapshot to be in a certain state. We used this to

estimate brain activity for each state, condition, and subject by multiplying the EEG data of each sample with its assigned probability, summing over the resulting values, and dividing by the summed probability over all samples. The resulting EEG pattern for each state, condition, and subject was then averaged over subjects. It is important to note that these patterns simply represent a weighted average of the EEG signal (Borst and Anderson, 2015).

2.3. Cognitive model

The cognitive model was designed to solve a simulated version of the experiment described above. By taking advantage of ACT-R's architecture, the processing of each individual module and their interactions are controlled to simulate task solving for a mental rotation task.

Initially, a simple cognitive process model was hypothesised, broadly depicting the steps necessary for task solving (Fig. 3). This process model is roughly modelled after the process analysis by Just and Carpenter (1976), who used eye tracking to discern separate phases during a mental rotation task, as well as the basic strategies theorised by Yuille and Steiger (1982). The presented stimuli are visually perceived and encoded into mental representations, which are then gradually rotated to align more closely to each other. Depending on if a certain threshold of similarity can be achieved, a human solver would then give a match or a mismatch response. Additionally, the model is allowed to bypass rotating if the stimuli can be visually compared directly, thereby creating "shortcuts" for non-rotated trials. The process formalisation was then developed into a production rule model, serving as the basis for an implementation of the cognitive model in ACT-R. Note that the process model is, by nature, an intuitive and subjective account of cognitive processes during mental rotation. Albeit resulting from a best effort to replicate models from established literature, it does not lay claim to be a complete nor a unique account of cognition during human mental rotation solving, and alternatives exist (e.g. Peebles, 2019a).

The model makes use of reinforcement learning and spreading activation as learning mechanisms, implemented in ACT-R as *utility learning* and *base-level learning*, to simulate strategy choice and implicit preference of productions. Base-level learning governs the strength of specific memory contents, like mental representations of stimuli from past trials. These mental representations gain stronger activation through repetitions, and at some point can be relied on instead of transforming the same object again, serving as a time-saving bypass and allowing faster responses. Utility learning mediates the gain of activation by helping the cognitive model gauge the usefulness of this reliance on task memory. To ensure cognitive plausibility and generalizability of the produced predictions, the cognitive model relies on a novel addition to the ACT-R architecture, the spatial module.

Based on an idea by Gunzelmann and Lyon (2007), the spatial module seeks to add cognitively plausible functionality for three-dimensional data in ACT-R, providing a unified way of simulating spatial transformation tasks. In contrast to Gunzelmann and Lyon, this framework relies heavily on standard ACT-R mechanisms. The spatial module acts as both storage for mental spatial representations and a means of transforming the latter. To enable spatial processing of information, chunks are extended by point clouds, representing an object in 3D space. These spatial chunks then allow for algebraic manipulation of objects while maintaining full compatibility with all default ACT-R mechanisms. Transformations are requested by the model and, if conforming to architectural and modular constraints, applied to the spatial object. The module's core function is calculating the time delay necessary for spatial transformation, based on the hypothesised linear process commonly found in mental rotation studies. This processing cost is generated by a function based on current research, constituting an attempt to find a common denominator underlying mental spatial transformations and thus facilitating explainability, validity, and generalizability. Currently, the following simplified formula is used:

$$Transformation\ delay = F * x$$

including a delay factor F and the raw input value of the transformation x. Two optional factors N representing the increased mental effort from successive transformations applied to the spatial representation, as well as M for assigning weights to different transformation modalities, were so far not used in this study.

Although a cognitive model cannot be meaningfully separated into cognitive stages, its productions can be coarsely grouped by their purpose. A model run for the experiment can thus be roughly divided into three approximate phases, analogous to Just and Carpenter (1976): visual encoding, transformation and comparison, and a motor response stage. Note that these phases are not identical to the aforementioned

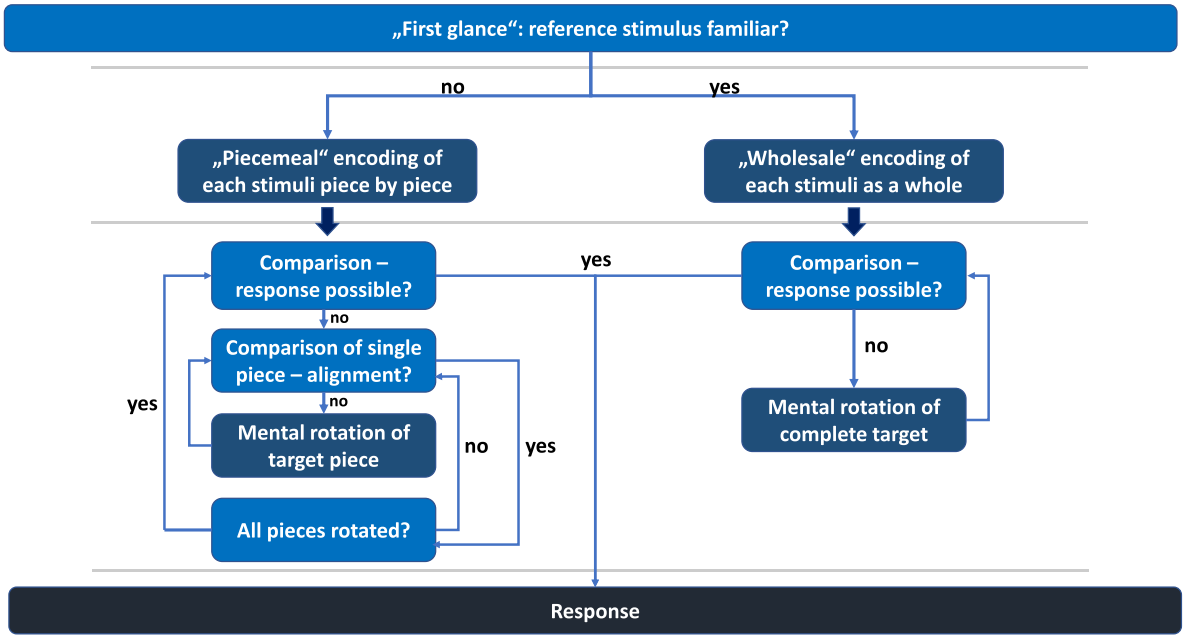


Fig. 3. Flow chart depicting a simple process model for mental rotation. Note the difference in complexity between the piecemeal (*left*) and the wholesale (*right*) strategy.

HsMM stages, but describe a grouping of theorised task-solving steps.

Visual encoding – At the start of a trial, the model is first presented with the reference stimulus. The model visually perceives the complete object and encodes it as a mental image. Prior research (Yuille and Steiger, 1982) suggests a familiarity mechanism to be the deciding factor in a solver's strategy choice for the rest of the trial. The model implements this in the form of instance-based learning (Gonzalez et al., 2003), relying on ACT-R's base-level learning mechanism to determine if a prior representation can be retrieved from declarative memory or not: to see if the encoded stimulus is already familiar to the model, its memory is queried for an existing representation of the structure from earlier trials. If the stimulus is familiar, i.e. presented often and recently enough, the model proceeds with the faster wholesale strategy, while a novel stimulus starts the piecemeal strategy. In the latter case, the reference object is encoded again, but instead of the whole structure its individual features are now one after another saved as a separate mental image. Afterwards the model waits for the target stimulus to appear, which it then also encodes in the same manner as the reference object.

Transformation and comparison – Once both stimuli are encoded and stored in the imaginal and spatial modules, respectively – simulating a solver actively maintaining both mental representations – the transformation process starts. By using the spatial module, the two representations are compared with each other: the similarity of both representations is measured and either can be declared a "match" or a "mismatch" directly (representing a visual shortcut) or understood to require mental rotation: depending on its choice of strategy, only part of the target or the complete object is rotated. Each rotation is followed by a comparison process, in which the model again gauges the similarity of the two representations: if similarity has improved, an additional rotation is initiated; if similarity has worsened, the two objects are considered a "mismatch". If similarity reaches the threshold, both representations are considered equal and a "match". For the wholesale strategy, the appropriate response is prepared after one iteration of transformation and comparison, while during the piecemeal strategy, the degrees of rotation necessary to reach an informed decision are remembered and applied to all subsequent pieces. When all pieces reach the similarity threshold after being rotated in this manner, the piecemeal strategy concludes a "match", otherwise a "mismatch".

Motor response – After collecting enough information to decide on a match or mismatch response, the model enters its response via a virtual keyboard. The model can reach this point at several times during a run:

- "obvious" matches or mismatches are discerned directly after visual encoding,
- a mismatch is returned if most, but not all pieces are spatially equal,
- a remembered rotation value during the piecemeal strategy does not work on a subsequent piece and produces a mismatch,
- rotation yields no improvement on the similarity leading to a mismatch,
- all pieces of the target object match all pieces of the reference object after rotation during the piecemeal strategy or
- the target object matches the reference object after rotation during the wholesale strategy.

This triggers a motor action initiated by the manual module, simulating a key press on the respective answer key for "match" or "mismatch". Note that this means that the transformation and comparison phase is not a prerequisite of the motor response stage, as a decision can be made purely by visual comparison. The model was fitted to the subjects' RTs, aggregated for each experiment block and target rotation, to ensure a sufficient predictive quality of the model before generating module activity. Model parameters were chosen such that they produced high correlation and low root mean square error (RSME) between model and human data. Parameters used were memory retrieval latency factor, retrieval threshold of memory content, added retrieval request noise, added production utility noise, and spatial transformation time cost.

For the generation of model predictions, each human subject in the experiment was individually replicated, that is, their respective order of trials was followed to control for sequence or learning effects. During each trial, the activity produced by each module is saved in binary form, i.e. a module is active or inactive at a specific time: a module is considered active when it is explicitly processing cognitive content. While module output is binary by nature, aggregation over multiple trials enables a gradient output. Module activity for this study was created with a 20 Hz sample rate. Only trials that were answered correctly by both subject and model were considered, leading to rejection of about 3.2% of model data. A more in-depth look at the model, and the underlying spatial module, can be found in Preuss and Russwinkel (2021).

2.4. HsMM-EEG – cognitive module activity comparison

Module activity of a cognitive model can not be considered a straight analogue to HsMM-EEG results: while cognitive processes are simulated by the model, they are not clearly separable stages as those resulting from HsMM-EEG. So far, there are no established approaches to relate both model and HsMM data to each other. For this reason both a visual and an exploratory statistical analysis of the two datasets was conducted. On one hand, the two time series are compared visually, assessing potential similarities and common features. On the other hand, the relation between module activity patterns and HsMM stages was analysed statistically. As the resulting HsMM stages model data up to ~3200 ms, only cognitive model data from 0 to 3200 ms was included in the analyses.

2.4.1. Visual comparison

As a first step, the processing stages as resulting from the HsMM-EEG analysis were contrasted with ACT-R module output in a purely visual manner. With this approach we aim to demonstrate what kind of information can be deduced from the two types of data and how they can be aligned. The results from the visual comparison shall be used as a first insight into semantic stage content.

2.4.2. Statistical comparison

For insight into the predictive quality of module activity patterns on HsMM processing stages, a chi-square test was conducted, with Cramér's V offering a measure of effect size. Module activity generated by the cognitive model was sampled with a rate of 20 Hz and assigned to its respective HsMM stage in regards to trial time. Subsequently, the amount of active samples per module and HsMM stage were counted.

2.5. Method summary

We attempt to get deeper insight into mental rotation processes on a neuropsychological level by application of the HsMM-EEG method. Moreover, we aim to deliver proof-of-concept for linking the HsMM-EEG method with ACT-R output and propose that it advances semantic interpretability of distinct processing stages in a cognitive process.

3. Results

In this section, we will first briefly describe the behavioural results of the mental rotation experiment as they were originally presented by Raddatz (2014). With the following results of the HsMM-EEG analysis we aim to uncover differences in number and duration of processing stages between rotated and non-rotated conditions, respectively. What follows are the results of the model fitting process, that is, the correspondence between the ACT-R model output and the experimental results. With the subsequent comparison of the HsMM-EEG results with the ACT-R module activities we aim for an interpretation of the semantic content of the identified processing stages and thereby also at a better understanding of the identified differences between rotated and

non-rotated conditions.

3.1. Experiment results

Raddatz (2014) conducted an ANOVA for the effect of factors rotation disparity, sameness of presented stimuli, sex of subject, and sample source on RT. The author reported significant effects for rotation ($F(3, 102) = 126.634$; $p < .001$) as well as sameness ($F(1, 34) = 53.117$; $p < .001$). Post-hoc tests revealed a significant contrast for linearly increasing RTs with increasing rotation disparity, and significantly longer RTs for mirrored than same stimuli. Furthermore, the author conducted an ANOVA for the effect of the same factors on error rate. Again, significant effects were found for rotation ($F(3, 102) = 41.074$; $p < .001$) and sameness ($F(1, 34) = 11.704$, $p = .002$). Post-hoc tests revealed a significant contrast for linearly increasing error rates with increasing rotation disparity, except for the difference between 0° and 50°, and significantly higher error rates for mirrored than same stimuli.

These findings replicated the commonly reported linear effect of rotation condition on subject performance, which indicates that subjects indeed performed mental rotation. Fig. 7 shows behavioural results in conjunction with results from the cognitive model.

3.2. HsMM-EEG results

Since we assume mental rotation processes to differ between non-rotated conditions and rotated conditions, we conducted separate analyses for rotated and non-rotated trials. For non-rotated conditions a 7-state HsMM was identified as providing the most likely description of the data while for rotated conditions, a 6-state HsMM was found to best account for the data. Fig. 4 shows the average gain in log-likelihood of k-state HsMMs over a 1-state HsMM. Table 1 shows the number of subjects whose log-likelihood improved from k-state HsMMs over ($l < k$)-state HsMMs. Note that the average log-likelihood can decrease with an increasing number of states because the likelihood estimation was independent from fitting the models (Borst and Anderson, 2015).

In a sign-test with 23 subjects, the increase becomes significant when for at least 16 out of 23 subjects the likelihood for a certain number of stages improves over another ($p < .05$). For non-rotated conditions, the 7-state HsMM had the highest log-likelihood and was more likely than HsMMs with fewer states for at least 16 subjects, corresponding to a significance level of $p < .05$. For rotated conditions, the 9-state HsMM had the highest log-likelihood, but only the 6-state HsMM turned out to reach significance at the level of $p < .05$, being more likely than HsMMs with fewer states for at least 16 subjects. Note that it might be possible to sub-divide these states into smaller states, however, there are always at least as many states as is indicated by the optimal number of states resulting from the sign test (Borst and Anderson, 2015). We will therefore focus on the 7-state solution for non-rotated conditions and the 6-state solution for rotated conditions.

Figs. 5 and 6 show the properties of the identified processing stages, corresponding to the states in the HsMMs. Fig. 5 shows the brain signatures for the identified stages for non-rotated and rotated conditions, respectively. Fig. 6 shows stage durations adding up to entire experimental trials, averaged over subjects, for each experimental condition.

The brain signatures indicate that the stage-wise aggregated EEG activity patterns widely match between non-rotated conditions and rotated conditions for the first four stages as well as for the last stage. Moreover, the sixth stage in the non-rotated conditions seems to match with the fifth stage of the rotated conditions, while the sixth stage in the non-rotated conditions shows a pattern that does not match with any of the stages in the rotated conditions, indicating that the additional stage in the non-rotated conditions is in the fifth place. Given the apparent correspondence between the stages 1, 2, 3, 4, 5, and 6 in both kinds of conditions, we will hereinafter refer to a stage numeration from 1 to 6 for all conditions, with the additional stage 4b only being present in the non-rotated conditions.

To analyse the effect of experimental condition on processing stage duration, we performed a two-way ANOVA with the factors rotation disparity and sameness of reference and target figure. Table 2 reports the effects of experimental condition. To account for the seven stages, we consider results significant when the p-values were smaller than a Bonferroni-corrected threshold of $\frac{.05}{7} \approx .007$.

3.3. Cognitive model results

As a close fit between cognitive model and behavioural data is paramount for accurate module activity prediction, parameters of the cognitive model were adjusted to maximise correlation and minimise prediction error. A match threshold value of 20 yielded the best results during initial testing. For initial model fitting, experiment data was aggregated by condition and experiment block, resulting in 24 mean values which served as a basis of comparison between human and model data. A conservative grid search of selected parameters resulted in 162 possible parameter combinations. Testing these combinations with the aim of replicating the aggregates as closely as possible resulted in an optimal fit to the experiment data with the following parameter values:

- Latency factor: 0.2 (default: 1.0)
- Retrieval threshold: −1.0 (default: 0.0)
- Retrieval noise: 0.5 (default: none)
- Utility noise: 1 (default: none)
- Spatial delay: 0.03 (default: n/a)

3.3.1. Correlation

The parameter-fit cognitive model correlated well with behavioural data. A strong correlation of RTs over rotation and experiment block was achieved, with little deviation between predicted RTs ($r(22) = 0.92$; $p < .001$; RMSE = 0.23). Rotation disparity by itself, aggregated over all blocks reached a strong correlation ($r(2) = 0.97$; $p < .05$, RMSE = 0.139), as does experiment block aggregated over all rotation conditions ($r(4) = 0.84$; $p < .05$, RMSE = 0.22) (see also Fig. 7). Standard deviation of model-predicted RTs is close to that of the behavioural data, albeit slightly lower ($SD_H = 1.953$; $SD_M = 1.421$).

3.3.2. Regression analysis

To test the similarity between cognitive model and behavioural data, a multiple linear regression analysis was performed, gauging the influence of experiment block, rotation disparity and data source (human or model) on RTs. The three predictors explained 53.9% of the variance ($R^2 = 0.539$, $F(47, 840) = 23.04$, $p < .001$). Rotation disparity was a significant predictor for RTs ($\beta = 0.66$, $p < .001$), as was the interaction between experiment block and rotation disparity ($\beta = -0.29$, $p < .05$). Data source showed no influence on RTs ($\beta = 0.02$, $p = .43$), showing no significant difference between human and model results. Fig. 7 shows a linear effect in the results, with increased rotation disparity effecting increased RTs. Over the course of the experiment blocks, RTs are generally lowered, with this effect being more pronounced for higher rotation disparity.

3.4. HsMM-EEG – cognitive module activity comparison results

To gain insight into the semantic content of the processing stages as identified by the HsMM-EEG method, we compared the results of our HsMM-EEG analysis to the module outputs of our ACT-R model. In the following, we first describe the results of an initial visual comparison between the two, followed by a statistical analysis.

3.4.1. Visual comparison

Since to date no established convention for the visual comparison of processing stages resulting from HsMM-EEG and ACT-R module output

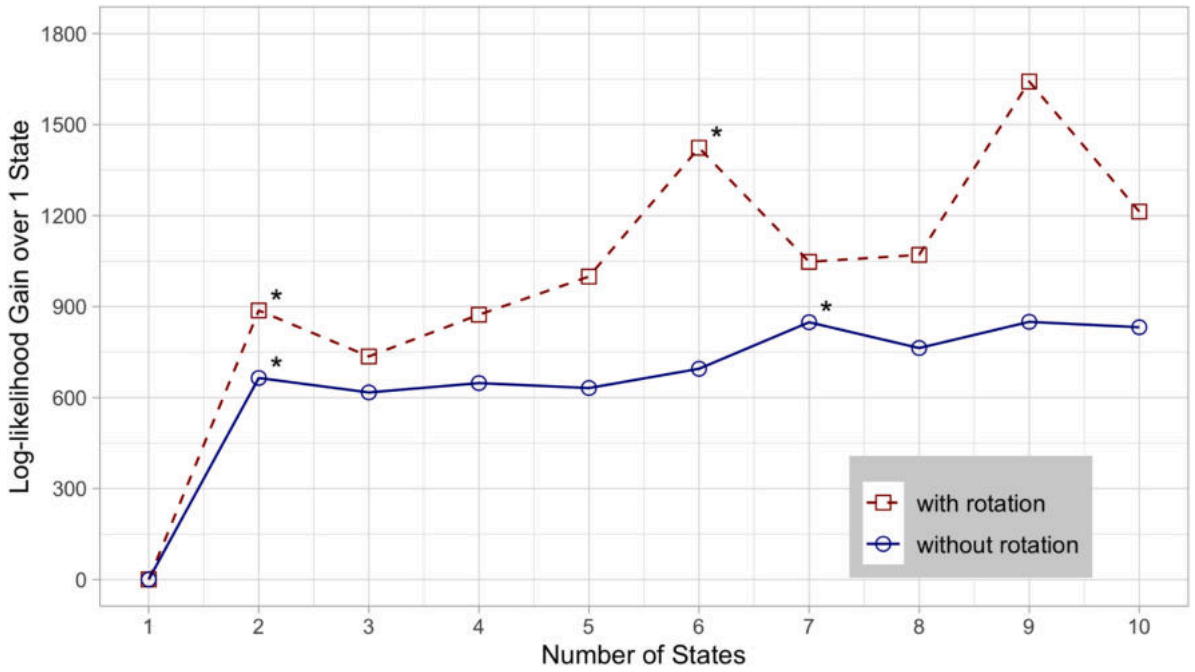


Fig. 4. Results of the LOOCV-procedure. The average gain in log-likelihood of a k-state HsMM over a 1-state HsMM. "*" marks significant increases in log-likelihood over all fewer-state solutions as indicated by the sign test.

Table 1
Number of subjects with an increasing log-likelihood over fewer state solutions.

Number of states	Number of subjects with an increasing likelihood								
	>1 state	>2 states	>3 states	>4 states	>5 states	>6 states	>7 states	>8 states	>9 states
Without rotation									
2	23	–	–	–	–	–	–	–	–
3	23	6	–	–	–	–	–	–	–
4	22	13	15	–	–	–	–	–	–
5	22	11	13	8	–	–	–	–	–
6	22	14	16	14	16	–	–	–	–
7	22	16	17	16	18	16	–	–	–
8	22	15	17	15	17	12	3	–	–
9	22	18	20	18	18	18	13	19	–
10	22	15	16	15	17	14	11	14	11
With rotation									
2	23	–	–	–	–	–	–	–	–
3	22	6	–	–	–	–	–	–	–
4	22	12	17	–	–	–	–	–	–
5	22	13	18	13	–	–	–	–	–
6	22	16	20	18	18	–	–	–	–
7	19	15	17	13	10	3	–	–	–
8	20	13	14	13	12	5	8	–	–
9	21	16	18	17	16	12	18	17	–
10	19	14	14	14	13	7	12	9	7

exists, we decided for an explorative approach described in the following. Fig. 8 shows a comparison that aligns human RT distribution with mean RTs, HsMM-EEG results and ACT-R module activity. Module activity is aggregated over all ACT-R modules, but separated by all eight possible conditions to match it with the HsMM-EEG bar plots. HsMM-EEG processing stages are displayed as background colour for module activity to allow for interpretations. Human RT distributions by trend become wider through the progression from condition 0°/same to 0°/mirrored, to 50°/same, to 50°/mirrored, to 100°/same, to 100°/mirrored, to 150°/same, to 150°/mirrored (for the sake of simplicity, this progression will hereinafter be referred to as "increasing condition difficulty"). However, distributions are relatively matching the time span of stage 6. As expected, aggregated module activity does hardly, if at all, differ between conditions in the time span from 0 ms until 1000 ms, which is the time span from reference stimulus appearance until target stimulus appearance. This already underpins the assumption that in this time period only processes occur that do not differ between conditions. In the time span after 1000 ms, slight deviations in the distribution of peaks and flats are visible, especially for conditions 0°/same and 0°/mirrored. The fact that aggregated module activity does not exhibit rightward shifts on the time axis with increasing condition difficulty, although the HsMM-EEG results suggest that the beginning of the stages 4, 4 b, 5 and 6 are linearly delayed with increasing condition difficulty, makes semantic interpretations for these stages difficult at this point however. Also, the module activity patterns for the 0°/same and the 0°/mirrored condition do not provide information yet about why there is an additional stage in these conditions. In a next step the ACT-R module activity will be functionally differentiated, i.e., separated

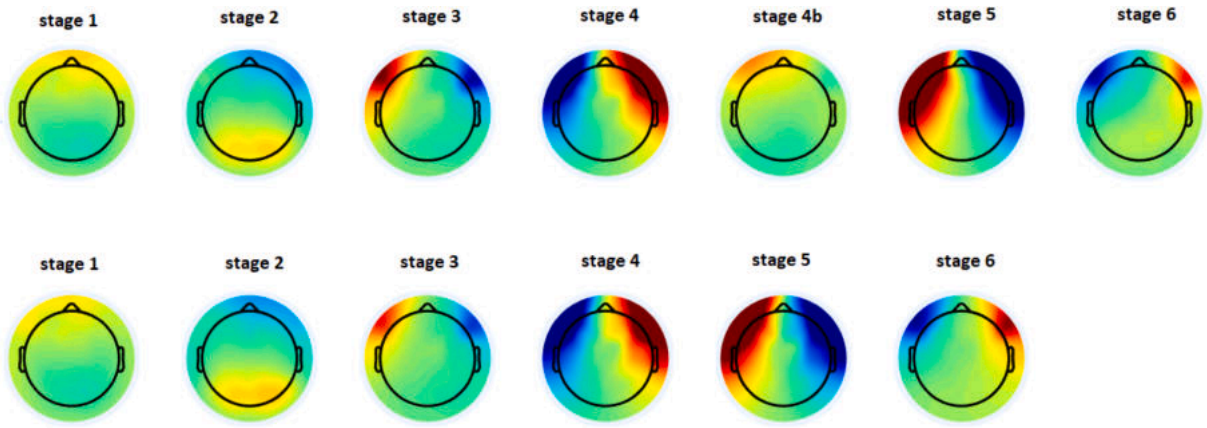


Fig. 5. State signatures for non-rotated (*top row*) versus rotated conditions (*bottom row*).

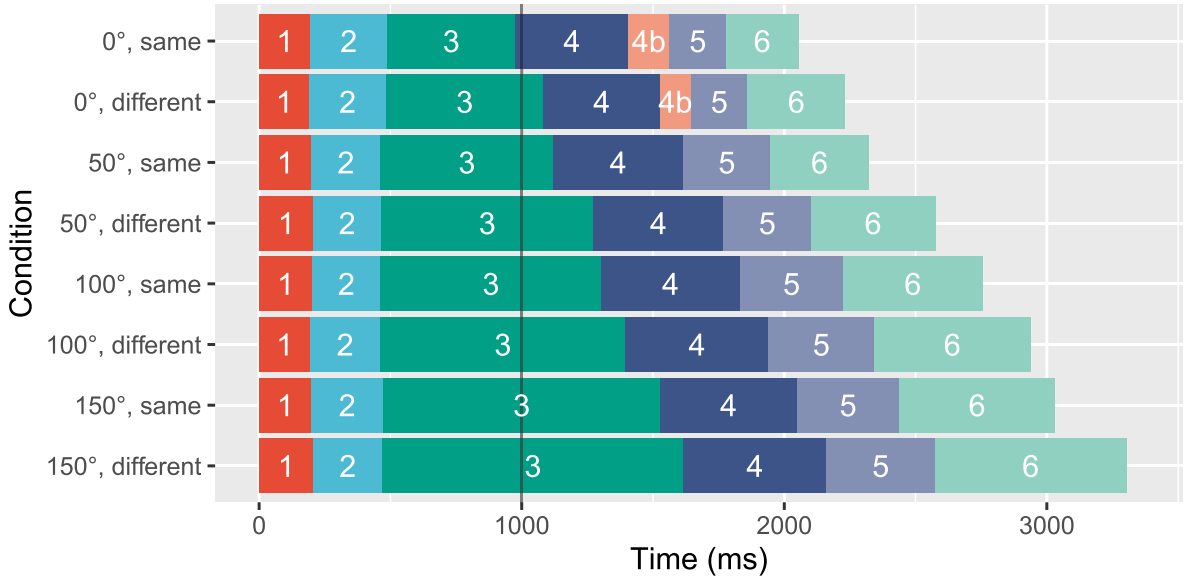


Fig. 6. HsMM stages over the course of a trial, separated by all 8 condition combinations. The additional stage 4b only applies to non-rotated conditions (0°, for same and mirrored stimuli).

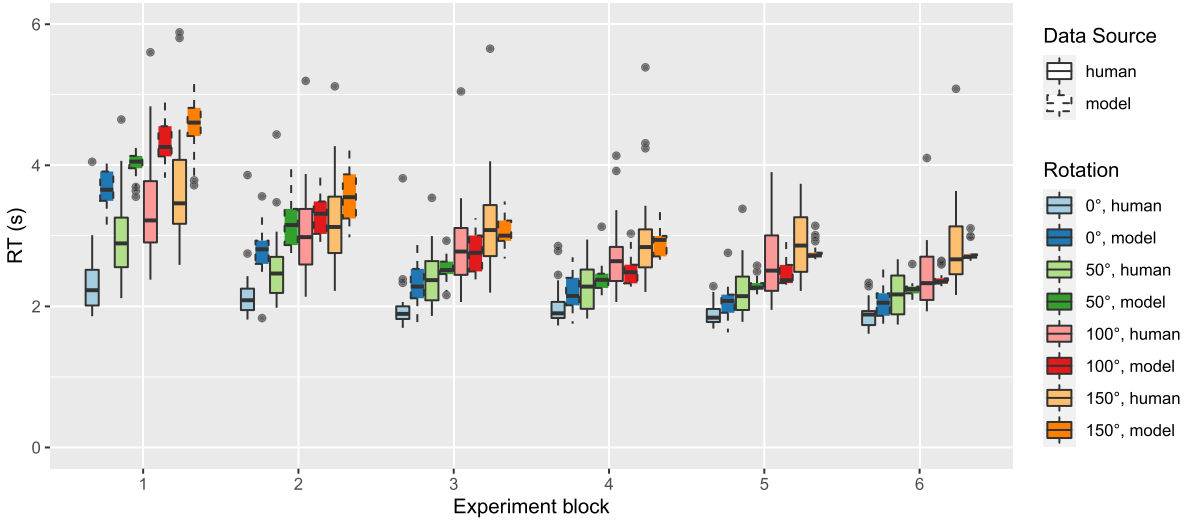


Fig. 7. RTs of subjects (*solid black outline*) and cognitive model (*dashed red outline*), grouped by rotation condition and experiment block.

Table 2
Effects on stage durations for subject-wise data. The significance level was chosen as a Bonferroni-corrected threshold of $\frac{.05}{7} \approx .007$. Significant effects are highlighted in bold type.

Source	F	p	η^2
stage 1			
degree	.245	.87	.004
sameness	.014	.91	< .001
degree * sameness	.197	.90	.003
stage 2			
degree	1.80	.15	.03
sameness	.10	.75	< .001
degree * sameness	.10	.96	.002
stage 3			
degree	**21.46**	**<.001**	**.26**
sameness	4.93	.03	.02
degree * sameness	.07	.98	< .001
stage 4			
degree	**6.23**	**<.001**	**.10**
sameness	.49	.49	.002
degree * sameness	.07	.98	< .001
stage 4b			
sameness	1.63	.21	.04
stage 5			
degree	**28.71**	**<.001**	**.33**
sameness	.030	.59	< .001
degree * sameness	.13	.95	.001
stage 6			
degree	**19.97**	**<.001**	**.24**
sameness	**8.70**	**.004**	**.04**
degree * sameness	.19	.91	.002

by different modules.

Fig. 9 shows the activity of six ACT-R modules separated by all eight possible conditions. The modules have been selected for their functional relevance for cognitive processes. At first sight, the following findings can be identified: **(i)** Until about 1100–1500 ms (depending on module), module activities do not differ at all between conditions, further underpinning the assumption that before target stimulus presentation cognitive processes do not differ between conditions. **(ii)** The most obvious deviations become apparent for the 0°/same condition. This condition shows more pronounced activity peaks than all other conditions for (1) the visual-location module, after about 1500 ms, (2) the visual module, after about 1500 ms, (3) the imaginal module, after about 1700 ms, and (4) the spatial module, with irregular deviations across the time span. **(iii)** Apart from the 0°/same condition, the most pronounced difference between conditions becomes apparent for manual module activity which shows by tendency incrementally rightward-shifted activity peaks with increasing condition difficulty.

Following our intention to extract from ACT-R module output interpretations for the processing stages that are identified by HsMM-EEG, we will now go through each of the identified stages for finding matches with phases of heightened module activities.

Stage 1: As shown in 3.2, stage 1 has nearly the same duration across all conditions. During this time span, there is high, albeit decreasing visual-location module activity. Also, there is constant, very high visual module activity.

Stage 2: As shown in 3.2, stage 2 also has nearly the same duration across all conditions. During this stage, there is only very little visual-location module activity. There is only a very short time period of visual module activity at the beginning of this stage. Noticeable, the global activity peak of the retrieval module falls within this stage. Imaginal module activity is slowly rising at the end of this stage.

Stage 3: This stage differs substantially in length between conditions. For all conditions, until 1000 ms, stage 3 matches increasing visual-location module activity, decreasing retrieval module activity, and peaking but then decreasing imaginal module activity. With increasing duration of stage 3, parts of the following patterns match with this stage after 1000 ms: a short phase of visual-location module activity which is then falling off, another visual module activity peak, a very low imaginal module activity peak, a first spatial module activity peak, no manual module activity (since the manual module activity peak around 1500 ms is only expressed for the non-rotated conditions, and in these conditions this time phase does not fall into stage 3).

Stage 4: Since for this stage both the time points of beginning and end differ substantially between conditions, a differentiated visual comparison with module activities becomes quite challenging. Therefore, we restrict the comparison to the most obvious patterns here. During stage 4, there is very little or no visual module activity, except for the non-rotated conditions. The spatial module shows pronounced activity with alternating peaks and troughs during stage 4 for all conditions. Manual module activity is rising up to a peak in the non-rotated conditions and the 50°/same condition, and to a smaller extent in the 50°/mirrored condition.

Stage 4b: Both non-rotated conditions show peaks of visual-location and visual module activity roughly matching with stage 4b, where both peaks are slightly more (visual-location) or much more (visual) pronounced in the 0°/same condition. Also, for the 0°/same condition, a very slight peak in spatial module activity roughly falls into stage 4b. The 0°/mirrored condition also expresses a pattern of fluctuating spatial module activity during this time phase. In both conditions, stage 4b is also characterised by considerable manual module activity.

Stage 5: For the 0°/same condition, there is a pronounced peak of visual-location module activity during stage 5. The 0°/same condition shows a pronounced imaginal module activity during stage 5. This peak is also present but more flattened out in the 0°/mirrored condition. There is considerable spatial module activity for all conditions except the 0°/same condition (absent) and the 50°/same condition (very low). There is rising, very high or even peaking (50°/same condition) manual module activity for all rotated conditions.

Stage 6: The 50°/same condition shows a peak of visual-location module activity during stage 6. Visual module activity is peaking in the 0°/same condition during stage 6. Imaginal module activity is falling off in the 0°/same condition and increasing in the 50°/same condition. Spatial module activity is oscillating on a high level for all conditions except 0°/same and 50°/same. Manual module activity is increasing or peaking for all rotated conditions.

Fig. 10 shows the activity of the manual module separated by experimental conditions. For the "same" conditions there is an activity peak which is delayed with increasing rotation disparity. For each of the "mirrored" conditions there are multiple smaller activity peaks and activity exhibits an oscillating pattern across the time span. At the beginning of the activity peaks of the non-rotated conditions, there is a concentrated usage of the "match" and "mismatch" shortcuts by the ACT-R model during the wholesale strategy, as indicated by the notches at the bottom line. These shortcuts also occur shortly before the additional stage 4b in the non-rotated conditions. It becomes apparent that in the non-rotated conditions manual module activity is not only high but its peaks roughly match the time phase of stage 4b. Additionally, later "mismatch" shortcuts during the piecemeal strategy imply late visual shortcuts for both 100° and 150° mismatch conditions.

3.4.2. *Chi-square test*

A chi-square test of independence was conducted to gauge the relation between ACT-R's module activity output and the HsMM's cognitive stages, that is if the pattern of active modules is distinctive for each cognitive stage. A strong relation was found and was highly significant ($\chi^2(42, 1312524) = 887404$, $p < .001$). Cramér's V shows a moderate effect between the two factors ($\varphi_c = 0.336$, 95% CI [0.335, 0.3364]). Post-hoc tests on the predictive quality of module activity for each

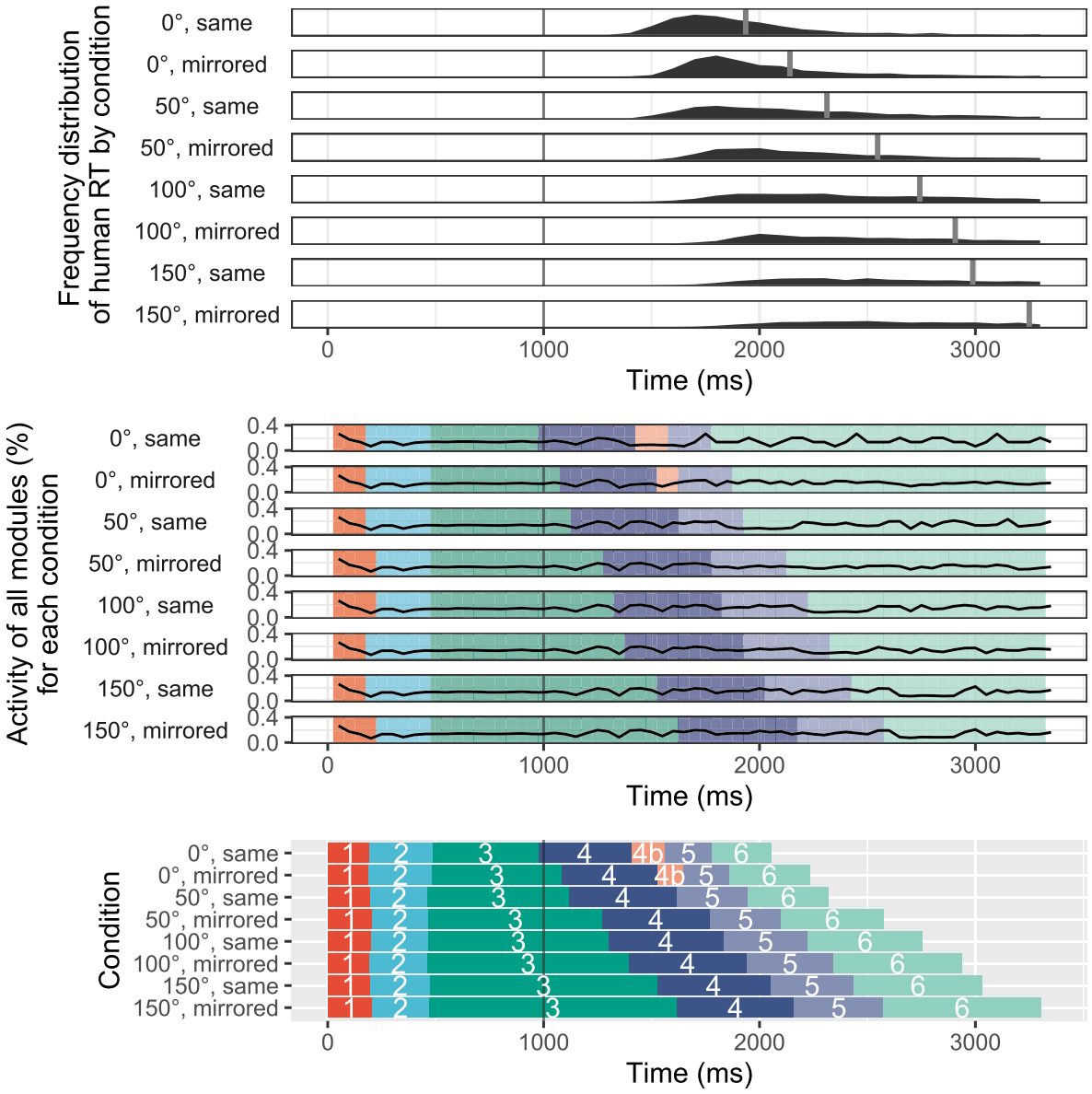


Fig. 8. Comparison of experiment RT (*top*), aggregated module activity (*middle*) and HsMM stages (*bottom*) over the course of a trial, separated by all 8 possible conditions. Vertical line at 1000 ms marks target stimulus appearance. For orientation, colour of individual HsMM stages is used as background for module activity, oriented by trial time. Note the additional stage for the non-rotated conditions (0°).

individual stage were highly significant for all (**Stage 1**: $\chi^2(7, 131798) = 182331$, $p < .001$, **Stage 2**: $\chi^2(7, 135868) = 159273$, $p < .001$, **Stage 3**: $\chi^2(7, 412911) = 400271$, $p < .001$, **Stage 4**: $\chi^2(7, 245503) = 464435$, $p < .001$, **Stage 4b**: $\chi^2(7, 9407) = 17703$, $p < .001$, **Stage 5**: $\chi^2(7, 130100) = 200694$, $p < .001$, **Stage 6**: $\chi^2(7, 246937) = 191985$, $p < .001$).

Although chi-square tests with large N have a well-known bias towards high significance, these results nonetheless show a non-random weighting of simulated activity and therefore a distinct module activity pattern for each stage indicated by standardised residual errors (see Fig. 11). More explicitly, stage 1 is mainly predicted by the goal, visual-location, and visual modules (166.37; 93.7; 323.69), stage 2 by the retrieval module (495.62), stage 3 by the visual-location and imaginal modules (300.75; 218.05), stage 4 by the spatial and production[7] modules (264.47; 107.51), stage 4b by the manual module (183.8), stage 5 by the spatial and manual modules (223.41; 135.06), and stage 6 by the spatial and manual modules as well (albeit with reversed weighting: 106.36; 281.32).

3.5. Summary

The experiment replicated main effects seen e.g. in Shepard and Metzler (1971), namely linear increase of RT with rotation disparity and learning over the course of the experiment. The effects of the experiment were in turn closely replicated by the cognitive model regarding RTs as well as effects of rotation and experiment block. An investigation of ACT-R module activity separated by experimental conditions revealed insightful differences between conditions. Comparing HsMM-EEG stages and ACT-R module outputs visually, some stage-specific activity patterns become apparent across all experimental conditions. However, for some HsMM-EEG stages there are differences in corresponding ACT-R module activity between experimental conditions. Finally, a

[7] The production module manages information exchange between all other ACT-R modules, comparable to the basal ganglia's function in the brain.

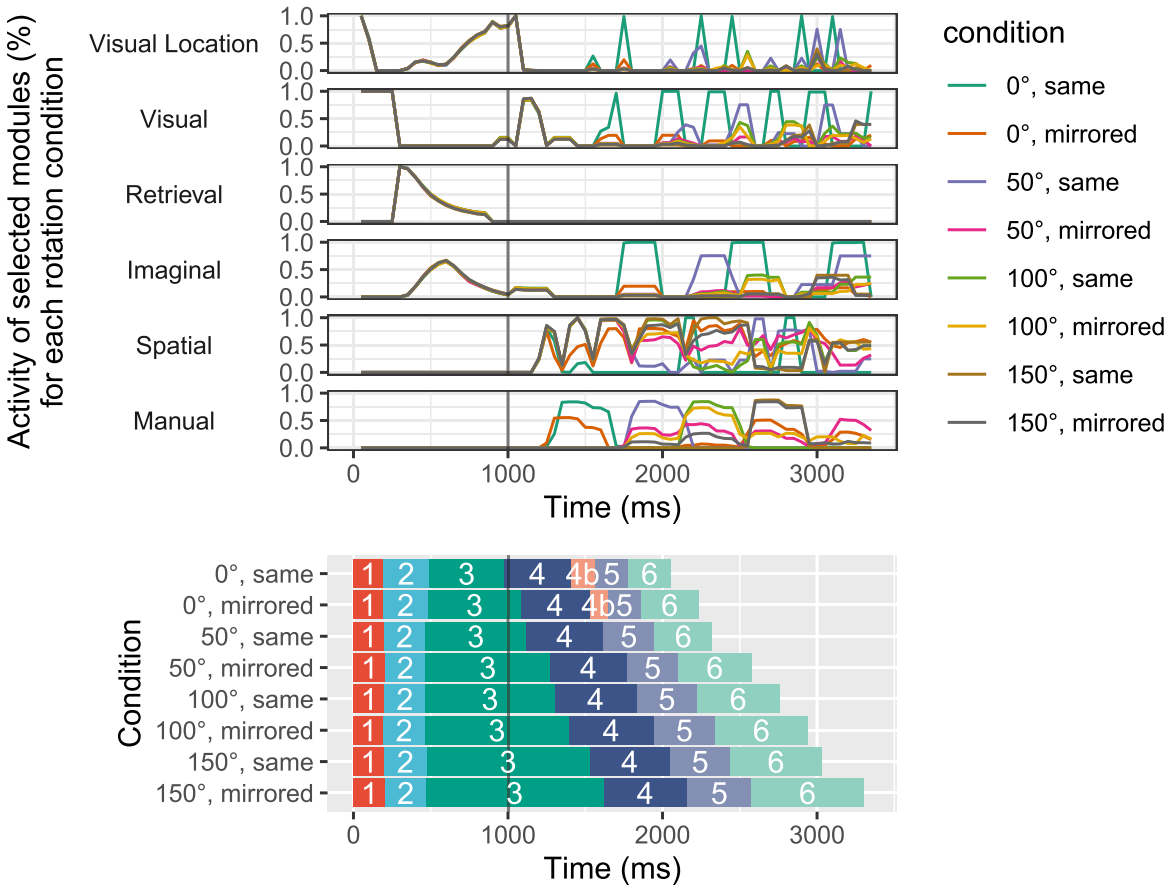


Fig. 9. Comparison of selected module activity (*top, in descending order: visual location, visual, retrieval, imaginal, spatial and manual module*) and HsMM stages (*bottom*) over the course of a trial, separated by condition.

chi-square test revealed a strong relation between module activity patterns and HsMM stages, suggesting high predictiveness of one on the other.

4. Discussion

This section deals with potential interpretations of these results, especially how they relate to our introductory hypotheses and how they integrate into current cognition research, as well as an outlook on further research.

4.1. Results discussion

4.1.1. HsMM-EEG

HsMM-EEG results for the mental rotation task reveal seven processing stages for non-rotated conditions, but six stages for rotated conditions. The brain signatures resulting from the HsMM-EEG method indicate that the extra stage 4b in the non-rotated conditions is an additional stage while all other stages seem to match between conditions. This points to an additional process happening in non-rotated conditions, which is, while being counter-intuitive, a valuable pointer with regards to different requirements in non-rotated compared to rotated mental rotation tasks. Further, the HsMM-EEG method revealed that stages 3, 4, 5, and 6 increase significantly in duration with increasing rotation disparity. Stage 6 also increases significantly in duration in mirrored compared to same conditions. Consequently, stages 3, 4, 5, and 6 can be ascribed functional relevance for the mental rotation process, while stages 1 and 2 are likely to contain processes that do not differ between conditions.

4.1.2. Cognitive model

To adequately compare the mental rotation cognitive model to EEG data, a high fit to behavioural experiment data is of high importance: meaningful comparisons of model output to cortical events can only reach high fidelity if the two time series match closely. This requirement was reached as seen in Fig. 7 and substantiated statistically in Section 3.3. The linear increase of RT for rotation disparity, as well as the general decrease of RT over experiment blocks put forth by the model, mirror the experiment results well as shown by high correlation and low deviation, lending credence to the validity of the postulated process model. A good fit to increasing RTs over conditions validates the concept of a dedicated cognitive module for mental spatial transformations, while the learning effect is explained competently by the instance-based strategy choice and utility learning mechanisms.

4.1.3. HsMM-EEG – cognitive module activity comparison

Visual comparison – Until about 1000 ms, when the target stimulus is presented, module activities do not differ between conditions. In all conditions, stages 1, 2, and a part of stage 3 (in condition 0°/same: the entire stage 3) fall within this time span. **Stage 1** is dominated by visual-location and visual module activity. **Stage 2** is instead mainly characterised by retrieval module activity. These visual interpretations mark clear functional differences between stage 1 and 2, which allows much higher differentiation than it would be possible without ACT-R module activities. Apparently, during stage 1, visual processes dominate, while stage 2 contains mainly memory retrieval processes. Regarding **stage 3**, in the HsMM-EEG results there surprisingly is no boundary between stages at 1000 ms for most conditions, which marks the appearance of the target stimulus. Still, the ACT-R module activity provides

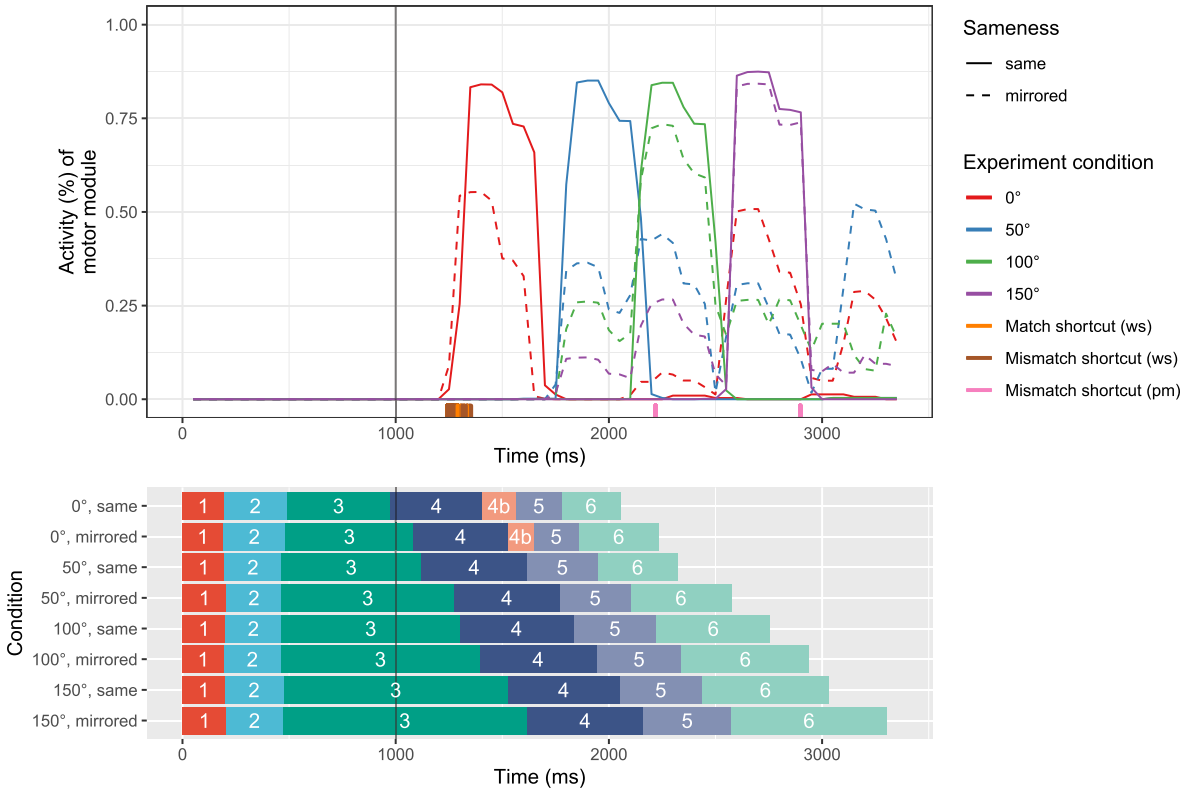


Fig. 10. Comparison of manual module activity (*top*) and HsMM stages (*bottom*) over the course of a trial, separated by condition. Notches underneath module activity mark shortcut usage by the cognitive model.

information pointing to sub-stages within stage 3: Until about 1000 ms, the stage is dominated by visual-location, retrieval, and imaginal processes. After 1000 ms, there is a peak in visual module activity and, for those conditions with the longest durations of stage 3, spatial module activity. This suggests that stage 3 is a mental image building stage which, after target stimulus presentation, becomes a visually guided rotation stage. Alternatively, this effect could also point towards a prolonged mental image building stage with no differentiation between stimuli, thereby negating the cognitive model predictions. **Stage 4** is characterised by spatial module activity, but with high visual and manual activity for the non-rotated conditions. This points to stage 4 being a rotation stage with an increased visual emphasis in the non-rotated conditions and which contains some manual preparation in the non-rotated conditions. **Stage 4b** appears to contain mostly visual and motor processes for both non-rotated conditions with some spatial processes involved. While its function does not become apparent through module activity, additional markers produced for the cognitive model's non-rotation shortcuts suggest this stage to mark visual bypassing of mental spatial transformation for simple comparisons (see Fig. 10). **Stage 5** matches with spatial and imaginal module processes, with the latter by tendency being more pronounced with increasing rotation disparity. Interestingly, the non-rotated conditions differ from the remaining conditions in that for these conditions, there appear to be additional visual processes, and that manual activity mostly culminates for them. This points to stage 5, possibly similar to stage 4, being a rotation stage with a different pattern in conditions 0°/same and 0°/mirrored where concluding visual processes play an important role and the manual response had already been executed. **Stage 6** is challenging to interpret. For this stage, we found visual-location, visual, imaginal, and manual module activity for some conditions, the latter providing peaks for both 150° conditions. Spatial module activity could be found for all conditions. A possible interpretation is that during stage 6 mental rotation continues, with different degrees of visual and manual processes being involved, pointing towards a late response stage.

Critically, considering visual inspection of the HsMM-EEG results, the brain signatures obtained for stages 4 and 5 in both the non-rotated as well as in the rotated conditions cannot be ruled out to contain eye movement patterns. During EEG data preprocessing, independent components reflecting eye movements were excluded, still, visual identification of such components always bears some risk of uncertainty. However, we argue that for our approach it is not crucial whether the brain signatures purely reflect brain activity since in the end, by applying the HsMM-EEG method to mental rotation and by bringing together HsMM-EEG with cognitive modelling, we want to enhance the understanding of the cognitive process as a whole. Comparison with the ACT-R module output shows that stages 4 and 5 are indeed characterised by spatial module activity, which is likely to correlate with some degree of eye movement. Careful conduction of EEG data cleaning appears to be especially important for methods such as HsMM-EEG which is based on weighting EEG data points (in this case, the 'snapshots') according to some criteria. Remember that for brain signature creation, snapshots are weighted according to their probability to belong to a certain stage. This is to be taken into careful consideration in further studies. Still, we believe that our findings regarding stage interpretations apply regardless of the exact source of the physiological data.

Statistical comparison – The chi-square test results serve as a proof-of-concept for the statistical comparison of the two data sets: processing stages as predicted by HsMM-EEG show a defining pattern of module activity as predicted by the cognitive model. Vice versa, specific modules seem to imply the onset of specific stages. Especially both visual modules, the retrieval module, the imaginal module, the spatial module, and the manual module display a high selectivity, as do stages 1, 2, 3, 4 and 5. In addition, high activity of the production module especially during stage 4 points towards rapid processing of short transformative processes.

Conclusion – To the best of our knowledge, this seems to mark the

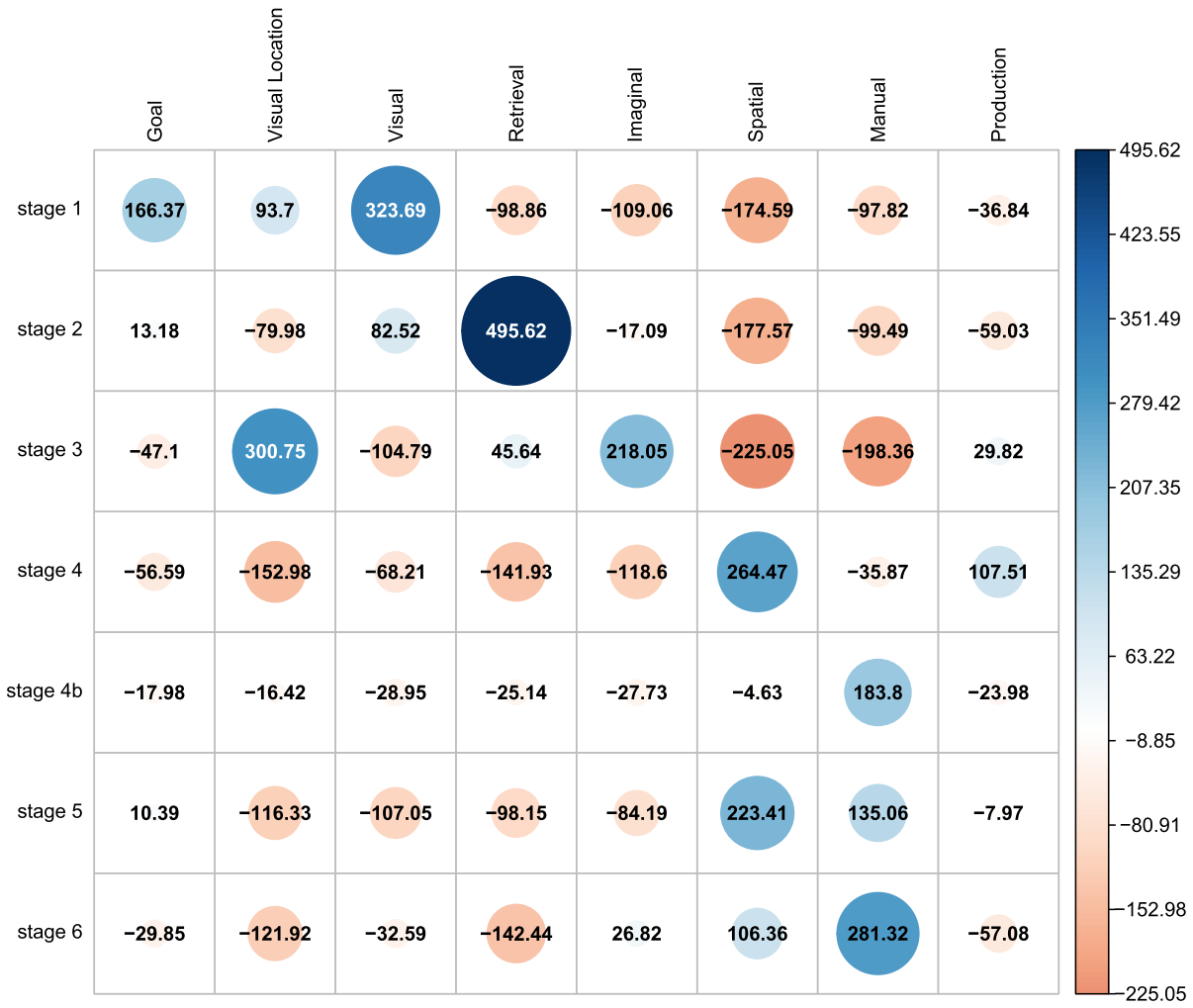


Fig. 11. Standardised residuals for each possible module-stage combination as predicted by the chi-square test. Columns ordered by presumed order of activity during a trial.

first time that patterns of module activity can be clearly matched to temporally distinct cognitive processing stages. On the basis of visual and statistical comparison an interpretation of these stages becomes evident:

- a **visual search and encoding stage** (*stage 1*),
- an **instance retrieval** or **familiarity stage** (*stage 2*),
- a **mental image forming stage** (*stage 3*),
- a **mental spatial transformation stage** (*stage 4*),
- an optional **"shortcut" stage** (*stage 4b*),
- a **response initialisation stage** (*stage 5*) and
- a **motor response stage** (*stage 6*).

This extends prior neuropsychological research on cognitive processing stages (e.g. Just and Carpenter, 1976) quite well and differentiates the functional processing carried out in high detail.

4.2. General discussion

In the following we discuss whether our hypotheses have been confirmed by our analysis.

Hypothesis 1. The HsMM-EEG method is a powerful tool to extract the number, sequence, and duration of processing stages during a cognitive process from EEG data. Indeed, the application of the HsMM-EEG method to EEG data from a mental rotation task resulted in meaningful likelihood estimates describing the most likely number and order of stages. Clear-cut results were found for average durations of these stages. Still, HsMM-EEG does not offer means to deal with the methodological challenge of temporal variability between trials and subjects. It remains challenging for HsMM-EEG methods to overcome this hurdle since possible solutions like onset-locked averaging or event-locked averaging, as proposed by Anderson et al. (2008a), would require additional steps in preprocessing of the EEG data, like averaging and linear detrending, that seem to be incompatible with the algorithms used in this study, especially with the snapshot creation that is required for HsMM-EEG estimation. Altogether, these HsMM-EEG results extend the findings of previous applications of stage-identifying methods on different tasks, e.g. Walsh et al. (2017), Zhang et al. (2017), Anderson et al. (2018), Portoles et al. (2018), Zhang et al. (2018)a, Zhang et al. (2018b), and Berberyan et al. (2020).

Hypothesis 2. ACT-R models provide a useful source of information about cognitive processes that can guide the semantic interpretation of these processing stages. The cognitive model was able to produce differentiated module activity during trials for each condition. As a high level of model fit to behavioural data was reached, theories underlying its design were validated. The time series produced by the model thus provided an overall good fit to stages predicted by the HsMM-EEG method. A comparison between the two data sets revealed not only visual similarities, but statistically verifiable patterns of module activity distinct for each processing stage.

Hypothesis 2a. Combining data-driven HsMM-EEG and theory-driven ACT-R increases interpretability of processing stages. These dependent patterns in turn add semantic meaning to stages otherwise only discernible in distinct neuronal firing patterns: by adding symbolic information to intra-trial phases, the HsMM stages now offer a definite statement on the time span of semantic cognitive processing stages during task solving. This solidifies our claim that linking the HsMM-EEG method with ACT-R models extends the interpretability of HsMM-EEG results.

Hypothesis 2b. By bringing together the HsMM-EEG method with an ACT-R model of mental rotation, further insights into cognitive processing stages are possible for mental rotation task solving. We successfully confirmed this hypothesis. Our results expand on research in cognitive processing during mental rotation tasks, e.g. Just and Carpenter (1976) or Yuille and Steiger (1982). Applying ACT-R's module information on EEG-derived HsMM stages, we were able to extract an additional layer of interpretability not available through cognitive modelling or EEG analysis alone. Additionally, we provided an indication of shortcut use, potentially explaining the existence of a stage 4b for non-rotated experiment conditions. A more differentiated analysis could subsequently shed light on a few aspects still remaining ambiguous.

The linearity of the increase in RTs dependent on increased rotation between the presented stimuli is a commonly found effect in mental rotation research. Interestingly enough, our results seem to indicate a more intricate process: while the spatial transformation stage (*stage 4*) does increase in duration for more difficult conditions, the mental image forming stage (*stage 3*) increases even more so. One possible interpretation would ascribe a combined mental imagery and spatial process to stage 3 instead, although the distinctiveness of stages inherent in HsMM-EEG speaks against that. Another interpretation would put emphasis on the preparation of an appropriate mental representation for the success of a trial. Further differentiation of these stages is necessary for a more exact answer.

Hypothesis 2c. Common features are apparent visually or statistically in both data sets, and our conclusions are applicable to spatial cognitive processing in general. It is as of yet unclear to what degree our results are generalizable, i.e. applicable to experimental paradigms other than mental rotation. The results presented here do nevertheless imply a task solving sequence commonly found in neurocognitive research: visual encoding, reasoning and (motoric) response (e.g. Just and Carpenter, 1976). Additional research will be needed to discern this in detail. Our results could therefore prove useful as a guide towards comparing model data with brain-imaging data like EEG, outlining a sensible approach to detect common features or notable differences.

While the application of our approach to a mental rotation task yielded satisfying results, it should be noted that after only a single use of the methodology, its falsifiability is not yet proven, i.e. that HsMM-EEG semantically enriched by a cognitive model would reach plausible, but ultimately false conclusions. This concern is however alleviated by the good, but imperfect fit of the method displayed in this paper: not every attribute of behavioural data was explained, instead explanations were provided for prominent features only. By avoiding overfitting of the combined approach on a specific task, its viability for other neuropsychological experimental paradigms is highly probable (see Anderson et al. (2008a) for a similar argument).

The approach used in this paper establishes a method of analysis that allows for a more intricate connectivity between ACT-R predictions and EEG data. Notably, HsMM-EEG adds a clear-cut temporal dissection to cognitive processes that is not offered by the temporal continuity of neither raw EEG data nor ACT-R module output and which substantially facilitates a temporally structured comparison. In addition, HsMM-EEG validates the cognitive architecture's modular concept even further by successfully allocating specific modules to time periods suggested by HsMM-EEG, thereby externally verifying it. By bringing together the HsMM-EEG method with an ACT-R model, we demonstrated a way to extract a high amount of information from trial-wise EEG data, extending established methods of EEG data analysis that focus on a fixed period within an EEG data segment and that do not allow for an examination of qualitatively different stages of information processing within one analysis. Beyond that, our approach demonstrated how the lacking interpretability of HsMM-EEG results can be alleviated by adding semantic information from a cognitive model, clearly extending previous research based on neurostatistical stage-identifying methods, e.g. MVPA-HMM (Anderson and Fincham, 2014a), HsMM-MVPA (Anderson et al., 2016) or standalone HsMM-EEG (Borst and Anderson, 2015). Consequently, we proved that by linking HsMM-EEG as a data-driven method on one hand and ACT-R modelling as a theory-driven method on the other hand, clear and interpretable stages of cognitive processing can be provided.

4.3. Open questions and outlook

The original source of the EEG data used in this study were a master's and a bachelor's thesis, addressing an altogether different problem, with no intention of further post-processing. Given that, the data was optimised neither for HsMM-EEG nor for cognitive model comparison. More bespoke data could therefore improve on the application of both methods. Relying on a maximum likelihood estimation approach, HsMM-EEG is a stochastic tool, i.e., its results always come with a quite high degree of uncertainty. As was described, the LOOCV procedure results in the most likely minimum number of HsMM states, meaning that the identified processing stages could possibly be further subdivided, at least for some subjects. Note that by relying on sign test results for significant numbers of HsMM states, we ruled out higher state solutions that did not become significant but that still were the most likely solutions for a number of subjects. In general, since the LOOCV procedure requires a number of subject data sets in order to produce stable results, HsMM-EEG is not well suited for analyses of inter-individual differences. For our approach, stage numbers and durations were always averaged across subjects and trials, making more differentiated insights into inter-individual differences or individual strategies difficult. The retrieval stage that the ACT-R model pointed at brought out the relevance of memory retrieval processes for task paradigms relying on a limited pool of stimuli that are each presented multiple times. Our HsMM-EEG analysis was not sensitive enough for these possible differences since EEG data was not segregated into different experimental blocks.

Over the course of analysis, shortcomings of the cognitive model (and by extension, ACT-R) became evident. All modules except the manual module show little to no differentiation along the time axis between conditions. At times, aggregated module activity shows 100% strength for a given time, meaning that the same module fired at the exact same time in every single trial. These effects should be considered unintended artefacts of ACT-R – its predictions are not entirely stochastic and often occur in quantified, discrete time steps; at times, this can lead to an extensive overlap of activity during trials. The model itself on the other hand relies on a spatial module outlined above for the eponymous mental rotation. While this module aims to represent a universally valid approach to mental spatial transformation, the mental rotation task is to date the only paradigm fully tested with the spatial module. This entails that as of now, it remains specialised to the task at hand and is kept relatively simple. Still, we maintain that it offers a sound, reliable method of introducing spatial capabilities into a symbolic architecture like ACT-R. Additional studies with other paradigms of spatial cognition aim to improve on this claim. The current model fit is close, but still leaves room for improvement – for instance, the model starts out much slower than human subjects in the first block, and becomes increasingly deterministic. These differences seem to be negligible for our purposes, alternative approaches modelling human mental

rotation behaviour could potentially improve on this nevertheless.

The retrieval stage pointed out by the ACT-R model brought out the relevance of memory retrieval processes for task paradigms relying on a limited pool of stimuli, each being presented multiple times. Further studies should apply a stage-identifying method to an EEG data set that is not only structured by conditions but also segregated into different experimental blocks. Thereby, it could be investigated whether the retrieval stage changes in duration over the course of the experiment.

In the future, we plan to apply the approach outlined in this paper to other experimental paradigms. Currently, an application to a mental folding study is underway (extending and updating Preuss et al., 2019), including a cognitive model that also incorporates the spatial module presented here as well as new HsMM-EEG models. Since mental folding requires spatial reasoning in a similar fashion to mental rotation, those results will be especially interesting regarding differences and similarities in cognitive processing.

Non-spatial tasks could also benefit from this approach. Certain requirements for this method did however become apparent: the task needs to be separable into distinct phases, preferably relying on a wide array of cognitive functions; it should be easily generalizable across subjects and not be highly individualised; it should offer a sufficient amount of experimental conditions; and it should ideally provide several behavioural measures.

Whereas comparing cognitive model output to HsMM-EEG stages can further inform and infer data, it can inversely also serve as an advanced option for model fitting, with an emphasis on intra-trial model behaviour. Crucially, HsMM stages with established cognitive correlates could subsequently serve as benchmarks for the model to approximate in time frame and function, going beyond fitting models to end-of-trial RT to verify accurate reproduction of cognitive stages.

In parallel to the approach outlined herein, the cognitive model is currently analysed in relation to cerebral clusters resulting from an Independent Cluster Analysis (ICA) on the same EEG data set. Contrary to HsMM models that show cognitive processing over time, ICA yields a spatial differentiation, meaning module activity can then be assigned to functional structures of the brain. This promises a validation of both module usage with already documented functions of certain clusters, as well as our theories on mental spatial processing.

While we cannot make the claim that our results are generalizable to overall cognition, i.e. that our proposed cognitive stages are universally valid across experimental task paradigms, they do however offer an intriguing cornerstone for further research: with the exception of the spatial transformative stage, the processing stages do seem to entail basic cognitive functions: visual search and encoding, memory retrieval, mental imagery, visual comparison and motor response. Yet even so, spatial transformative processing could be a task-specific instance of a general reasoning stage detectable across task paradigms. The possibility that generalizable cognitive processing stages could after all be reflected in a wide array of neuropsychological task environments appears very likely. This paper provides a clear outline for applying EEG monitoring and cognitive modelling to reach results not accessible by either method alone. Exploring facets of complex cognition seems especially viable using this methodology. We hope these results serve to facilitate similar approaches in other fields of neuropsychology and cognition research.

Credit author statement

Linda Heimisch and Kai Preuss contributed equally in: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Writing, Visualization. Author Nele Russwinkel contributed in: Conceptualization, Writing, Supervision.

Funding

This work was funded by the German Research Foundation (DFG) as part of project #396560184 (https://gepris.dfg.de/gepris/projekt/396560184).

Declaration of competing interest

None.

Data availability

See appendix.

Acknowledgements

The authors would like to thank Jelmer Borst, Christopher Hilton, and Eike Richter for their support.

Appendix A. Available data and software

- Original behavioural and EEG data ()(Hilton et al., 2022): https://doi.org/10.17605/osf.io/gfy5j
- HsMM-EEG data and analysis: https://doi.org/10.14279/depositonce-16143
- Cognitive model of mental rotation: https://doi.org/10.14279/depositonce-15701
- Spatial module for ACT-R: https://doi.org/10.14279/depositonce-11386
- Mental rotation model data and analysis: https://doi.org/10.14279/depositonce-17785

References

Anderson, J.R., Borst, J.P., Fincham, J.M., Ghuman, A.S., Tenison, C., Zhang, Q., 2018. The common time course of memory processes revealed. Psychol. Sci. 29 (9), 1463–1474.
Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y., 2004. An integrated theory of the mind. Psychol. Rev. 111 (4), 1036–1060.
Anderson, J.R., Carter, C.S., Fincham, J.M., Qin, Y., Ravizza, S.M., Rosenberg-Lee, M., 2008a. Using fMRI to test models of complex cognition. Cognit. Sci. 32 (8), 1323–1348.
Anderson, J.R., Fincham, J.M., 2014a. Discovering the sequential structure of thought. Cognit. Sci. 38 (2), 322–352.
Anderson, J.R., Fincham, J.M., 2014b. Extending problem-solving procedures through reflection. Cognit. Psychol. 74, 1–34.
Anderson, J.R., Fincham, J.M., Qin, Y., Stocco, A., 2008b. A central circuit of the mind. Trends Cognit. Sci. 12 (4), 136–143.
Anderson, J.R., Fincham, J.M., Schneider, D.W., Yang, J., 2012. Using brain imaging to track problem solving in a complex state space. Neuroimage 60 (1), 633–643.
Anderson, J.R., Lebiere, C.J., 2014. The Atomic Components of Thought. Psychology Press.
Anderson, J.R., Qin, Y., Sohn, M.-H., Stenger, V.A., Carter, C.S., 2003. An information-processing model of the BOLD response in symbol manipulation tasks. Psychon. Bull. Rev. 10 (2), 241–261.
Anderson, J.R., Zhang, Q., Borst, J.P., Walsh, M.M., 2016. The Discovery of Processing Stages: Extension of Sternberg's Method. Psychological Review.
Ashby, F.G., Crossley, M.J., 2010. Interactions between declarative and procedural-learning categorization systems. Neurobiol. Learn. Mem. 94 (1), 1–12.
Bell, A.J., Sejnowski, T.J., 1995. An information-maximization approach to blind separation and blind deconvolution. Neural Comput. 7 (6), 1129–1159.
Berberyan, H., van Maanen, L., van Rijn, H., Borst, J., 2020. Electroencephalogram-based identification of evidence accumulation stages in decision-making. J. Cognit. Neurosci. 33 (3), 510–527.
Bethell-Fox, C., Shepard, R., 1988. Mental rotation: effects of stimulus complexity and familiarity. J. Exp. Psychol. Hum. Percept. Perform. 14 (1), 12–23.
Borst, J.P., Anderson, J.R., 2015. The discovery of processing stages: Analyzing EEG data with hidden semi-markov models. Neuroimage 108, 60–73.
Borst, J.P., Schneider, D.W., Walsh, M.M., Anderson, J.R., 2013. Stages of processing in associative recognition: evidence from behavior, EEG, and classification. J. Cognit. Neurosci. 25 (12), 2151–2166.
Brain Products GmbH, 2014. BrainAmp (64 Channels) [Apparatus]. Gilching, Germany.
Byrne, M.D., Anderson, J.R., Douglass, S., Matessa, M., 1999. Eye tracking the visual search of click-down menus. Proceedings of the SIGCHI conference on Human Factors in Computing Systems 402–409.
Cassenti, D., Kerick, S., McDowell, K., 2011. Observing and modeling cognitive events through event-related potentials and ACT-R. Cognit. Syst. Res. 12 (1), 56–65.

Donders, F.C., 1868. Over de snelheid van psychische processen. Onderzoekingen gedaan in het Physiologisch Laboratorium der Utrechtsche Hoogeschool (1868–1869) 2, 92–120.
Funke, J., 2010. Complex problem solving: a case for complex cognition? Cognit. Process. 11 (2), 133–142.
Gonzalez, C., Lerch, J.F., Lebiere, C., 2003. Instance-based learning in dynamic decision making. Cognit. Sci. 27 (4), 591–635.
Griffiths, G., West, R., D'Anguilli, A., 2011. Cognitive modeling of event-related potentials. In: Proceedings of the Annual Meeting of the Cognitive Science Society, vol. 33 (33).
Gunzelmann, G., Lyon, D.R., 2007. Mechanisms for human spatial competence. In: Barkowsky, T., Knauff, M., Ligozat, G., Montello, D.R. (Eds.), Spatial Cognition V: Reasoning, Action, Interaction. Springer, pp. 288–307.
Harris, J., Hirsh-Pasek, K., Newcombe, N.S., 2013. Understanding spatial transformations: similarities and differences between mental rotation and mental folding. Cognit. Process. 14 (2), 105–115.
Hilton, C., Raddatz, L., Gramann, K., 2022. A general spatial transformation process? Assessing the neurophysiological evidence on the similarity of mental rotation and folding. Neuroimage: Reports 2 (2), 100092.
Jo, S., Myung, R., Yoon, D., 2012. Quantitative prediction of mental workload with the ACT-R cognitive architecture. Int. J. Ind. Ergon. 42 (4), 359–370.
Just, M.A., Carpenter, P.A., 1976. Eye fixations and cognitive processes. Cognit. Psychol. 8 (4), 441–480.
Khooshabeh, P., Hegarty, M., Shipley, T.F., 2013. Individual differences in mental rotation. Exp. Psychol. 60 (3), 164–171.
Klaproth, O.W., Halbrügge, M., Krol, L.R., Vernaleken, C., Zander, T.O., Russwinkel, N., 2020. A neuroadaptive cognitive model for dealing with uncertainty in tracing pilots' cognitive state. Topics in Cognitive Science 12 (3), 1012–1029.
Klemm, P., Russwinkel, N., Prezenski, S., 2021. A General Method for Matching Cognitive Models and EEG Processes - towards a Holistic Understanding of Cognition [Manuscript Submitted for Publication]. Technische Universität Berlin.
Laird, J.E., Lebiere, C., Rosenbloom, P.S., 2017. A standard model of the mind: toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. AI Mag. 38 (4), 13–26.
Lebiere, C., R Anderson, J., Bothell, D., 2001. Multi-tasking and cognitive workload in an ACT-R model of a simplified air traffic control task.
Marr, D., 1982. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.
Newell, A., 1990. Unified Theories of Cognition. Harvard University Press.
Peebles, D., 2019a. Modelling alternative strategies for mental rotation. In: Stewart, T. (Ed.), Proceedings of the 17th International Conference on Cognitive Modelling. University of Waterloo, pp. 138–143.
Peebles, D., 2019b. Modelling mental imagery in the ACT-R cognitive architecture. In: Goel, A., Seifert, C., Freksa, C. (Eds.), Proceedings of the 41st Annual Conference of the Cognitive Science Society. Cognitive Science Society, pp. 2550–2556.
Peters, M., Battista, C., 2008. Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. Brain Cognit. 66 (3), 260–264.
Portoles, O., Borst, J.P., van Vugt, M.K., 2018. Characterizing synchrony patterns across cognitive task stages of associative recognition memory. Eur. J. Neurosci. 48 (8), 2759–2769.
Preuss, K., Raddatz, L., Russwinkel, N., 2019. An implementation of universal spatial transformative cognition in ACT-R. In: Stewart, T.D. (Ed.), Proceedings of the 17th International Conference on Cognitive Modelling. University of Waterloo, pp. 144–150.
Preuss, K., Russwinkel, N., 2021. Cognitive modelling of a mental rotation task using a generalized spatial framework. In: Stewart, T.D. (Ed.), Proceedings of the 19th International Conference on Cognitive Modelling. University Park, PA, pp. 220–226.
Prezenski, S., Russwinkel, N., 2016. A proposed method of matching ACT-R and EEG-data. In: Reitter, D., Ritter, F. (Eds.), Proceedings of the 14th International Conference on Cognitive Modeling. University Park, pp. 249–251.
Rabiner, L.R., 1990. A tutorial on hidden markov models and selected applications in speech recognition. In: Readings in Speech Recognition. Elsevier, pp. 267–296.
Raddatz, L., 2014. *EEG Correlates of Sex Differences during Mental Rotation* [Unpublished Bachelor's Thesis]. Leuphana Universität Lüneburg.
Schischigin, K., 2014. *The Influence of Navigation Strategies on Mental Rotation: an EEG Study* [Unpublished Master's Thesis]. Technische Universität Berlin.
Shepard, R., Metzler, J., 1971. Mental rotation of three-dimensional objects. Science 171, 701–703.
Smith, E.R., 1994. Procedural knowledge and processing strategies in social cognition. Handbook of Social Cognition 1, 99–151.
Sternberg, S., 1969. The discovery of processing stages: extensions of Donders' method. Acta Psychol. 30, 276–315.
Stevens, C.A., Morris, M.B., Fisher, C.R., Myers, C., 2019. Cognitive metrics profiling of a complex task: toward convergent validity with behavioral and EEG workload indicators. In: Stewart, T.D. (Ed.), Proceedings of the 17th International Conference on Cognitive Modeling.
Stocco, A., Anderson, J.R., 2008. Endogenous control and task representation: an fMRI study in algebraic problem-solving. J. Cognit. Neurosci. 20 (7), 1300–1314.
Sudre, G., Pomerleau, D., Palatucci, M., Wehbe, L., Fyshe, A., Salmelin, R., Mitchell, T., 2012. Tracking neural coding of perceptual and semantic features of concrete nouns. Neuroimage 62, 451–463.
Tenison, C., Anderson, J.R., 2016. Modeling the distinct phases of skill acquisition. J. Exp. Psychol. Learn. Mem. Cognit. 42 (5), 749.
van Vugt, M.K., 2012. Relating ACT-R buffer activation to EEG activity during an attentional blink task. In: Russwinkel, N., Drewitz, U., van Rijn, H. (Eds.), Proceedings of the 11th International Conference on Cognitive Modelling, p. 218.
van Vugt, M.K., 2013. Towards a dynamical view of ACT-R's electrophysiological correlates. In: West, R., Stewart, T.C. (Eds.), Proceedings of the 12th International Conference on Cognitive Modelling, pp. 11–16.
van Vugt, M.K., 2014. Cognitive architectures as a tool for investigating the role of oscillatory power and coherence in cognition. Neuroimage 85, 685–693.
Walsh, M.M., Gunzelmann, G., Anderson, J.R., 2017. Relationship of P3b single-trial latencies and response times in one, two, and three-stimulus oddball tasks. Biol. Psychol. 123, 47–61.
Yuille, J.C., Steiger, J.H., 1982. Nonholistic processing in mental rotation: some suggestive evidence. Percept. Psychophys. 31 (3), 201–209.
Zhang, Q., van Vugt, M., Borst, J.P., Anderson, J.R., 2018a. Mapping working memory retrieval in space and in time: a combined electroencephalography and electrocorticography approach. Neuroimage 174, 472–484.
Zhang, Q., Walsh, M.M., Anderson, J.R., 2017. The effects of probe similarity on retrieval and comparison processes in associative recognition. J. Cognit. Neurosci. 29 (2), 352–367.
Zhang, Q., Walsh, M.M., Anderson, J.R., 2018b. The impact of inserting an additional mental process. Computational Brain & Behavior 1 (1), 22–35.