

**DOES COGNITIVE PROCESSING „ADD-UP“? ASSESSING THE ASSUMPTION OF PURE
INSERTION WITH A HIDDEN SEMI-MARKOV MODEL ANALYSIS**

Julius Kricheldorf

August 2018

Major Thesis MSc Programme Behavioral and Cognitive Neuroscience

Faculty of Science and Engineering

Supervised by: dr. Jelmer Borst and Oscar Portoles

Authors Note: I would like to thank Jelmer Borst, Oscar Portoles and Hermine Berberyan for their constant help and guidance throughout this project. I greatly appreciated the advice and support given to me in particular during the more challenging periods of the project. I have learned a tremendous amount during the past six month and while at times it was challenging, I enjoyed it every step of the way. I also want to thank my family for all their support over the years and making it possible for me to be part of this project.

CONTENTS

<u>ABSTRACT</u>	4
<u>INTRODUCTION</u>	5
METHODS TO IDENTIFY LATENT COGNITIVE PROCESSING STAGES	6
METHODS BASED ON BEHAVIORAL DATA	6
METHODS AIDED BY NEURAL/PHYSIOLOGICAL DATA	7
HIDDEN SEMI-MARKOV MODELS	8
RESEARCH QUESTION AND EXPERIMENTAL PARADIGM	10
<u>METHOD</u>	11
PARTICIPANTS	11
EXPERIMENTAL DESIGN	11
LEARNING PHASE	12
EXPERIMENTAL PHASE	12
EEG-RECORDING	13
EEG-PREPROCESSING	13
GENERAL ANALYSIS	14
ERP-ANALYSIS	14
HSMM-MVPA ANALYSIS	15
<u>RESULTS</u>	16
BEHAVIOR RESULTS	16
ERP RESULTS	19
HSMM RESULTS	20
NAMES PRESENTATION	20
EXPERIMENTAL CONDITIONS	21
PURE INSERTION	24
<u>DISCUSSION</u>	25

ASSESSING PURE INSERTION USING HSMM

IDENTIFIED STAGES 26

DISTANCE EFFECT..... 27

ASSESSING PURE INSERTION WITH HSMMs..... 28

LIMITATIONS 29

FUTURE RESEARCH 31

CONCLUSION 31

REFERENCES 32

APPENDIX 35

ABSTRACT

Many theories in cognitive psychology have the premise that information processing in the brain is facilitated in distinct cognitive stages. A controversial assumption known as “pure insertion” maintains that in a chain of processing stages one can add or remove any processing stage without having an effect on other processing stages. In this study we attempted to assess pure insertion, using EEG to record participants’ brain activity, while having them solve an information processing task. Experimental conditions of the task required the insertion of either an additional memory retrieval, the manipulation of order information, both, or neither for successful problem-solving. Effects on individual processing steps were assessed by parsing the EEG task data into processing stages with distinct neural signatures using a hidden semi Markov Model multivariate pattern analysis (HSMM-MVPA). We were successfully able to identify processing stages specific to each experimental manipulation. However, while behavioral analyses indicate pure insertion to be violated, due to low sample size we could not conclusively track this violation to a distinct processing stage. We identified an encoding-, retrieval-, as well as stages specific to manipulating the name order. Unexpectedly, the manipulation of order information turned out to contain also a memory retrieval stage. However, we were unable to consistently identify and assess effects on shared stages related to response and response preparation. Our results illustrate the validity of the HSMM-MVPA method and but also highlight the importance of careful task design and sample size considerations when using a HSMM-MVPA analysis.

Keywords: HSMM-MVPA, pure insertion, EEG

DOES COGNITIVE PROCESSING “ADD-UP”? ASSESSING THE ASSUMPTION OF PURE INSERTION WITH A HIDDEN SEMI-MARKOV MODEL ANALYSIS

An assumption shared by much of classical cognitive science either explicitly or implicitly is that information processing in the brain happens in discrete processing stages. This perspective dates back to Donders (1868; 1969). Faced with the problem that mental processes cannot be studied by observing the physiology of the brain, Donders came up with the idea to study mental phenomena by isolating individual processing steps and quantifying their processing time. Recognizing – and assuming – that complex mental processes consist of a chain of simpler processes, he devised the so-called subtraction method. This method isolates mental processes by comparing the response time (RT) of a baseline task versus the baseline task plus an additional processing stage. Using this method, Donders developed a series of tasks. For example, a task in which participants on the display of a white light had to make a response with their right hand and on the display of a red light a response with their left hand. He contrasted this task with a similar task where on display of red light the response had to be withheld (the prototype of the go/noGo task). Donders reasoned that in the conditions requiring both hands to respond, this would insert a selection and preparation process requiring participants to determine which hand was subsequently needed. Indeed, measures of reaction time showed an increase in the inserted condition, letting Donders conclude that the selection and preparation process took 154 ms.

For the results of Donders (1868; 1969) task to be reflective of cognitive processing stages according to Ulrich, Mattes and Miller (1999), at least three assumptions must hold true:

- 1) Processes in the brain must occur sequentially with each subsequent stage further processing the output of the previous process as input.
- 2) At every point in time only one process is active with each process beginning once the previous process has ended.
- 3) As a consequence of the previous assumptions one can add or remove each cognitive processing step without affecting the duration of all other processing steps – termed “pure insertion”.

While there are plausible alternative possibilities for the serial sequence models of cognitive processing (Sackur & Dehaene, 2009), these will not be addressed here (the interested reader may be referred to e.g. McClelland, 1979). For the purpose of our study we are interested in the assumption of pure insertion. In essence this assumption excludes the possibility of interaction between different cognitive processes (Friston et al., 1996). Evidence for or against the hypothesis is at present mixed. On the one hand, Ilian and Miller (1994) performed a series of experiments to assess if the assumption holds for the insertion of a mental rotation stage. However, their results suggested that insertion of a mental rotation interacts with response selection. Ulrich,

ASSESSING PURE INSERTION USING HSMM

Mattes and Miller (1999) on the other hand, assessed the effects of the insertion of an additional processing stage on the motor response stage by using response force as a dependent variable in addition to RT. At least in the context of a motor response, pure insertion appears to hold as response force was indifferent to the insertion of a response choice stage. However, as we will show, a problem of all of these studies is that they relied only on compound measures of cognitive processing (i.e. whole trial RT) or only assessed a subset of processing stages. The purpose of this study is to observe if the addition or removal of a cognitive processing stage directly affects the duration of all cognitive processing steps involved as reflected by their neural correlates.

Methods to identify latent cognitive processing stages

Methods based on behavioral data

If the assumption of pure insertion may be problematic, how can we then directly observe cognitive processes in reaction time? The problem with RT is that it is difficult to discern the contribution of individual processing stages to its total. Sternberg (1969) developed the additive factors method, which is able to relate differences in RT to individual processing stages and is not reliant on pure insertion. By observing the effects of specific experimental manipulations on RTs, one can infer the existence of independent latent processes. That means, if the effects on RT of experimental manipulation A and experimental manipulation B are additive and independent from each other, one can infer that they affect different latent processing stages. Similar to Donders (1868; 1969) subtraction method, Sternberg's (1969) method depends on three assumptions to hold true:

- 1) There are successive processing stages (Sternberg himself used the more nuanced term “functional components”) between stimulus and response whose durations are additive.
- 2) The RT distributions of these stages are stochastically independent.
- 3) The RT distributions of these stages are known (i.e. it is known if they are normally distributed, exponentially distributed etc.).

An advantage of this method is that it does not require the assumption of pure insertion. A disadvantage of the method is that it only allows to infer a lower-bound limit of the estimation of all cognitive processing stages present in a task. For every latent process, whose presence is to be determined, at least one experimental manipulation is required to confirm its existence. This is in particular a problem for tasks that are more complex. Further, a processing stage can only be identified if it can be experimentally manipulated.

Another problem that Sternberg, 1969 acknowledged himself is that the additive factors method does not necessarily allow to infer the presence of an underlying cognitive processing stage (Stafford & Gurney, 2011). According to Stafford and Gurney (2011), instead of inferring processing stages at the implementational

ASSESSING PURE INSERTION USING HSMM

level, the additive factor method at most would only allow discovery of processing stages at the functional level. Therefore, with the additive factors method we are able to isolate functional architectural stages. However, it does not follow that they are implemented in a strictly serial manner. Stafford and Gurney (2011) used a discrete two-stage model and a continuous processing model, which conceptually consists of a single "stage" to model reaction time results of a Stroop-task. In the Stroop-task stimulus intensity (word- and color to the same degree) and word-color congruence were varied producing additive effects on reaction time and thus indicating at least two latent processing stages. However, RT data produced by the two models in a simulation were both in agreement with experimental data. Thus, while an additive factors analysis predicted at least two stages, a single stage model was able to account for the data. The underlying problem of Sternberg's and RT-based methods in general, is that they do not allow to make inferences on individual stages. They only allow to assess the cumulative effect of all cognitive processing stages involved. Hence, they rely on whole-trial RTs (Anderson, Zhang, Borst, & Walsh, 2016).

Methods aided by neural/physiological data

As a consequence, we may have to accept that behavioral data alone are insufficient. However, unlike Donders (1868; 1969) we are able to relate brain activity to cognitive processing. Hawkins, Mittner, Forstmann and Heathcote (2017) outlined the importance of including neural data to identify latent cognitive stages. To assess the assumption of pure insertion, EEG measures are to be preferred over fMRI measures, as fMRI is restricted to relatively long processes due to its low temporal resolution (Anderson et al., 2016). Augmenting analysis of Donders (1868; 1969) original tasks with EEG data led Smid, Fiedler and Heinze (2000) to conclude that the inserted choice stage used did not adhere to the assumption of pure insertion. First, Smid et al. (2000) argued what compared to the previously introduced go/noGo baseline task by Donders (1868; 1969) was considered to be an inserted decision-selection stage (a known response in the go/noGo condition vs. an unknown response in the choice condition) was confounded by response preparation (preparing for a known response vs. preparing for a response that is unknown). Moreover, they argued that the difference in task requirement may affect not only motor preparation, but already encoding of the relevant stimulus information. To correct for this confound, they adjusted the tasks by making both tasks a go/noGo paradigm. The decision selection stage was inserted by requiring two distinct responses for two stimuli, which required a Go-response versus only one response for both stimuli in the baseline condition. They used EMG (electromyogram) measures and RT to quantify the motor response and event-related potentials of EEG recordings to quantify response selection and response preparation (lateralized readiness potential), but also stimulus identification (selection-negativity). EMG and ERP results indicated that increase in RT of the choice tasks was at least partially due to

ASSESSING PURE INSERTION USING HSMM

advance preparation of motor responses as indicated by EMG and LRP amplitudes. Smid et al. (2000) consequently concluded that the insertion of the choice stage is not pure, also because a go/noGo task already includes a choice stage. This demonstrates how neural data can provide information about individual processing stages by which pure insertion can be tested.

While the addition of ERPs as neural covariates of cognitive processing stages provides an additional tool to assess the assumption of pure insertion, it is not without drawbacks. ERP results alone are insufficient as markers of cognitive processing. In order to provide information about ordering, onset, duration and functionality of cognitive processing it requires converging experimental evidence (Hillyard & Kutas, 1983). With regard to the study by Smid et al. (2000), for instance, the selection negativity (SN) ERP intended to measure stimulus identification is only an indirect measure of that processing stage. The authors themselves acknowledged that it only starts after the stimulus has been identified. Further, ERPs do not contain any information with regard to the distribution of duration/onset/offset of the cognitive process of interest.

Hidden semi-Markov models

A method capable of addressing many of these problems is the so called hidden semi-Markov model multivariate pattern analysis (HSMM-MVPA – from here on referred to as HSMM) by Anderson and colleagues (2016). It uses hidden Markov models, which describe a system that is always in one of a particular set of states, which cannot be observed directly (Rabiner, 1989). State probabilities can be estimated by outputs generated by the hidden state. The "semi-" distinction here is added indicating that the duration of the state-durations is not fixed, but allowed to vary in every sequence; these HMMs are also known as variable-duration HMMs (Rabiner, 1989).

Applied to neural data the method is supposed to identify stages of cognitive processing by identifying significant cognitive events present in the data. Originally, the HSMM analysis was applied to fMRI data (Anderson, Fincham, Schneider, & Yang, 2012), where significant cognitive events for every trial were identified as periods of sustained hemodynamic response. Sustained activity indicated a processing stage. Different processing stages then could be distinguished by different patterns of sustained activity. A drawback of using fMRI is its low temporal resolution which only allows the identification of cognitive processing stages that last relatively long (10+ seconds, (Anderson et al., 2016). In order to identify stages in the millisecond range, Borst and Anderson (2015) applied the method to EEG data. Similar to the fMRI version, significant cognitive events were identified based on constant patterns of EEG data. Since then the method has been further refined.

The classical theory of ERP generation suggests that significant cognitive events produce brief bursts in activity that add on top of uncorrelated sinusoidal activity (Shah et al., 2004). When averaging large amounts of

ASSESSING PURE INSERTION USING HSMM

trials, this added phasic activity (if not jittered) produces the typical deflections observed in ERPs. Thus, instead of trying to identify periods of sustained activity, Anderson et al. (2016) modelled transitions between stages as so called “bumps”, half-sinusoidal multidimensional peaks lasting 50ms. In the HSMM context, such a “bump” marks the onset of a new stage and the end of the previous one. Variable stage durations are modelled as “flats”, the return of the ongoing signal to sinusoidal noise with mean zero. Flat durations are modelled as gamma distributions with a shape parameter of two and a scale parameter to be estimated by the HSMM.

The HSMM method by Anderson et al. (2016) estimates these parameters by estimating “bump” latency and amplitude for every trial, maximizing likelihood of the data. The parameter estimation is limited by the number of bumps. For a n -bump model, every possibility that a trial might be divided into $n+1$ stages is estimated and the model that maximizes the log-likelihood across all trials is selected. To address overfitting and avoid the addition of meaningless “bumps” that only inflate model fit, leave-one-out-cross validation (LOOCV) is used. In order to justify an additional bump, a sign-test is used to determine if an additional bump increases model-fit increase is consistent for all participants. Furthermore, the number of bumps that can be fitted is limited by the minimum trial duration. Hence, it is not possible to estimate more than 10 bumps in a trial that only lasts 500ms. The same is true for the flat durations, which are constrained by trial length and the number of bumps. While the method itself is data-driven, it also allows for top down-constrain. For example, Anderson et al. (2016) used ACT-R stage duration estimates to determine the shape parameter of the gamma estimates, and the total number of states.

HSMM analysis has been successfully used to extract cognitive processing stages in several studies (Portoles, Borst, & van Vugt, 2017; e.g. Zhang, Walsh, & Anderson, 2017; Zhang, van Vugt, Borst, & Anderson, 2018; Zhang, Walsh, & Anderson, 2018). For example, Anderson et al. (2016) used a HSMM to analyze data of a recognition-memory task, where the number of episodic associations with a word was experimentally manipulated. That is, the more associations an item had, the lower were the accuracy in a recognition memory task and the longer were the reaction time. Fitting a HSMM to a recognition-memory task identified a five “bump” – six stage model to fit the trial data best. The identified HSMM was in good agreement with cognitive theories of the task, roughly including an encoding stage, a retrieval stage, a decision stage and a response stage. The stages identified by the HSMM allowed to localize the effect of experimental manipulation in the variability in duration of stages associated with retrieval and decision, but not the encoding and response. In sum, a HSMM analysis allows to isolate signatures of cognitive processing stages and relate differences in overall RT back to underlying differences in individual cognitive processing stages.

Research question and experimental paradigm

In this study we use a HSMM analysis to assess the effects of an inserted stage in order to test the validity of the pure insertion assumption. Since there is ample criticism with Donders (1868; 1969) original task arguing that differences in RT may not be caused by the insertion of a processing stage but rather by differences in task difficulty (Smid et al., 2000), we used a task introduced by Anderson, Qin, Jung, and Carter (2007). The task is designed to accommodate Donders (1868; 1969) original assumptions that it requires distinct successive cognitive processing steps, where processing of an adjunct step should depend on the processed product of a prior step. The task first requires participants to encode three visual stimuli and remember their corresponding order. In a response phase afterwards, the order of the presented stimuli has to be indicated, which can differ based on prior displayed instructions. By manipulation of the task instruction, a memory retrieval (memory recall of a learned association required by displayed instruction), a transformation (order information of internal representations of the three stimuli has to be manipulated) or both can be inserted in the processing chain of the task between instruction and response. Using a HSMM analysis, we aim to assess if the assumption of pure insertion for the retrieval and transformation processes holds by assessing whether onset latencies and durations of individual processing stages vary between conditions.

We expect to detect an encoding stage which repeats for the three visual stimuli. This will also allow us to check the method as durations for the stimuli presentations are known. Moreover, we expect another encoding stage for the encoding of the instructions following name presentation. Depending on task condition, we expect the insertion of a memory retrieval, where a prior learned association has to be retrieved from memory based on the prior learned instruction. If both retrieval and transformation are required by the instruction, the retrieval has to occur before the transformation. Thus, for the final task we expect the following processing stages:

- 1) An encoding stage for each of the presented items where a mental representation of the item and associated order is created
- 2) Another encoding stage where task instructions are encoded
- 3) Possibly the insertion of a memory stage to retrieve the learned instructions
- 4) Possibly the insertion of transformation stage where associated order information of the encoded items is manipulated
- 5) Preparation and execution of a motor response – to indicate readiness to respond

With regard to determining the validity of our model analysis, we expect to identify the same sequence of processing steps with our HSMM analysis for each of the three item presentations. If the assumption of pure insertion is true, we will not observe any differences in the duration of the processing stages shared among all

conditions (i.e. encoding of the instructions, response preparation and response). To assess this assumption, we will fit a HSMM to each condition separately, to identify signatures of shared processing stages. Pure insertion could be tested by comparing gamma distributions of shared signatures as well as the differences between averaged bump onsets. If pure insertion is true, no differences should be observed for either of these measures.

METHOD

Participants

For the purpose of this study we collected data of 30 participants. For methodological reasons explained in the beginning of the results section, we excluded seven participants for a final sample of 23 participants (11 females, mean age = 20.74y). Two participants were left handed and all had normal or corrected-to-normal vision. Participants were undergraduate first year students of the Artificial Intelligence Bachelor program at the University of Groningen and received a monetary compensation of 20€ for their participation. Prior to testing all participants gave their informed consent in accordance with the declaration of Helsinki (World Medical Association, 2013).

Experimental design

The experiment consisted of an information processing task that required participants to recall and sometimes manipulate the order of a set of three names based on instructions. The instructions could come either in the form of a number combination (two different numbers from the range of 1 to 4, e.g. “13”) or a letter combination (e.g. “IT”, see Appendix Instruction sheet 1 for a complete list of stimuli). The two within-subject factors manipulated in the instruction were transformation (switching name order or not) and memory retrieval (no retrieval versus retrieving the association for a letter pair), leading to four experimental conditions:

- 1) Condition 1 (C1) “No retrieval – no transformation” (number pair associated with no action)
- 2) Condition 2 (C2) “No retrieval – transformation” (number pair associated with action)
- 3) Condition 3 (C3) “Retrieval – no transformation” (letter pair associated with no action)
- 4) Condition 4 (C4) “Retrieval – transformation” (letter pair associated with an action)

An instruction would require a participant to switch the order of the presented names if both digits associated with the instruction cue could correspond to a position of the three names displayed. For example, the cue “13” would require participants to switch the position of the first and third name, whereas the cue “24” would not require a subsequent reordering, because there is no fourth name. Further, a recall would be required if the instruction were to come in the form of a letter pair. All letter pairs were associated with a specific number

ASSESSING PURE INSERTION USING HSMM

combination, which were learned by the participants prior during the learning phase. When seeing a letter pair participants had to recall the corresponding number pair and complete the instruction associated with it.

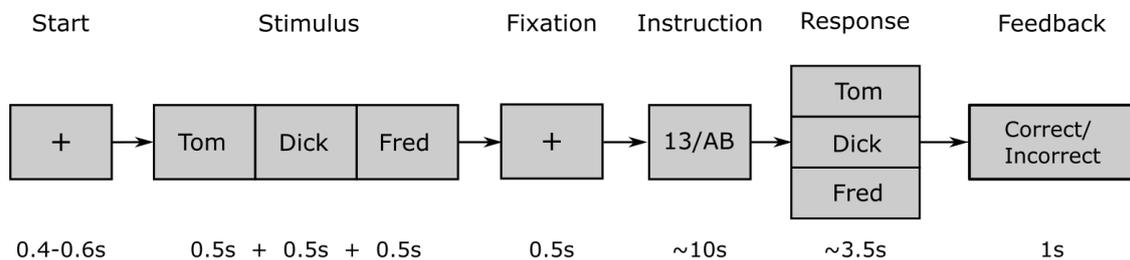


Figure 1. Schematic representation of the experimental task.

Learning phase

Prior to completing the actual experiment, participants were first given 15 minutes to learn associations between 12 letter and number pairs. After the initial learning period, participants were tested on their ability to recall the associations in a short test program. On screen participants would see a letter combination and had to type in the corresponding number pair and received feedback on whether the response was correct or not. Always only one letter pair would be shown until a response was indicated and the order of the letter pairs appearing on screen was randomized. In a given test cycle, a letter pair would appear until answered correctly. Participants had to complete four cycles of testing, with each cycle being completed once every letter pair had been answered with the correct number pair once.

Experimental phase

The experiment consisted of 6 blocks of 72 trials each. Presentation of conditions within each block was randomized and balanced (i.e. 18 trials per condition). In total, each participant thus completed 108 trials for each condition. Each trial started with the display of a fixation cross at the center of the screen for a duration randomly drawn between 400-600ms after which one by one each of the three names in random order was presented for 500ms (see Figure 1). After the three names were presented, another fixation appeared at the center of the screen for 500ms before the instruction cue (either a number or a letter pair) appeared on screen. Based on the instruction cue, participants had to either simply indicate the order that the three names were presented in or rearrange the order of the three presented names. The instruction cue remained on screen until participants indicated readiness to respond by a left click anywhere on screen. Responses were collected using a trackball as an input device, which was done to make results comparable to a separate study conducted in parallel in an fMRI environment. Upon seeing the instruction cue, participants had up to 10000ms to prepare their response and indicate their readiness to respond by a left click. If they exceeded this time limit, a trial would be rated as

ASSESSING PURE INSERTION USING HSMM

incorrect and feedback informing participants that their response was too slow would appear on screen. If they indicated readiness to respond within the time limit, the instruction disappeared and the three names stacked on top of each other in random order were displayed on screen. Upon presentation, participants had 3500ms to click each of the three names in the correct order as required by the instruction. The time limit of 3500ms was chosen to give participants enough time to find the correct names on the response screen, but not enough time to still process the instruction. Exceeding the time limit again would lead to a feedback prompt indicating that the response was too slow. Correct answers would be followed by feedback indicating that the response was correct and incorrect response by feedback indicating an incorrect response vice versa. A trial would be completed by the display of another fixation cross for 1000ms. After each block, participants would see a prompt on screen, informing them that a block was finished. In between blocks, participants were allowed to take short breaks and had to inform the experimenter if they wished to continue, who would start the next block. Upon completion of the experiment, participants received a prompt on screen, informing them that they reached the end of the experiment and thanking them for their participation. The experiment was designed and presented in OpenSesame (Mathôt, Schreij, and Theeuwes, 2012).

EEG-recording

EEG was recorded from 128 locations by using a “Biosemi Active Two” amplification system as well as two electrodes placed on the left and right mastoids. Further, horizontal electro-oculogram (HEOG) was recorded as the difference of two electrodes placed 1cm left and right to the eye. The vertical- electro-oculogram (VEOG) was recorded as the difference in activity of two electrodes placed 1cm above and below the left eye. Data were sampled and digitized at 512 Hz with a band-pass-filter of 0.16-100 Hz. Electrode offsets were kept within a range of $\pm 20\text{mV}$.

EEG-preprocessing

The data were re-referenced to the average of the right and left mastoid and filtered with a high-pass filter at 0.5Hz and a bandstop filter at 50Hz to remove line-noise. To decrease processing time, data were subsequently epoched into trials of varying length, starting from the disappearance of the first fixation and ending with the presentation of the feedback screen and down-sampled to 256Hz. The data were then inspected visually and trials containing artifacts excluded from further analysis. Further, eye artifacts were identified and rejected by decomposing the data into independent components (IC) as implemented by Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011). ICs with a time-course and topography indicative of saccades and eyeblinks were excluded and the signal recomposed with the remaining components. Prior to performing IC analysis,

ASSESSING PURE INSERTION USING HSMM

electrode channels containing too much noise as identified by single channel variance and subsequently confirmed by visual inspection were excluded from the IC analysis. After artifact removal, data were further down-sampled to 100Hz reducing the amount of data points to make computations with the HSMM-MVPA analysis more efficient. Further, baseline correction for the EEG signal was performed using the first 300ms prior to trial onset (i.e. the disappearance of the first fixation).

General Analysis

All subsequent analyses were applied to correctly answered trials only which were not more or less than three standard deviations away from the mean response time for each participants or shorter than 750ms in the condition requiring both a memory retrieval and a transformation. The motivation for this approach will be discussed in the HSMM-MVPA analysis section. Further, in the memory conditions we assessed accuracy for each item and participant individually. If accuracy was below chance (16.66%), correct answers for this item were excluded from further analysis. This was done to avoid including data in the analysis that was only answered correctly by chance and consequently did not include the processing stages required by the task. In total 31.7% of all trials were excluded from this analysis.

ERP-Analysis

To analyze differences between experimental conditions, EEG data were first averaged stimulus-locked (to instruction onset) for each participant separately. Averages were analyzed statistically using non-parametric cluster-level statistics to correct for multiple comparisons (Maris & Oostenveld, 2007) using Fieldtrip software (Oostenveld et al., 2011). First, we calculated dependent-samples t-tests between condition specific averages for every electrode- sample pair and participant separately. We selected all samples and electrode comparisons whose t-values crossed a significance threshold ($\alpha < .05$). Selected samples were clustered based on temporal and spatial adjacency. Next, the cluster-level statistics was calculated by summing the t-values within a cluster. The largest of cluster-level statistics was used to evaluate if there were differences between the experimental conditions. In order to do that, we first calculated non-parametric statistics by using a permutation test. With the test we assessed the null hypothesis that the probability distribution of the condition specific averages is independent of experimental condition. To calculate the permutation distribution, we randomly exchanged conditions within each subject and calculated the difference between both conditions. This procedure was repeated 500 times and for each permutation the cluster statistic with the largest sum was retained. Then, we compared the number of clusters from the random permutations that are larger than the actual cluster observed in

the data to determine our cluster-corrected p-statistics. We only report here clusters with significant p-statistics ($\alpha < .05$) in all analyses.

HSMM-MVPA Analysis

Prior to performing the HSMM-MVPA analysis, we band-pass filtered again (1-35Hz) and de-trended single trial data. This was done in order to allow the signal to return to sinusoidal noise with mean zero after the onset of significant cognitive event, so we could estimate the “flat”. Moreover, we performed a principal component analysis (PCA) retaining 10 components, which accounted for 91% of the variance. The retained principal components were z-scored to have a mean of zero and a variance of one.

The HSMM-MVPA analysis finds the most likely partition of each trial into a certain number of stages. Each stage starts with a bump, a multi-dimensional 50ms deflection, whose amplitude is estimated by the HSMM-MVPA. As mentioned, after a deflection neural activity is thought to return to sinusoidal noise with mean zero. This flat-interval between two bumps can vary between trials and is modelled by a gamma-distribution with a shape parameter of two and a scale parameter to be estimated by the HSMM-MVPA. Thus a cognitive-processing stages onset is marked by one bump and its offset by the subsequent bump. An exception to that rule are the first and the last processing stage. The first processing stage starts not with a bump but with the stimulus presentation and the duration of the flat represents the time until the stimulus is registered in cortical areas (Portoles et al., 2017). The last processing stage on the other hand starts with the last bump, but ends with the response, which does not necessarily indicate the end of the processing stage.

From this it follows that for an n bump, HSMM $n+1$ flats describing the processing stages and n bumps describing the transitions between states have to be estimated. For an n bump model, the most likely partition of each trial into $n+1$ stages has to be determined. Parameter estimation here is constrained by the total duration of each trial. That is the duration of all flats and bumps must sum up to the total trial duration ($F_1 + B_1 + \dots + B_n + F_{n+1} = \text{Trial duration}$). Based on this constraint, the HSMM calculates all possible ways that a trial can be parsed into $n+1$ stages and calculates the summed log-likelihoods of these partitions. It then selects bump magnitude parameter for the bumps and scale parameter for the flats that maximize the likelihood of the data for all trials. The procedure also returns probabilities of the bump occurring for every sample in every trial, which we use to determine the mean onset latency.

Further, in line with Portoles et al. (2017), prior to each n -bump configuration testing, we determined the best initial magnitude parameter. We performed 100 HSMM-MVPA analyses with different random onset amplitudes and gamma scale parameters, consisting of equal $n+1$ fractions of the duration of the longest trial. The random onset amplitudes with the highest likelihood subsequently was used for the actual HSMM-MVPA

analysis. This was done in order for the estimation algorithm not to converge on a local maximum that is sub-optimal (Portoles et al., 2017).

To determine the number of bumps that describe the data best and to avoid overfitting by adding bumps, we initially performed LOOCV. First, we estimated magnitude and scale parameter on all trials of all participants except one. Next, we determined the likelihood of the missing participants data based on the parameter estimated in the previous step. This procedure was repeated until the likelihood of every participant's data had been determined in this manner. In order to justify the inclusion of an additional bump, the likelihood of the data of each participant had to increase over all previous bump configurations. This was determined by a series of one-sided sign-tests assessing if the likelihood of the data of the current bump configuration increased over each configuration with fewer bumps. While this was the initial approach, the resulting bump configurations turned out to be insufficient explaining differences between experimental conditions. Thus we changed criteria slightly using visual inspection for a known response topology as the criteria. We will explain this in more detail in the general discussion.

RESULTS

For the final analysis we used a sample of 23 participants. Two participants had to be excluded due to technical difficulties during recording. Five more participants had to be excluded because not sufficient correct trials could be obtained in all experimental conditions to perform the HSMM analysis. That is, two participants did not answer any trial correctly in the conditions requiring a transformation (C2 & C4), whereas the remaining three subjects had too few correct trials in the retrieval-transformation condition (> 10 versus an average of 30+ for the other participants). The data of these participants are consequently not considered in any of the following analyses.

Behavior Results

To evaluate the effects of transformation and retrieval, we first assessed the effects on reaction time to confirm that added stages lead to increased processing time. Thus, we expected longer reaction times for conditions including at least one inserted processing stage, with the condition including both retrieval and transformation requiring the most time for successful problem solving. Reaction time in our analysis is considered as the time from the display of the instruction, until participants indicated their readiness to respond. The average reaction time in each condition and respective standard errors are displayed in Figure 2. For the calculation of the average reaction times, we only considered correctly answered trials. On average, insertion of

ASSESSING PURE INSERTION USING HSMM

an additional processing stage, increases reaction time with the longest mean reaction time for C4 (inserted retrieval stage + inserted transformation stage).

To confirm that these differences are statistically significant, we performed a mixed-model linear regression analysis with reaction time as the dependent variable and participants as a random factor. Since reaction times exhibit a floor effect (having a non-normal distribution, see Appendix Figure 9), we first performed a log-transformation, making their distribution approximately normal. To assess effects of our experimental manipulation, we did not only assess retrieval and transformation but also considered whether the to be transformed items were successive (i.e. the first and the second name, or the second and the third name – form hereon distance 1) or not (i.e. the first and the third name – form hereon distance 2). We decided to include the distance factor, since during the experiment multiple participants reported that they had a significantly harder time performing distance 2 transformations compared to distance 1 transformations.

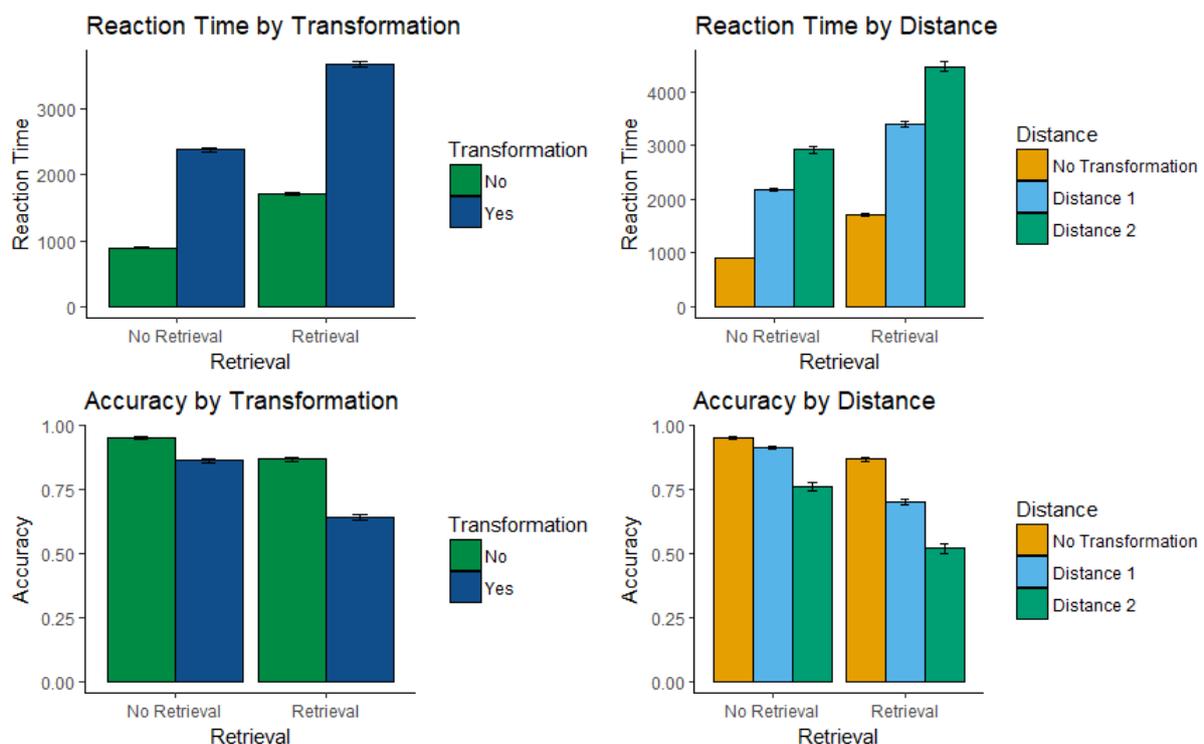


Figure 2. Behavioral results. On the left hand side we see the effects of the experimental manipulation and on the right hand side we see the same results but distinguishing distances in the transformation manipulation. In the top graphs the effects with regard to reaction time are described, with average reaction time in ms on the y-axis and retrieval effect on the x-axis. In the bottom graphs we see the effects with regard to accuracy with mean accuracy per condition in the y-axis and retrieval effect on the x-axis. Color of the bars distinguishes the transformation effects. Standard errors had between-participant variance removed (Morey, 2008).

Based on model selection criteria, Akaike's Information Criterion (AIC), a model containing the factor “retrieval” ($\beta_{\text{Retrieval}} = 0.57$; $t = 15.42$; $p < .0001$), the factor “distance” ($\beta_{\text{Distance1}} = 0.8$; $t = 14.09$; $p < .0001$; $\beta_{\text{Distance2}} = 1.09$; $t = 14.06$; $p < .0001$), the interaction ($\beta_{\text{Distance1*Retrieval}} = -.1$; $t = -5.29$; $p < .0001$; $\beta_{\text{Distance2*Retrieval}} = -.15$; $t = 5.31$; $p < .0001$) and random effects with random slopes for factors participants was determined to have

ASSESSING PURE INSERTION USING HSMM

a better fit than a simpler model including the factor transformation instead of distance. Table 1 depicts the final model (see Table 1). This model predicts increased reaction times, when a retrieval had to be performed compared to when no retrieval had to be performed. It foretells also an increase when a transformation had to be performed with a larger increase when the transformation was a distance 2 transformation. Further, as indicated by the significant interaction terms, overall RT increases when both a retrieval and a transformation had to be performed (slightly more for a distance 2 transformation). The final model explained a significant proportion in the variation of reaction times ($R^2 \approx .68$) (Xu, 2003) and is depicted in Equation 1 (Appendix).

Fixed Effects	Estimate	Std.Error	t-value	p-value
Intercept	6.7339	.0315	214.0087	>.0001
Dist_D1	.7982	.0567	14.0882	>.0001
Dist_D2	1.0922	.0777	14.0591	>.0001
Ret_Retrieval	.5748	.0373	15.4201	>.0001
Dist_D1- Ret_Retrieval	-.1021	.0193	-5.2869	>.0001
Dist_D2- Ret_Retrieval	-.1463	.0276	-5.305	>.0001

Table 1. Mixed-effects linear regression results on logarithmically transformed reaction time data with significant fixed-effects for transformation distance, retrieval, a significant interaction term for retrieval and transformation distance 1 and transformation distance 2.

To assess if the difference in added processing stages also affected task difficulty, we assessed effects on accuracy. Based on model selection criteria Akaike's Information Criterion (AIC) a model containing the factor "retrieval" ($\beta_{Retrieval} = -.99$; $z = -5.88$; $p < .0001$), the factor "distance" ($\beta_{Distance1} = -.72$; $z = -4.24$; $p < .0001$; $\beta_{Distance2} = -1.90$; $z = 14.01$; $p < .0001$), the interaction ($\beta_{Distance1*Retrieval} = -.48$; $z = -3.09$; $p < .005$) and random effects with random slopes for factors participants was determined to have a better fit than a simpler model including the factor transformation. Table 2 depicts the final model (see Appendix). This model predicts a decrease in accuracy when a retrieval had to be performed compared to when no retrieval had to be performed. It foretells also a decrease when a transformation had to be performed with a larger decrease when the transformation was a distance 2 transformation. Further, the significant interaction indicates that accuracy decreases more when both a retrieval and a distance 1 transformation had to be performed. The final model is displayed in Equation 2 (Appendix).

Fixed Effects	Estimate	Std.Error	z-value	p-value
Intercept	3.1961	.1993	16.038	>.0001
Dist_D1	-.7169	.169	-4.242	>.0001
Dist_D2	-1.8967	.216	-8.783	>.0001
Ret_Retrieval	-.9889	.1683	-5.875	>.0001
Dist_D1- Ret_Retrieval	-.4792	.1553	-3.085	.002
Dist_D2- Ret_Retrieval	-.0208	.1651	-1.26	.2028

Table 2. Mixed-effects logistic regression results on accuracy data with significant fixed effects for transformation distance, retrieval and a significant interaction term for retrieval and transformation distance 1.

ERP Results

To assess statistical differences between ERPs with inserted conditions, we performed nonparametric cluster-level statistics for the period after instruction display. To see if we could identify the inserted conditions, we compared the condition without any inserted processing stage (C1) against each condition including inserted processing stages (i.e. C2, C3, C4). Moreover, we also calculated cluster-level statistics for the conditions with one inserted processing stage versus the condition with two inserted processing stages (C2 vs. C4 & C3 vs. C4 respectively). However, we could only identify statistically significantly different clusters for the comparisons between C1 vs. C3 and C1 vs. C4.

For the comparisons of C1 vs C3 we found a statistically significant difference between conditions (see Figure 3 left) between 600-700ms after the display of the instruction around right parietal areas of the scalp. In this interval on average amplitudes in C1 were higher compared to C3. For the comparison C1 vs C4 we found a statistically significant difference between 400-450ms after the display of the instruction (Figure 3 right). The identified cluster consisted of three electrodes in right frontal positions. During this interval amplitudes of C4 were on average higher than C1.

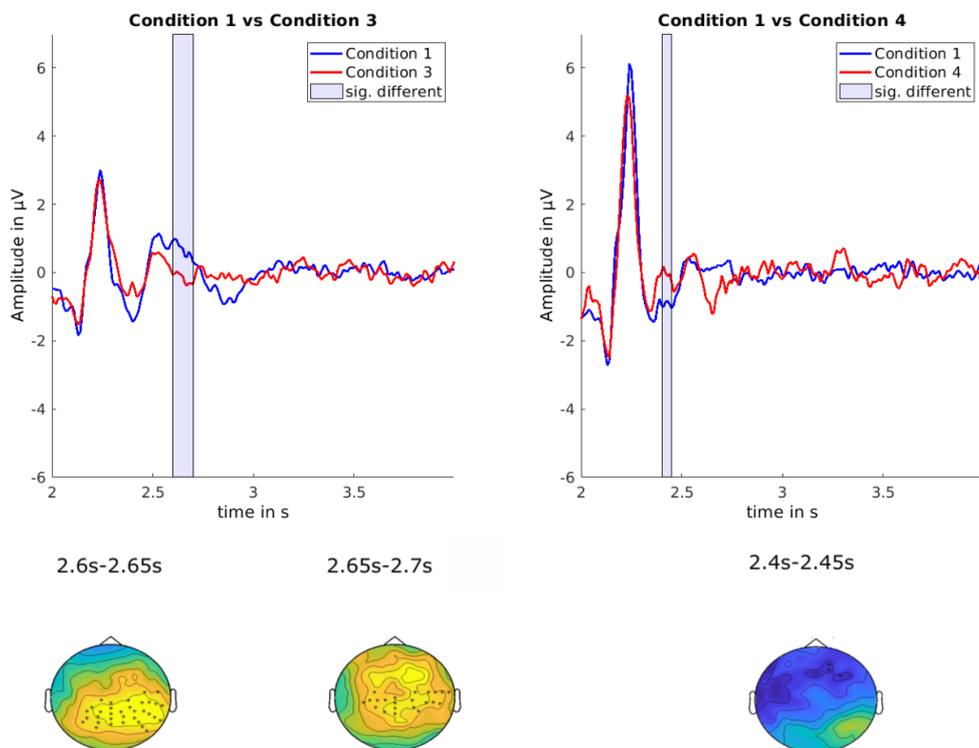


Figure 3. ERP results. On top we see the averaged EEG activity of all electrodes with a significant difference in the comparison. The grey bar indicates the period during which conditions differed significantly from each other. Below we see the corresponding electrode clusters of the significant comparison. The underlying topographical plot is the algebraic

ASSESSING PURE INSERTION USING HSMM

difference between the ERPs of C1 and C3 or C1 and C4 respectively for the significant duration. For all figures, the displayed time axis is in references to the start of the trial (the display of the first name), thus two seconds marks the display of the instruction.

HSMM Results

Names Presentation

To determine how well the HSMM analysis is able to identify stable processing stage signatures, we first analyzed the presentation of the three names. Figure 4 shows the relative increase in model log-likelihood of each bump configuration compared to a one-bump HSMM model of the name presentation. For each bump model results of the LOOCV are displayed below, indicating for how many participants inclusion of an additional bump improved model log-likelihood compared to the previous bump configuration. Based on the results of the LOOCV, we determined a model with 16 bumps to fit the data of the presentation of the three names best.

The results of the 16 bump model are displayed in Figure 5. After an initial bump at 70ms after the disappearance of the fixation cross and the display of the first name, we see a first bump with a negative topology. Subsequently, we see a repetition of five bumps in the same order three times with comparable topographical distributions. A first bump with a broad negative topology, followed by a positive deflection around the frontal areas. The third bump appears to differ slightly between the presentation of the first-, vs the second and third name. Whereas in the latter two names we can observe a positive deflection in the frontal areas, in the first name we see a positive deflection more around occipital sites. The fourth and fifth bumps again have a rather negative distribution with the fourth showing slightly more positive distribution in the frontal areas. There is very little variation in the duration of the flats between all of the bumps, which all have a median duration around 90ms. The only exception is the last flat, which however as mentioned does not indicate that the processing stage ends with the disappearance of the last name.

ASSESSING PURE INSERTION USING HSMM

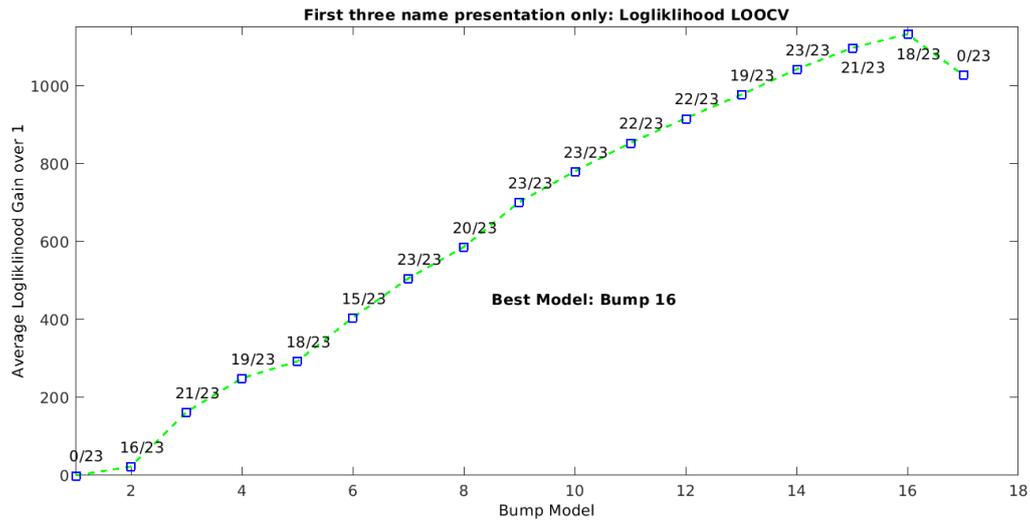


Figure 4. LOOCV results of all bump-configurations for the display of the three names. On the y-axis is the average gain for each bump configuration over the average log-likelihood of a bump one model. On the x-axis we see the respective number of bumps of each model. For each bump configuration we see for how many participants adding an extra bump increased the likelihood significantly compared to the previous bump configuration. The best model is the last model where the increase in likelihood over all previous bump-models is significant.

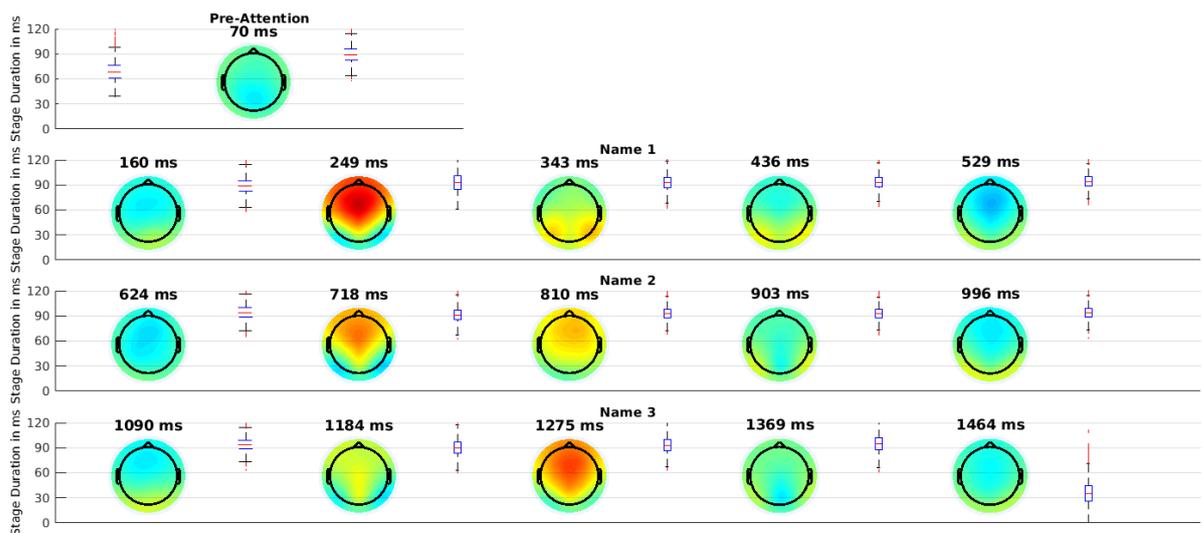


Figure 5. Best Bump configuration of the three names as determined by LOOCV. On each y-axis we see the duration of the flats in ms. Displayed we see boxplots of the durations of the flats calculated as the difference between the average onset between two adjacent bumps for each trial. Between the flats we see the averaged magnitudes of the bump distribution of scalp activity. For better comparison we split the model in four parts. The top-graph displays the first bump and first two flats which appear to be not specific to the presentation of each name. In the graphs below the subsequent five bumps and flats are always displayed, showing a pattern which appears to repeat itself for each name presentation.

Experimental Conditions

The results of the LOOCV varied between experimental conditions (see Figure 6.). For C1 (top left Figure 6.) we see a gradual increase in likelihood gain over bump 1 until a bump 6 model and afterwards a gradual decline. However, for C2 - C4 a maximum in likelihood gain is more difficult to identify. Whereas increasing bump configurations appear to lead to an increase in likelihood gain on average this is not the case for

ASSESSING PURE INSERTION USING HSMM

all participants. Thus, the best models as determined by LOOCV paradoxically had fewer bumps in the conditions with inserted processing stages, than the condition without processing stages (the resulting model comparison can be found in the Appendix Figure 11) and were consequently not very informative. Subsequently, as previously announced we decided to select models based on visual selection, matching processing stages shared amongst the conditions. We will discuss this further in the discussion.

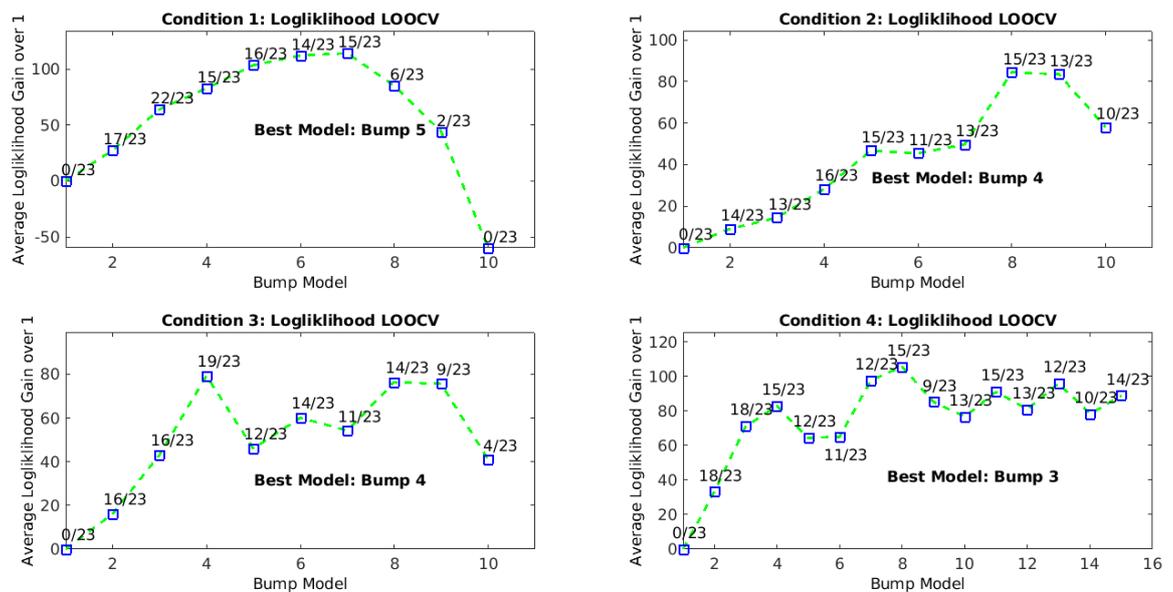


Figure 6. LOOCV results of all bump configurations for each experimental condition. On each y-axis is the average gain for each bump configuration over the average log-likelihood of a bump one model. On each x-axis we see the respective number of bumps of each model (note condition 4 was tested up to 15 bumps vs. 10 bumps in the other conditions). For each bump configuration we see for how many participants adding an extra bump increased the likelihood significantly compared to the previous bump configuration. The best model is the last model where the increase in likelihood over all previous bump-models is significant.

In Figure 7 underlined in red we see that the first three bumps are shared amongst all conditions. The onset of the first bump in C4 occurs slightly later compared to the other conditions. The signatures and latencies of the first and third bump are very similar to the bump topologies identified by Zhang, Walsh and Anderson (2018). They identified the first bump as resembling the N1 ERP component, which reflects visual attention to the stimulus (Luck, Woodman, & Vogel, 2000). The third bump on the other hand with a positive anterior distribution is thought to correspond to the P2 ERP, which reflects the encoding of visual stimuli (Finnigan, O'Connell, Cummins, Broughton, & Robertson, 2011). From the literature it is difficult to infer what process the second bump could be related to, but it seems reasonable to assume that it is also involved in visual encoding giving its latency and resemblance to the third bump. Overall, the bumps we identified in the experimental conditions to reflect encoding closely resemble the first three bumps we identified in the encoding of each of the three names.

ASSESSING PURE INSERTION USING HSMM

The last two bumps distributions in C1 – indicated by the blue line in Figure 7 – can also be identified in all other conditions but C2. The duration of the flat following the last of these two bumps seems to be much longer and to vary significantly in C4. It has a median duration of 1100ms compared to 250ms in the other conditions.

Underlined in yellow we see a bump that is shared between C2 - C4. As it is the only bump that distinguishes C3 from C1, we believe to represent a memory retrieval. It therefore also naturally occurs in C4. Surprisingly, this stage appears to be also involved in the transformation condition. However, the following flat varies significantly in duration between C2 ($M = 677.8$ ms, $SD = 176.9$ ms) and C3 ($M = 909.5$ ms, $SD = 243.8$ ms) indicating qualitative differences between both retrieval stages; $t(44) = -3.69$, $p < .001$. It is very likely that both of these stages reflect the same process as the same stage because C4 ($M = 1058.6$ ms, $SD = 319.4$ ms) is significantly longer than C2 ($t(44) = -5.01$, $p < .001$) but not as C3 ($t(44) = -1.78$, $p < .081$) alone. The magnitude distribution and latency most closely resembles the FN400 ERP, which has been mostly associated with processes of familiarity but also appears during recollection (Tsivilis et al., 2015).

Lastly, underlined in green we see four bumps of the transformation condition (C2) that also reappear in the transformation and retrieval condition (C4). However, in C2, the last bump appears to be followed by a significantly different flat. In terms of stages that indicate a response, unfortunately, we have little agreement between the conditions. The last bump in C4 is what we believe to be a response stage. The same stage can also be identified in the other conditions, however, at the cost of either losing consistency in the first three stages or the addition of more bumps, whose presence is difficult to explain as they cannot be identified in C4 anymore.

ASSESSING PURE INSERTION USING HSMM

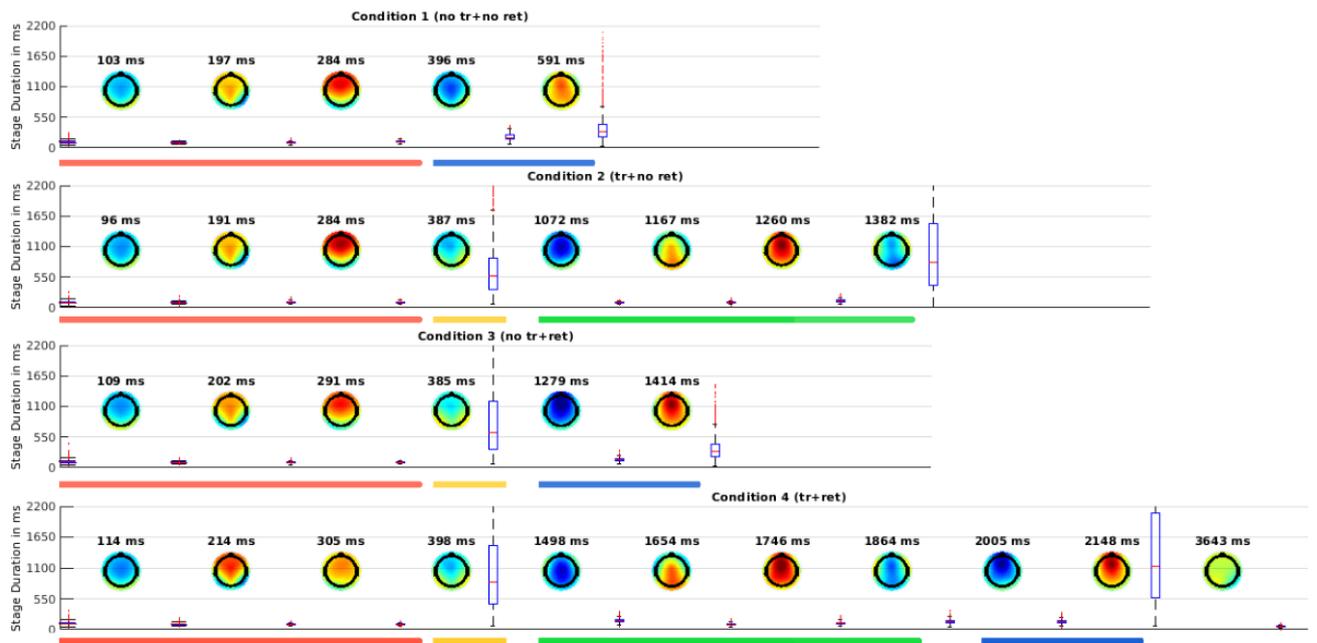


Figure 7. Best Bump configurations of each experimental condition from instruction display to response as determined by visual inspection. From top to bottom condition one to condition four. On each y-axis we see the duration of the flats in ms. Displayed we see boxplots of the durations of the flats calculated as the difference between the average onset between two adjacent bumps for each trial. Between the flats we see the averaged magnitudes of the bump distribution of scalp activity. Underlined in red we see the three bump topologies that are shared amongst all conditions. Underlined in yellow we see a bump topology that is shared amongst the conditions that have an added processing stage. Marked in blue we see two bump topologies that is shared among condition one, three and four but not two. And lastly marked in green we see a set of topologies shared among condition two and four alone.

Pure Insertion

We tried to determine whether the insertion of an additional processing stage affected the duration of other shared processing stages across conditions. However, we were only able to consistently identify the first four stages across all conditions. In the first three stages, stage onset and offset as indicated by the respective bumps with similar topologies does not differ between conditions. However, the fourth stage is significantly different in C1 compared to C2 ($t(44) = 6.36, p < .001$), C3 ($t(44) = 12.9, p < .001$), and C4 ($t(44) = 14.13, p < .001$). Stage four durations for C1 ($M = 111.3$ ms, $SD = 5.5$ ms) are consistently longer in contrast to C2 ($M = 102.6$ ms, $SD = 3.4$ ms), C3 ($M = 94.1$ ms, $SD = 3.2$ ms), and C4 ($M = 92.9$ ms, $SD = 2.9$ ms). Figure 8 displays the average stage durations, which across all four stages vary between 90ms to 110ms. With regard to a response or response preparation stage, we were unable to isolate it in the conditions and thus are unable to compare them. Nevertheless, we also compared inserted stages shared by C2 and C4 and C3 and C4. For C2 vs C4, the stage 4 and stage 9 seem to differ, whereas for C3 vs C4 the last stage (stage 7 or 11 respectively) differs. Results here however need to be interpreted with caution, as the identified HSMMs were selected based on plausibility as determined by LOOCV and visual inspection. While the first four stages were reliably identified across conditions in a majority of all HSMMs considered, the later stages (in particular in the conditions including a

ASSESSING PURE INSERTION USING HSMM

transformation) were relatively unstable and could differ from one bump configuration to the next. It is likely that fewer or more stages are involved in the conditions.

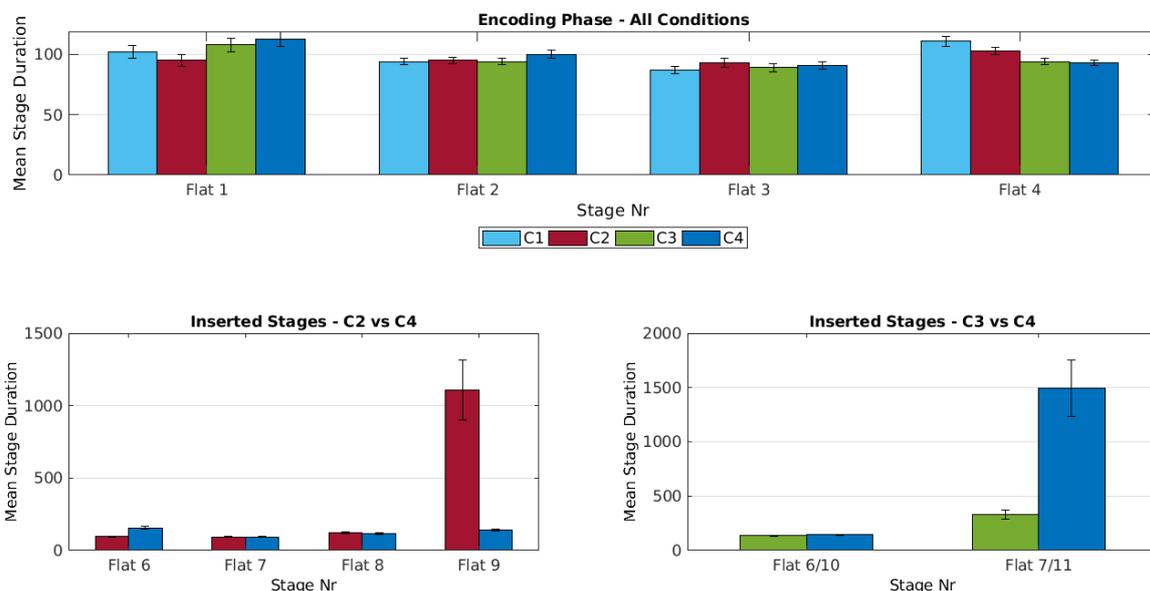


Figure 8. Duration of the shared stages between experimental conditions calculated as the average of the difference between the average onsets between two adjacent bumps for each trial. At the top, the first four stages shared among all conditions, at the bottom the inserted stages shared between C2 and C4 (bottom left), and C3 and C4 (bottom right). On the x-axis we see the corresponding stages of each comparison indicated as flats. Note that for the comparison C3 vs C4 the inserted stages (specific to C3) are thought to occur later in C4 than C3. Hence, here two flats are indicated (left = C3, right = C4). On the y-axis the average duration of the stages is indicated in ms.

DISCUSSION

In this study we set out to investigate the assumption of pure insertion by using an HSMM analysis to parse EEG data of a name ordering task into independent stages. We first validated the HSMM analysis on the presentation of the three names. We were able to distinguish the presentation of each name by a sequence of five processing steps, which repeated for every name presentation. The behavioral results indicated that our experimental manipulations were successful. The insertion of an additional mental process increases the difficulty of the task and requires more time by the participants to successfully solve it. To identify the locus of this increase in processing time, we performed a HSMM analysis. Compared to the baseline condition (C1), we were able to identify the inserted retrieval process, which likely corresponds to the FN400 ERP. Results were less clear with regard to the transformation condition. Our results suggest that the transformation of the order of the three names may consist of more than one inserted processing stage, of which one may consist of a retrieval process. Further, we were able to identify all bump magnitudes specific to the conditions with one inserted stage also in the experimental condition that contained both insertions. With regard to the stages shared across conditions, we were successful to isolate shared stages of the encoding of the instruction. Unfortunately for the

response stages no parsimonious solution consistent across conditions could be identified, likely due to a lack of data. Thus, to assess pure insertion, we were only able to compare the encoding stages that did not differ significantly across conditions.

In the remainder of this thesis, we will first discuss conclusions to the stages identified in the task, then problems with the experimental design to assess pure insertion, and lastly in general whether HSMMs are a suitable tool to assess this assumption.

Identified Stages

The pre-attention and encoding stages identified in this task (bump 1-3 and corresponding flats) are comparable to the ones identified by HSMMs of other tasks by Anderson et al. (2016) and Zhang et al. (2018). An exception is the second bump identified across all conditions. It remains to be determined if this is an independent processing stage or just a subdivision of the third bump. It is not uncommon that the HSMM subdivides states, whose functional significance is not apparent at first sight. Borst and Anderson's (2015) results initially indicated a 7-bump instead of a 6 bump solution. However, the authors determined that the added stage in the 7-bump solution was just a subdivision of the first bump and that a 6-bump solution was more parsimonious and congruent with the cognitive model of the task. To determine if this is a subdivision in our data, we could analyze the correlation of the estimated bump magnitudes for every trial between the two adjacent stages in temporal proximity as used by Zhang et al. (2018).

The last stage which we assume occurs prior to a response in C1 and C3 has a relatively long duration. Moreover, compared to Anderson et al. Zhang et al. (2018) it has a stronger anterior positivity suggesting more frontal involvement. J. Borst (personal communication 24.07.2018) suggested that according to his ACT-R model of the task, participants may rehearse the three names before they respond. This may explain the relatively long response preparation stage for a relatively simple response (a simple click). However, while response preparation explains the long duration of this last stage what is still missing in our analysis consistently across conditions is a response stage. In addition to response preparation stages, other HSMM analyses (e.g. Anderson et al., 2016; Zhang et al., 2018) were also able to identify brief stages that reflect the response.

The inserted bumps in condition two and three have different durations but a very similar topology. For a future analysis a connectivity analysis in the manner of Portoles et al. (2017) to the bump-locked EEG data should be performed to confirm that these are indeed distinct processes that do not differ in the relevant areas involved in the respective computations. The retrieval process identified here varies from the retrieval processes identified by other HSMM-MVPA (Anderson et al., 2016; Zhang et al., 2017). It appears to reflect more a FN400 associated with familiarity rather than the stage in other studies that was believed to be related to the parietal

“old-new effect” (Tsivilis et al., 2015). Familiarity may be involved, when participants recognize the instruction, however the process requires recall and is relatively long. The process is not involved in C1, which suggests that if this stage corresponds to the FN400, it does not occur necessarily after instruction display. However, it does occur also during C3 although shorter and increases, when retrieval and transformation are both required in C4. It is not apparent why a familiarity process should be involved in the transformation C2 but not C1.

Distance Effect

We decided to disregard differences between the transformation of names that did not follow each other in our analysis. This was done in order to have sufficient data to perform the HSMM analysis. The inclusion of all distances led to overall more stable bump patterns and flat durations in the first half of a trial, but not the second. However, this may have been problematic for three reasons: 1) The results of our behavioral analysis clearly indicate that reaction times as well as accuracy is impacted by the distance of the two names that have to be transformed, 2) excluding transformations of names with distance 2 improved overall model log-likelihoods of all tested bump models (see Appendix Figure 12) in the transformation condition, 3) upon reflection we could not exclude the possibility that there are two ways to solve the distance 2 problem, whereas there can be only one for the distance 1 problem. To explain the latter we need to consider the original ACT-R model of the task (see Anderson et al. 2007). According to the model a transformation task is solved by keeping the two names that have to be switched in working memory and then rebuilding the new sequence in working memory. This is the only way the distance 1 problem can be solved. However, in the distance 2 problem you can either use this approach for problem solving or simply recall the whole sequence backwards. This presents a completely different type of problem solving. This may also explain why previous studies (Anderson et al., 2007) employing the same task were unable to detect a distance effect in RT and accuracy data.

We tried to determine if the behavioral data contained evidence for participants using different strategies to solve different transformation problems. We tested this by assuming that we should be able to observe differences in RT between participants for solving distance 2 transformations. Specifically, we propose that some participants may be able to solve distance 2 transformations on average faster than or at least as fast as distance 1 transformations. Figure 10 (Appendix) shows RT graphs for every participant individually. However, it appears that only some participants (e.g. participant 1 and participant 5) are able to solve distance 2 transformations at least as fast as distance 1 transformations. Hence, at least in our dataset there appears to be no plausible reason to assume that participants varied in the strategies that they used.

It would be interesting to determine the source of the distance effect, whether is it a difference cognitive processing stages or task difficulty? We attempted to probe this question using separate HSMMs for each

ASSESSING PURE INSERTION USING HSMM

distance, but unfortunately did not have sufficient trials for consistent results. However, we assume that if the source is task difficulty then we should identify the same set of bumps in two datasets containing only distance 1 and distance 2 data, respectively. In that case the models can be distinguished by variable flat durations in the identified stages. If on the other hand, different transformation distances involve different processing stages then we should see varying bump magnitudes between both datasets. Our preliminary results showed very similar stages for the first 5-6 processing stages, which were consistent across analyses and thus differences may be expected in the later stages.

Assessing Pure Insertion with HSMMs

In this study we identified differences in the encoding of the instruction between the experimental conditions. The fourth stage was slightly longer in C1 compared to C2-C4. However, these differences were very small (on average ~15ms at most). Provided that our data consisted of 10ms samples, we would not treat this as conclusive proof upon which we would reject pure insertion. Comparisons of the inserted stages (C2 vs C4 and C3 vs C4) also yielded significant differences between conditions, specifically for stage 6 and stage 9 for C2 vs C4 and stage 7 and 11 for C3 vs C4. Nevertheless, again the differences cannot be seen as conclusive evidence for a violation of pure insertion. For one, due to a lack of data it is not certain whether these are indeed valid bump models of each condition. In particular bumps specific to the inserted transformation were quite variable in bump topologies and flat latencies. Furthermore, the stages which showed the most significant differences were the last of the inserted stages. For C2 vs C4 this is a problem because this stage in C2 probably also includes the processing durations of a motor preparation and motor response, as those bumps are still missing in the model. In the comparison C3 vs C4 on the other hand, while the topologies seem similar, it is not clear if the 11th stage in C4 is really related C4 to the inserted retrieval stages. It seems implausible that the inserted retrieval stages would be relevant after the transformation (as indicated by the intermediary inserted bumps specific to the transformation) seemingly has occurred. The stage equivalent stage in C3 seems to be related to response preparation and response. While differences in the duration of this stage may be conceivable between both conditions it is not clear why they should quadruple in C4 compared to C3. It is more likely that part of this stages duration in C4 is still related to the transformation as the last of the inserted transformation stages is as mentioned notably shorter in C4 than C3. Perhaps even part of the response preparation of the transformation condition is included here. However, overall without sufficient data to properly determine the validity of our models it would be speculative to reject pure insertion based on the HSMM results.

While in this study we were unable to conclusively assess pure insertion using HSMMs, our results nonetheless revealed how inherently problematic experimental designs that rely on this assumption are. In one of

ASSESSING PURE INSERTION USING HSMM

our conditions we intended to insert a transformation process and ended up with multiple processing stages. Interactions between transformation distance and retrieval according to the additive factors method indicate that pure insertion probably does not apply to this task. While we were not able to show this with our HSMM analyses, nonetheless we could determine that differences in reaction times reflected the presence of multiple processing stages instead of one. This highlights the importance of methods such as the HSMM, which rely on fewer assumptions and allow a more direct assessment of cognitive processing stages. While the HSMM analysis of this task may not have been successful in all respects, it yet is also superior to the nonparametric ERP analysis we performed. The results of the ERP analysis appear to mainly capture differences that relate to a retrieval operation. Only comparisons including the insertion of a retrieval were significant and the identified latencies correspond to the times that the retrieval stage we identified supposedly may have been active. More subtle processes of the transformation operations could not be detected. Further, as mentioned previously without supplementing the results of our HSMM analysis it would be difficult to draw conclusion about the functionality of the ERP differences.

A great example of how HSMMs can be used to assess if a task is adhering to pure insertion was recently published study by Zhang et al. (2018) Their task required participants to perform mathematical operations (inserted stages consisted of calculation of a product or the substitution of an unknown variable) with a specific response for each solution. They identified a pre-attention stage, an encoding stage, a response-preparation and a response stage to be present within all conditions. The HSMM result revealed an interaction between insertion and response preparation as well as an interaction with response, which went unnoticed in an ERP and RT analysis, thus rejecting pure insertion. A reason for their clear results may have been: a) a larger sample size as the average accuracy across conditions was around 90%, b) fewer processing steps as average response times per condition were at most 1.57 seconds, and the bump solution with the most processing steps being a five bump model. Hence, while HSMMs can be successfully used to assess pure insertion, it is important to establish a measure of power for a given task beforehand to enable accurate detection that considers both the expected amount of viable data for analysis as well as the expected length/number of processing steps in a task.

Limitations

A limitation of using a HSMM approach is that it is restricted to sequential latent cognitive stages. An HSMM does not allow for processing to be occurring in parallel. In our case most likely the task design protects from processes to occur in parallel with each stage requiring the process of the previous stages to be finished. To be ready for a response in simplest task means visual information must have been encoded and understood. The same is true for the retrieval stage and the transformation stage. Also in C4, transformation could not possibly

ASSESSING PURE INSERTION USING HSMM

occur before the retrieval of the association was finished. Hence, in this experiment this is not considered to be a problem. Nonetheless, depending on the task it is useful to consider that a sequential explanation for the modelling results of the HSMM is not the only viable account. In case of an ambiguous task design, other approaches may be more lenient with regard to the seriality assumption of cognitive processes and allow for processes to occur at the same time (Hawkins et al., 2017).

Another problem of the HSMM methodology that should be discussed is that the model validation criteria currently used was not ideal and depends largely on sample size. Ultimately to be able to compare conditions meaningfully, we decided to use the LOOCV results only as guidance. We selected for each decision a more complex model to be able to accommodate encoding stages shared across conditions. However, the temporal resolution of a winning model as determined by LOOCV significantly depends on the sample size of the dataset as exemplified in the higher temporal resolution of a much simpler cognitive process (visual encoding) compared to more complex operations such as rearranging the three names. Nevertheless, there are also other means to validate identified bump solutions such as AIC, BIC etc. All of these methods have advantages but also disadvantages. For instance, using AIC to validate the number of stages in a Hidden Markov model analysis has the tendency to overestimate the number of actual states according to a simulation study by Pohle, Langrock, van Beest, & Schmidt (2017). Their results further suggest that the ability of a validation method to identify the correct model is largely dependent on the amount of data that is available. All estimation methods identified the correct model if sufficient data was available. In our case, when there is little data available for the LOOCV per participant (C4 – some datasets included less than 30 trials), it might be preferable to compare the results of different validation criteria.

Moreover, the missing response- and in C2 also response preparation stages are problematic. We attribute this as previously mentioned to the lack of data in the experimental conditions. The number of stages that one can discover heavily depends on the sample size. This is best exemplified by our HSMM analysis of the encoding of the three names. Although the encoding of a name is arguably less complex than reading an instruction, retrieving an association from memory and preparing and performing a response (C3), the encoding of a name only consisted of one stage less than C3. This can be attributed to having more than three times the sample size for the encoding of the names than C3, which allows for a higher temporal resolution in the processing steps involved. With a larger sample size we could increase also the number of stages we can discover in our experimental condition and it seems likely that we would also find the currently missing response stages. At present in the absence of a formal methodology, we would recommend for this task more than 90 trials per participants per condition (the average amount of trials in C1, where we could not identify a response stage)

ASSESSING PURE INSERTION USING HSMM

up to around 300 trials. The latter number stems from the average number of trials per participant for the name presentation task where we successfully were able to identify a model using LOOCV that contained in total 17 stages. However, this does not consider the actual duration of each stage so for a complex task with processing stages that have long durations even more trials may be required.

An issue whose importance cannot be stressed enough is the question whether the signatures of the processing stages we identified actually are the processes that we were initially interested in. It is not certain whether the topology of bump X indeed is the corresponding neural covariate of process X or just a proxy of neural instantiation of process X . This is not a question that can be conclusively answered by this study alone. In particular it would be interesting to see if the HSMM method can recover the same processing stages in a MEG dataset, which records activation from a different orientation. A recently published MEG analysis by Anderson et al. (2018) of a previously used associative recognition task (e.g. Anderson et al., 2016) gives us hope in that regard, as they were able to identify the same stage structure as the EEG analysis.

Future research

In the future it would be interesting to identify bump signatures that are shared across tasks such as encoding and response. Perhaps it is possible to reduce the degrees of freedom in the parameter estimation in the HSMM analysis when the magnitudes of the first and last stages (known to be present in the task) are already known and only additional stages have to be estimated in the best case. For the very least, establishing such criteria would help evaluate the results of the LOOCV model estimation. For instance, the retrieval stage identified in this task does not correspond to retrieval stages identified in other tasks. The familiarity interpretation based on the similar FN400 ERP does at present not make clear why this distinction is there. Further, for other processing stages in particular for more complex tasks later in the trial there may be no ERP data available. Latency jitter in the onset of cognitive stages may obscure its presence in the average signal (Ouyang, Sommer, & Zhou, 2016).

Conclusion

Using HSMM analysis we were unable to conclusively assess the assumption of pure insertion in a complex information processing task. Nonetheless, our study highlights how HSMM analyses are able to detect subtle differences in complex tasks that cannot be identified with conventional EEG analysis methods.

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. *Psychological Science*, , 0956797618774526.
- 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., Qin, Y., Jung, K., & Carter, C. S. (2007). Information-processing modules and their relative modality specificity. *Cognitive Psychology*, 54(3), 185-217.
- Anderson, J. R., Zhang, Q., Borst, J. P., & Walsh, M. M. (2016). The discovery of processing stages: Extension of sternberg's method. *Psychological Review*, 123(5), 481-509. doi:10.1037/rev0000030
- Borst, J. P., & Anderson, J. R. (2015). The discovery of processing stages: Analyzing EEG data with hidden semi-markov models. *NeuroImage*, 108, 60-73.
- Donders, F. C. (1868). Over de snelheid van psychische processen. *Onderzoekingen Gedaan in Het Physiologisch Laboratorium Der Utrechtse Hoogeschool, Tweede Reeks II*,
- Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, 30, 412-431.
doi:[https://doi.org/10.1016/0001-6918\(69\)90065-1](https://doi.org/10.1016/0001-6918(69)90065-1)
- Finnigan, S., O'Connell, R. G., Cummins, T. D., Broughton, M., & Robertson, I. H. (2011). ERP measures indicate both attention and working memory encoding decrements in aging. *Psychophysiology*, 48(5), 601-611.
- Friston, K., Price, C., Fletcher, P., Moore, C., Frackowiak, R., & Dolan, R. (1996). The trouble with cognitive subtraction. *NeuroImage*, 4(2), 97-104. doi:10.1006/nimg.1996.0033
- Hawkins, G. E., Mittner, M., Forstmann, B. U., & Heathcote, A. (2017). On the efficiency of neurally-informed cognitive models to identify latent cognitive states. *Journal of Mathematical Psychology*, 76, 142-155.
- Hillyard, S. A., & Kutas, M. (1983). Electrophysiology of cognitive processing. *Annual Review of Psychology*, 34(1), 33-61.
- Ilian, A. B., & Miller, J. (1994). A violation of pure insertion - mental rotation and choice-reaction time. *Journal of Experimental Psychology-Human Perception and Performance*, 20(3), 520-536. doi:10.1037//0096-1523.20.3.520
- Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, 4(11), 432-440.

ASSESSING PURE INSERTION USING HSMM

- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177-190.
- McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*(4), 287.
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Reason*, *4*(2), 61-64.
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 1.
- Ouyang, G., Sommer, W., & Zhou, C. (2016). Reconstructing ERP amplitude effects after compensating for trial-to-trial latency jitter: A solution based on a novel application of residue iteration decomposition. *International Journal of Psychophysiology*, *109*, 9-20.
- Pohle, J., Langrock, R., van Beest, F., & Schmidt, N. M. (2017). Selecting the number of states in hidden markov models-pitfalls, practical challenges and pragmatic solutions. *arXiv Preprint arXiv:1701.08673*,
- Portoles, O., Borst, J. P., & van Vugt, M. K. (2017). Characterizing synchrony patterns across cognitive task stages of associative recognition memory. *European Journal of Neuroscience*,
- Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257-286.
- Sackur, J., & Dehaene, S. (2009). The cognitive architecture for chaining of two mental operations. *Cognition*, *111*(2), 187-211. doi:10.1016/j.cognition.2009.01.010
- Shah, A. S., Bressler, S. L., Knuth, K. H., Ding, M., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2004). Neural dynamics and the fundamental mechanisms of event-related brain potentials. *Cerebral Cortex*, *14*(5), 476-483.
- Smid, H., Fiedler, R., & Heinze, H. J. (2000). An electrophysiological study of the insertion of overt response choice. *Journal of Experimental Psychology-Human Perception and Performance*, *26*(3), 1053-1071. doi:10.1037//0096-1523.26.3.1053
- Stafford, T., & Gurney, K. N. (2011). Additive factors do not imply discrete processing stages: A worked example using models of the stroop task. *Frontiers in Psychology*, *2*, 287.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of donders' method. *Acta Psychologica*, *30*, 276-315.

ASSESSING PURE INSERTION USING HSMM

- Tsivilis, D., Allan, K., Roberts, J., Williams, N., Downes, J. J., & El-Deredy, W. (2015). Old-new ERP effects and remote memories: The late parietal effect is absent as recollection fails whereas the early mid-frontal effect persists as familiarity is retained. *Frontiers in Human Neuroscience*, *9*, 532.
- Ulrich, R., Mattes, S., & Miller, J. (1999). Donders's assumption of pure insertion: An evaluation on the basis of response dynamics. *Acta Psychologica*, *102*(1), 43-75. doi:10.1016/s0001-6918(99)00019-0
- World Medical Association. (2013). World medical association declaration of helsinki: Ethical principles for medical research involving human subjects. *Jama*, *310*(20), 2191-2194. doi:10.1001/jama.2013.281053 [doi]
- Zhang, Q., van Vugt, M., Borst, J. P., & Anderson, J. R. (2018). 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. *Journal of Cognitive Neuroscience*, *29*(2), 352-367.
- Zhang, Q., Walsh, M. M., & Anderson, J. R. (2018). The impact of inserting an additional mental process. *Computational Brain & Behavior*, , 1-14.

APPENDIX

$$\text{Log}_{RT} = 6.73 + 0.57\text{Retrieval} + 0.8\text{Distance1} + 1.1\text{Distance2} - 0.09\text{Retrieval} * \text{Distance1} - 0.15\text{Retrieval} * \text{Distance2}$$

Equation 1. Best model with random effects for participant explaining reaction time in every trial.

$$\text{Acc} = 3.2 - 0.99\text{Retrieval} - 0.72\text{Distance1} - 1.9\text{Distance2} - 0.48\text{Retrieval} * \text{Distance1}$$

Equation 2. Best model with random effects for participant explaining accuracy in every trial.

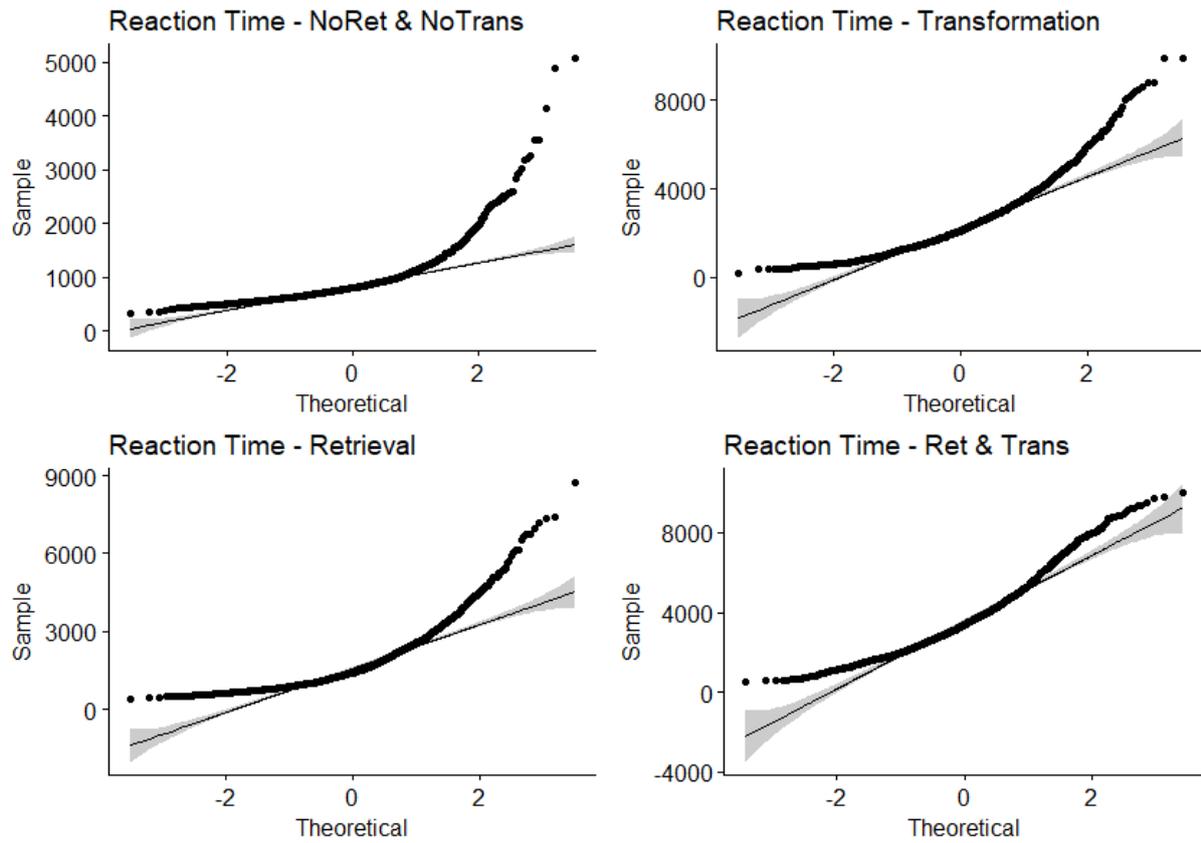
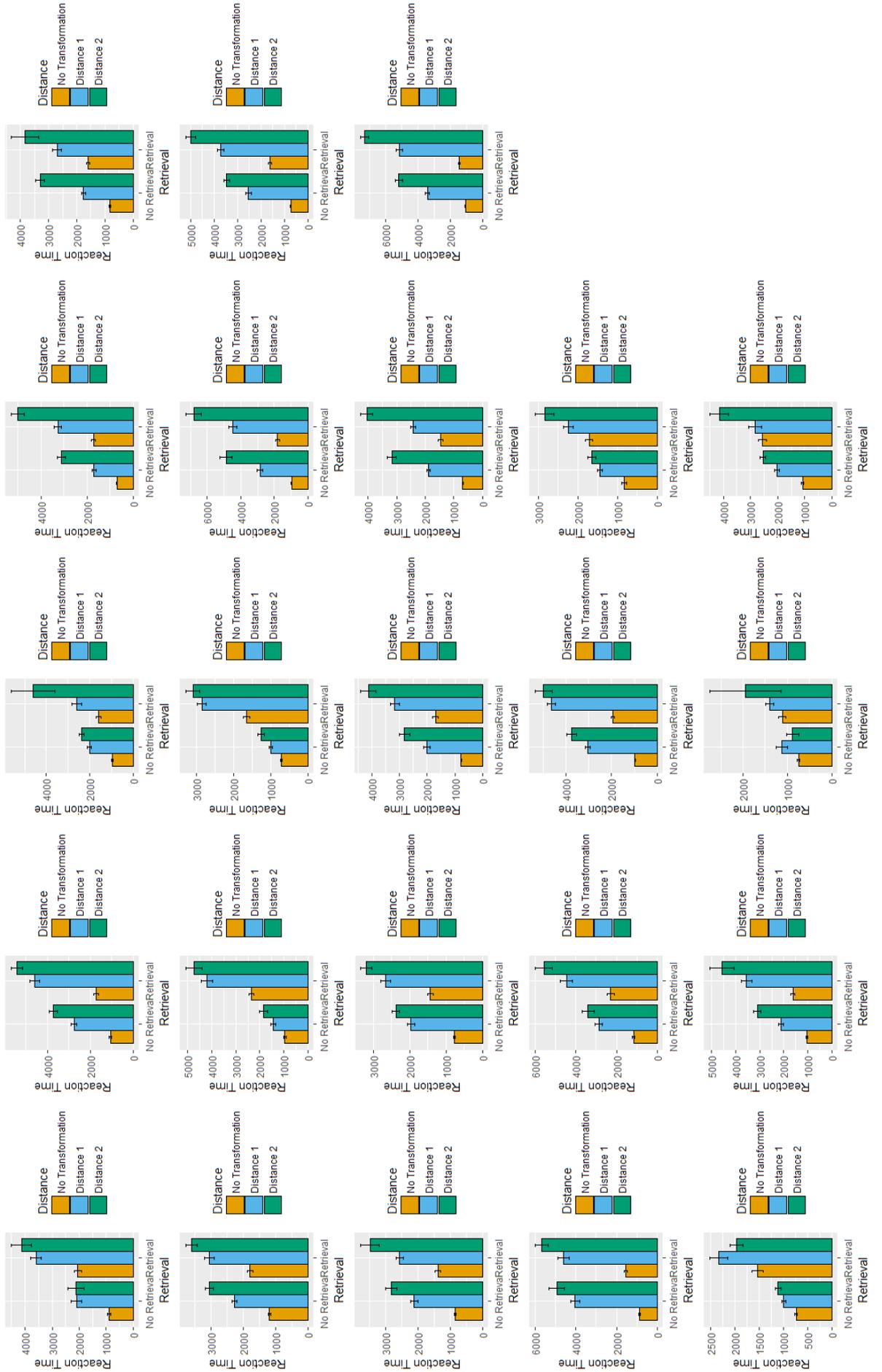


Figure 9. QQ-Plots of the data distribution for each experimental condition, indicating a violation of normality.

ASSESSING PURE INSERTION USING HSMM



ASSESSING PURE INSERTION USING HSMM

Figure 10. Reaction time results by retrieval and transformation distance for each participants individually. In the top graphs the effects with regard to reaction time are described, with average reaction time in ms on the y-axis and retrieval effect on the x-axis.

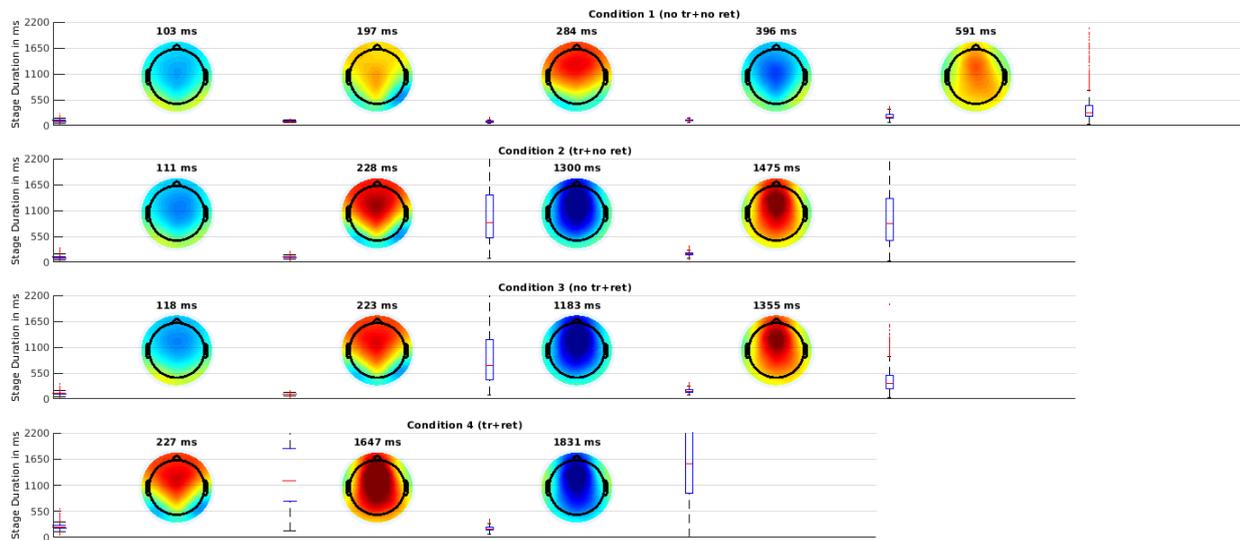


Figure 11. Best Bump configurations of each experimental condition from instruction display to response as determined by LOOCV.

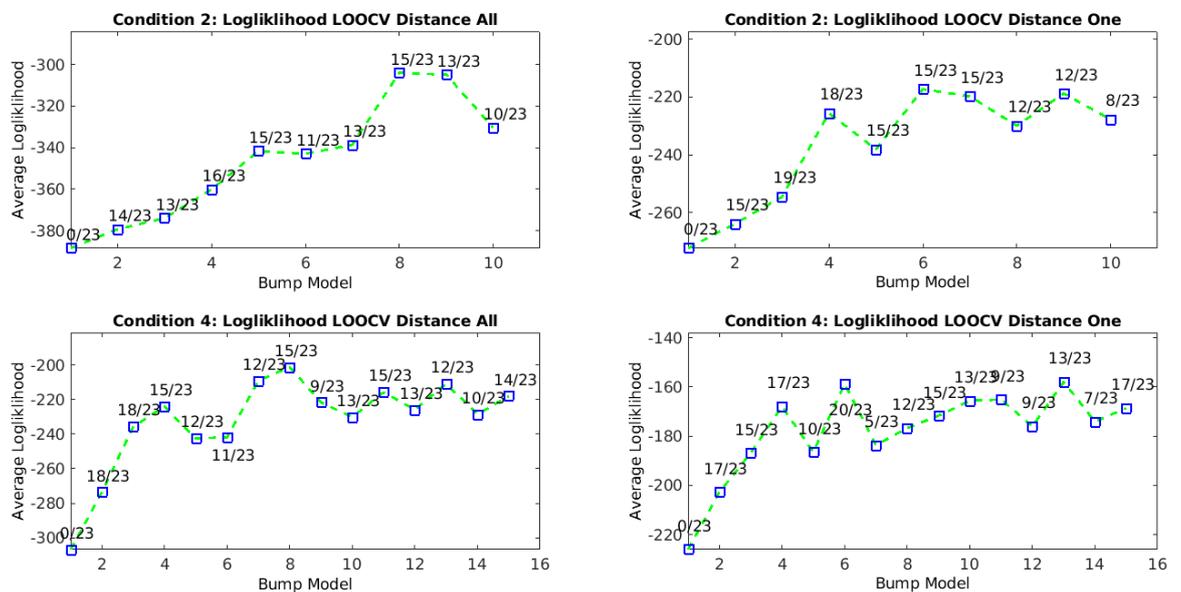


Figure 12. LOOCV results of all bump configurations for experimental conditions containing a transformation stage (C2 and C4). On each y-axis is the average log-likelihood each bump configuration. On each x-axis we see the respective number of bumps of each model (note C4 was tested up to 15 bumps vs. 10 bumps in C2). For each bump configuration we see for how many participants adding an extra bump increased the likelihood significantly compared to the previous bump configuration. The graphs on the left side contain data of all transformation distances, whereas the graphs on the right side contain only data of distance 1 transformations.

ASSESSING PURE INSERTION USING HSMM

IN – 12

OR – 13

AT – 14

DO – 21

OF – 23

BY – 24

BE – 31

AN – 32

WE – 34

TO – 41

HE – 42

IT – 43

Instruction sheet 1. Name Letter combinations to be learned by the participants.