university of groningen / faculty of science and engineering / artificial intelligence

# Eyes, Window to the Mind: Tracing the Structure of Cognitive Processing Stages in Associative Recognition using Pupil Dilatation Deconvolution

Bachelor's Project Thesis

Mara Nedelcu, s4245350, m.nedelcu.1@student.rug.nl,
Supervisors: Prof Dr J.P. Borst & J. Krause

**Abstract:** Associative Recognition is a cognitive task that can help shed light on the stages required for information processing. It has been studied from multiple perspectives, yet there is still no clear definition for the exact structure of its processing stages. Additionally, the manner in which the intensity level of mental effort impacts different stages has not yet been dealt with. It is known that the human pupil provides information regarding cognitive processes, its size increasing with mental effort. Here we applied a new approach employing Hidden semi-Markov Model combined with a Generalized Additive Mixed Model (HsMM-GAMM) to perform pupillary deconvolution on the data from an associative recognition experiment. Based on the results, we proposed an additional model of associative recognition. Results show a significant effect of fan on both response time (RT) and error rate (ER), while the effect of probe type is only significant on RT. Employing HsMM-GAMMs in the pupillary response analysis to create a new model for associative recognition provides insight into the variability and duration of the processing stages of the cognitive task through the lens of mental effort. The resulting model that proved to be the most probable is comprised of six stages and is closest in behaviour to a previously proposed MEG model. While recovering a sensible stage structure, the HsMM-GAMM model is not without fault. The greatest downsides of the chosen architecture were its lack of specificity, which influenced the manner in which it assigned intensity level of mental effort to specific stages, as well as its modelling of some individual stage responses as having negative amplitudes.

## 1 Introduction

The stage structure underlying human cognitive processing of inputted information has been a matter of interest and debate in the scientific community for more than a century (Donders, 1868/1969; Sternberg, 1969). Donders (1868/1969) was the first to consolidate the idea that information processing is a stage-based mechanism. Such mechanisms are apparent in numerous processes, namely solving a differential equation. There are certain steps — or stages — that need to be completed to reach the answer. Thus, Donders (1868/1969) brought into question three major issues concerning the stages underlying information processing: the matter of identifying them, their timing and the factors that impact them. He advanced an initial approach — the subtractive model — intended to formulate an answer to some of the questions. The model compares the Response Time (RT) of tasks hypothesized to share all but one stage, to infer the duration of said stage.

In his overview of the predecessors of experimental psychology, Boring (1929) made it clear that behavioural responses were found to consist of such stages. Reducing information processing to a mere succession of encoding, cognitive processing, and response elicitation, provides insight into the task consisting of at least three stages. Nevertheless, this is a rather rudimentary division of the process. A

more substantial approach was later formulated by Sternberg (1969), an approach that would eventually become the root for numerous other studies attempting to find answers to the questions formulated by Donders (1868/1969). Sternberg (1969) proposed an additive factors method highlighting that various manipulations on RT provide insight into the presence of different stages. For instance, manipulating two variables — or factors — that are independent of each other attests to the presence of two distinct stages that have been influenced by them. This becomes apparent with the difference in RT, which is a summation of stages' duration. Hence, the difference in RT between manipulations of the same added variables pinpoint discrepancies between the separate stages affected by each of them. Yet, RT manipulations alone simply bespeak the collective duration of all stages, lacking an actual delimitation between the individual ones (Borst, Schneider, Walsh, & Anderson, 2013).

As a solution, neuroimaging techniques have been used to distinguish between different stages, approximate their duration, and pinpoint their content and localization on the cerebral cortex (e.g. Anderson, Zhang, Borst, & Walsh, 2016; Borst, Ghuman, & Anderson, 2016; Sohn et al., 2005). As such, functional MRI (fMRI), electroencephalography (EEG) and magnetoencephalography (MEG) provide valuable input towards defining the structure of stages underlying information processing. Of these, MEG provides the highest temporal resolution and the most information regarding the cortical localization of such stages (Borst et al., 2016). Nonetheless, the intensity of mental effort elicited by different stages has not been the focus of these approaches, which paid close attention to delimiting the duration of different stages.

On the other hand, pupillary responses do provide insight into the intensity of mental effort required for a task, pupil size increasing with mental effort (Hess & Polt, 1964). More so, the pupil reacts to different events — or stages — generating a response to each (Hoeks & Levelt, 1993). Howbeit, the pupillary response is itself slow, its response to an event has a slow increase that peaks one second after the elicited event (Hoeks & Levelt, 1993). Pupil deconvolution breaks the pupil dilation time course into separate functions associated with these event-specific responses, underlining the intensity of mental effort associated with each stage (Hoeks & Levelt, 1993). Moreover, transforming the pupillary time course using pupil deconvolution provides a high-temporal resolution assessment of cognitive stages in information processing (Wierda, van Rijn, Taatgen, & Martens, 2012).

Yet, these past approaches to implementing pupil deconvolution fail to take into account differences between subjects or trials relative to the timing of the stages. The purpose of the current study is, thereafter, to fill this gap using a new approach (Krause, Borst, & van Rij, 2023) to performing pupil deconvolution. Together with pupillometry, it will be employed to analyse the intensity of mental effort apparent in the cognitive processing stages of associative recognition. The latter has been extensively studied in the past (e.g. Anderson et al., 2016; Borst et al., 2016), allowing for a pertinent comparison between the findings of the current study and previous ones. Concurrently, the study looks to evaluate the structure of the aforementioned stages as revealed by the human pupillary response.

## 1.1 Pupil Diameter Deconvolution

Pupil deconvolution has initially been proposed by Hoeks and Levelt (1993) as a method of dividing the pupillary time course into separate functions associated with each event-specific response elicited. Their model assumes each of these event-specific responses has an attentional pulse. These pulses then form a sequential system, each varying in temporal distribution, as well as amplitude. Additionally, the system has a constant-duration impulse response that is described by an Erlang gamma function. It, in turn, consists of the individual impulse responses of the attentional pulses, which are presumed to be identical, except for a particular amplification factor for each. Thus, the convolution of the inputted sequential system and the general impulse response output the predicted pupillary response assumed by Hoeks and Levelt's (1993) model. It takes the form of a vector having time as a dimension, and the pupil diameter's deviation from the baseline as values. The system is assumed to be linear, with the same event always generating the same pupillary response. Figure 1.1 captures the convolution process of the pupillary response (orange) from each of the event-specific responses (black). These events are each generated

from an attentional pulse (green) with a set amplitude and place in the time frame.

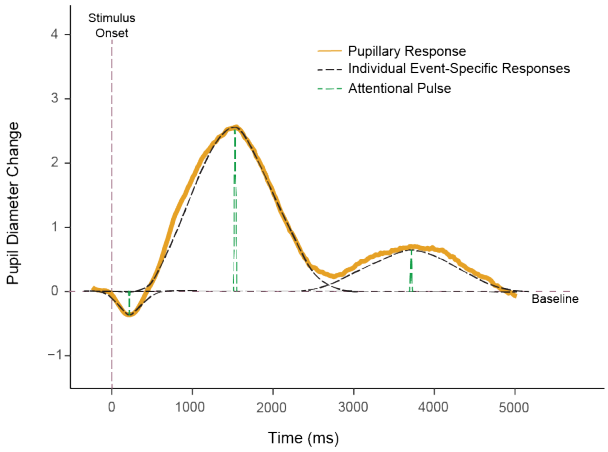


**Figure 1.1: The pupillary response (orange) as the convolution of event-specific responses (black). These responses are each generated from an attentional pulse (green). The pink lines mark the Stimuli Onset (vertical) and the baseline (horizontal).**

Furthermore, Wierda et al. (2012) extended the previous model to account for the pupil's delayed response to each event. They propose a model which provides a high-temporal resolution assessment of cognitive stages. To this end, they advance an optimization algorithm, which minimizes the difference between the recorded and predicted pupillary time courses. It is intended to capture the strength associated with the attentional pulse of each event, using the temporal locations vector of pulses. More so, the linear drifts in the data are accounted for with a slope which Wierda et al. (2012) estimate. This allows for longer pupillary responses to be analyzed. Concurrently, using the temporal locations of attentional pulses to determine the temporal resolution allows for an analysis that accounts for the delayed pupillary response.

Howbeit, both models compute the predicted pupillary response for different conditions using the averaged data over all subjects. This approach does not take into account that the same response to a specific event might occur at different moments for different subjects, or even for different trials. Additionally, the task chosen by Wierda et al. (2012) had stimuli be presented every 100 ms. This led to the potential events of every pupillary response being spaced every 100 ms, resulting in an overlap between the events of consecutive responses. With the data being averaged over all subjects, the aforementioned overlap is bound to have influenced the outcome of the model.

To account for these issues Krause et al. (2023) propose an approach that combines Hidden semi-Markov Models (HsMM) with Generalized Additive Mixed Models (GAMM). Hidden Markov Models (HMM) simulate a system which is always in one of its states. At times, the system transitions to a new state (Rabiner, 1989). As such, the system simulated by the model will be, in this case, the pupillary response, while the states are the event-specific responses or cognitive processing stages. The duration of each stage is variable, which motivates the use of the variable-duration HMM — HsMM (Yu, 2010). Concurrently, GAMM are function approximators which capture non-linear relationships between covariates (time) and signals (pupillary time course). These are used to reconstruct the shape of the pupillary time course from the different functions associated with each stage, generating the predicted pupillary response of each individual trial.

## 1.2 Associative Recognition

Associative recognition is a cognitive recall task which involves judging whether two items have been previously experienced together. The current study employs the task structure as proposed by Borst et al. (2013), where participants had to study a list of word pairs. This was followed by a recall task where they had to judge whether presented word pairs have been studied together, separately, or if they were studied at all. In order to provide an accurate response, participants needed to have both item and associative information. That is, a correct response required knowing whether the words were studied, and how these were studied, respectively. Namely, the response is impacted by the fan of the pair - the number of distinctive studied pairs the words in that specific probe appear in. As fan increases, RT has been found to increase as well (Anderson & Reder, 1999), thus the time for recalling a word increases with the number of episodic associations it has with other words in memory (Borst et al., 2013).

Numerous structural models of associative recog-

nition and its underlying stages have been proposed in the past couple of decades (Anderson, 2007; Anderson & Reder, 1999; Borst & Anderson, 2015; Borst et al., 2016). One that, as of late, has proven to yield a rather simplistic decomposition of the stages underlying associative recognition is the computational process model developed in the cognitive architecture Adaptive Control of Though-Rational (ACT-R; Anderson, 2007). While the most simple, this model has explained the patterns found in most data, offering a much more detailed analysis than allowed by either behaviour or fMRI (Borst et al., 2016). Thereafter, as seen in the top part of Figure 1.2, the ACT-R model assumes a sequential stage system to describe the cognitive mechanisms underlying associative recognition. That is, an initial encoding stage of the perceived stimuli, sensitive to physical features of the stimuli such as size. The encoded information is then used to perform associative retrieval, the second stage assumed by the ACT-R model. This stage is known to be influenced by the fan type of a pair. Seeing as, during the learning phase, the stimuli pairs are stored as sole information chunks in memory (Anderson & Reder, 1999), the model now retrieves from memory the closest match to the encoded pair. Once retrieval is done, the decision-making stage follows, where the participants determine whether the perceived stimuli have been studied together or not. This is done by comparing the stimuli pair to the retrieved pair from memory. As such, this stage is influenced by the probe type of the stimuli: target — studied pair —, re-paired foil — non-studied pair, but which consists of studied words —, or new foil — non-studied pair consisting of new words. Lastly, the ACT-R model assumes the motor response is the last stage of the process, occurring once the decision has been made.

The more complex EEG Model (Borst & Anderson, 2015) is constructed using a stage-discovery method employing HsMM on EEG data to identify specific stages and their duration through their neural signature. This allows for a more explicit analysis of the data than the inferences done on behavioural and fMRI data, on which the ACT-R model (Anderson, 2007) is based. As captured in the centre of Figure 1.2, the EEG model assumes an additional stage, compared to the ACT-R model. Familiarity does not consist of any information being retrieved from memory, but is a rather fast and automatic sense of how known — or familiar — an item is. This stage begins before encoding is complete and lasts until after the retrieval stage has begun. It is influenced by the early fan effect, pinpointing a swifter initial recognition of words appearing in higher fan probes. With this process

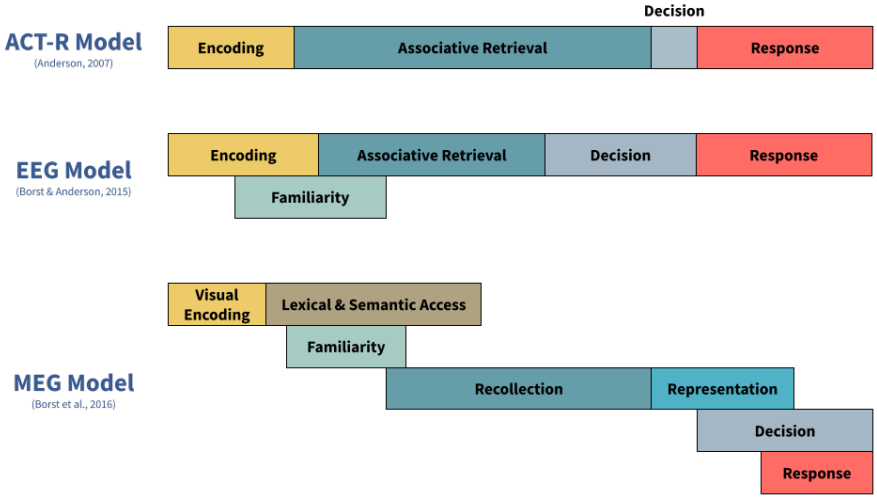


**Figure 1.2: The processing stages of associative recognition as described by the ACT-R model (Anderson, 2007) (top), the EEG model adapted by Borst and Anderson (2015) (centre), and the MEG model proposed by Borst et al. (2016) (bottom).**

beginning during encoding, words with a higher number of occurrences are rapidly recognised. This model also proposes the late fan effect to impact the retrieval stage, with probes of a lower fan being easier to recall.

The last model, the MEG one (Borst et al., 2016), assumes the most complex stage structure. The MEG data the model is based on provides both a high-temporal resolution (which fMRI data lacks), and a good spatial resolution (missing from EEG data). This allows for the spatiotemporal dynamics of associative recognition to be analysed using a multivariate classifier, along with non-parametric cluster analyses. The finer representations of the activity in the cortex, thus, justify the higher complexity of the MEG model. As captured in the bottom part of Figure 1.2, this model divides the encoding stage of the previous two models into two parts: Visual Encoding and Lexical & Semantic Access. The former is associated with the initial encoding of the visual stimuli as it appears on the screen, while the latter concerns processing the meaning of the words in the current probe. This stage is thought to start before familiarity, last throughout it, and end during recollection. The current recollection stage is the counterpart of the associative retrieval stages from the previous two models. An additional representation stage is assumed, where the recalled pair is stored during the decision stage, before the response. This pair is then compared to the stimuli during the decision stage. Ultimately, the aforementioned stage is now thought to last throughout the entire response stage.

### 1.3 Aim

This study will implement a HsMM-GAMM model for pupil deconvolution to analyse the underlying cognitive stage structure of associative recognition. Accordingly, it aims to observe whether the model implementing HsMM-GAMM will produce a similar stage structure to those observed in the previous associative recognition models (Anderson, 2007; Borst & Anderson, 2015; Borst et al., 2016). More so, it seeks to distinguish which of these models proves to be the most probable representation of the stage structure underlying the studied cognitive task. Furthermore, the effect of fan and probe type on the intensity of mental effort required in the stages of familiarity, recollection and decision-making will also be analysed.

The stage structure of Associative Recognition as reflected by the HsMM-GAMM model is hypothesised to reflect the structure captured in the previous three models, holding a similar number of stages. Additionally, fan and probe type, are also thought to impact the level of mental effort required by the stages of familiarity, recollection and decision-making. It is hypothesised that mental effort increases as the fan increases. Higher fan items are thought to be associated with increased pupil dilation as a response to the familiarity and recollection stages. On the other hand, it is hypothesised that mental effort increases when there is a match at the item level, but not at the associative level. It is believed that re-paired foils will be associated with increased pupil dilation, compared to targets, as a response to the recollection and decision stages.

## 2 Methods

### 2.1 Participants

This study gathered a total of 26 participants, howbeit, one had to be excluded as they were unable to complete the testing phase of the experiment. This was owing to the constant need for eye-tracker calibration resulting from the participant's coloured contacts. More so, one of the remaining participants failed to fill in the demographics data form. Of the remaining 24 participants, 11 were women, one preferred not to mention their gender and the rest were men. Their mean age was 21.46 with an age range of 20-26. 21 of them were right-handed, while the rest were left-handed. All participants were Bachelor's or Master's students at the University of Groningen and were approached personally for recruitment. All signed informed consent forms prior to participating in the study and received monetary compensation of 16 euros each for partaking in a single one and a half hours session.

### 2.2 Stimuli

The stimuli consisted of short (four or five letters long) noun word pairs divided into three distinct types, referred to as probes — targets, re-paired

foils & new foils. Targets were 16 word pairs, with eight pairs corresponding to each of the two fan conditions — Fan 1 & Fan 2. Thus, every word in a Fan 1 target pair exclusively occurred in that specific target probe, while all words in a Fan 2 target pair were coupled with one other Fan 2 word in a distinctive target probe. Likewise, re-paired foils consisted of 16 word pairs, equally split between the two fan conditions — eight for each. Fan 1 re-paired foils were generated by conjoining two words that appeared in different Fan 1 target probes. Similarly, Fan 2 re-paired foils comprised two words occurring in different Fan 2 target probes, each word appearing in two independent Fan 2 re-paired foil probes. Howbeit, new foils amounted to 104 word pairs, each word distinct from the ones used in targets and re-paired foils. All words in a new foil probe appeared only in one such pair. The targets were the only ones used during training, while all probe types were used during testing. In contrast to Borst et al. (2013), this study focused on short word pairs to account for the prospective confounding variable of eye movement that long word pairs ought to bring forth. Nevertheless, long word pairs were still used in the training phase, as well as the practice block of the testing phase. As such, an additional 16 long (seven or eight letters) word pairs were used during training. The practice block of the testing phase was designed with four of the long pairs learned, four pairs designed to mirror re-paired probes and two pairs with new words, matching the style of new foils. Withal, if all short word pairs were novel for each participant, the 24 long word pairs were unchanged across all. Being primely interested in the effect of probe and fan type, five conditions were determined by the short probes, in a 2 x 2 design — probe (target and re-paired foil) and fan (Fan 1 and Fan 2) type —, together with the one condition for new foils.

All probe types were generated from 260 words originally selected by Borst et al. (2013) from the MRC Psycholinguistic Database (Coltheart, 1981). Each of the words had a base imageability rating of 320 and a frequency between two and 97 occurrences per million. Of the total 260, 28 were long words, with a mean word frequency of 28 occurrences per million ($SD = 30.6$), a mean imageability rating of 509 ($SD = 86.1$) and a mean word length of 7.3 letters ($SD = 0.4$). The remaining 232 were short words, having a mean word frequency of 24.3 occurrences per million ($SD = 22.1$), a mean imageability rating of 539.3 ($SD = 55.3$) and a mean word length of 4.5 letters ($SD = 0.5$). The short words were randomly divided, for each participant, in a study list (of 24 words) and a new foil list (of 208 words), whereas the long words were divided once for all participants in the same two lists (24 in study and four in new foil). The two short word lists were matched for all participants on word length, frequency, and imageability based on six t-tests, all with $p > .1$, analogous to the method used by Borst et al. (2013). Each study list was generated such that all words began with a different three-letter sequence. Study lists were used to generate targets and re-paired foils, while the others generated new foil probes. While the first 20 pairs of short word lists were identical to those used by Borst et al. (2013), five new ones had to be generated for the remaining participants. The two long word lists were kept consistent across participants, being acquired from the generated word pair list of the first participant of Borst et al. (2013). Thus, all long word targets from Borst et al.' (2013) first participant list were used, while four pairs from their re-paired foil list and four from the new foils one were randomly selected.

## 2.3 Materials

Both phases of the experiment were implemented using the OpenSesame software (Mathôt, Schreij, & Theeuwes, 2012). All prompts had a white background, with a resolution of 1920x1080 px. All text in both phases was placed in the centre of the screen, written in black, size 18 px, using Arabic font for the word pairs in the training phase and Mono font for the instructions of the training phase and all throughout the testing one. Word pairs were presented at approximately 1.09 degrees of visual angle — both horizontal and vertical — for short pairs, and 1.75 horizontal degrees of visual angle and 1.09 vertical degrees for long word pairs, respectively. To ensure the consistency of the pupil dilatation samples, four bounds were set for the gaze location of the participant's left eye. Thus, if the left eye's gaze was outside of a 75 px wide square positioned in the centre of the screen, the experiment would trigger a calibration sequence of the eye tracker.

Turning to physical materials, if the training

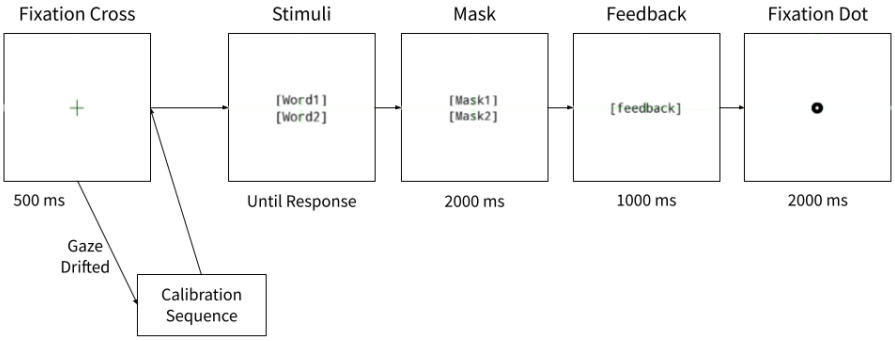


**Figure 2.1: The structure of a trial of the testing phase.**

phase simply required the use of a computer (desktop and keyboard), the testing phase made use of additional apparatus. In order to collect the eye-tracking data, the EyeLink Portable Duo was used to record the diameter of the participants' left pupil at a sampling rate of 500 Hz. During recording, participants placed their heads on a headrest to ensure that they would move as little as possible. The experiment was conducted in an eye-tracking lab at the University of Groningen. The room was free of any auditory or visual distractors, having no windows, the researcher and participant being the sole individuals inside during the experiment.

## 2.4 Procedure

The experimental procedure closely followed that of Borst et al. (2013), allowing for some changes that fostered the switch to an eye-tracking-based study. As such, each participant took part in one individual one-and-a-half-hour session for both the training and testing phase of the experiment. Before starting the experiment participants read an information sheet describing the course of the experiment, signed the informed consent form and filled in a demographic data form with their age, gender and handedness.

The training phase was split into two parts: acquisition and consolidation. During the acquisition phase, each of the 32 word pairs used during training were presented one word above the other for 5000 msec followed by a 500 msec blank screen. All participants were instructed to make an initial effort to remember the pairs when reading them. Once the acquisition phase was done, the consolidation one started. The latter consisted of a cued recall task where participants were prompted with all of the 48 words that appeared in the pairs studied during acquisition. Hence, one trial consisted of a randomly selected word from the target pool appearing on the screen, the participants being required to recall their one (for Fan 1 pairs) or two (for Fan 2 pairs) associated words. The response was bounded by no time constraints. If the correct response was provided, then the next trial started, otherwise, a 2500 msec feedback, that took the form of the word together with the expected response, appeared on the screen. More so, in case the response to the current trial was incorrect, the word appeared once again at the end of the block until the correct response was provided. One block consisted of the basis 48 trials and the repeated incorrect trials. The consolidation phase of the training consisted of three such blocks of trials, lasting for approximately 30 minutes.

Once the training phase was completed, the participants moved on to the testing phase, with the option to take a break between the two. Before starting the testing phase each participant assumed a comfortable position, placing their index and middle fingers on either the J and K keys — if they were right-handed or ambidextrous — or the D and F keys — if they were left-handed. Additionally, the eye tracker required one initial calibration prior to the beginning of this phase. The testing phase consisted of one practice block of ten trials and 13 testing blocks of 40 trials each, lasting for approximately one hour. If needed, participants could take breaks between each of the blocks. The practice block made use of long word pairs, two for each of the five conditions, that were kept the same for all

participants. Each testing block used short word pairs, eight for each of the five conditions, that appeared in random order. Thereafter, each condition corresponded to 104 of the total number of trials for each participant. All target and re-paired foil probes appeared once during each block — 13 times throughout the entire testing phase —, while eight randomly selected distinct new foil probes appeared in each block — each new foil pair appeared once throughout the testing phase.

The structure of a trial is captured in Figure 2.1. Thus, each trial began with a fixation cross positioned at the centre of the screen for 500 msec. In case the participant's left eye gaze was no longer in the preset bounds, the eye tracker would trigger a calibration sequence during the fixation. If such a sequence was triggered, the fixation period was repeated. Following the fixation, one word pair would appear on the screen — one word above the other. The participants were instructed to press either the J or F key — depending on their handedness — if the word pair had been studied during the training phase, or the K or D key — depending on their handedness — if the pair had not been studied. Withal, the participants were encouraged to respond promptly and accurately. The response triggered a hash mask — matching the length of both words — of 2000 msec, which allowed for the delayed pupillary response to the stimulus to be accurately captured before the feedback was presented (Wierda et al., 2012). The feedback — *Correct/Incorrect* — was presented for 1000 msec, followed by a fixation dot. The latter appeared for 2000 msec, to allow for the pupil dilation levels to return to baseline following feedback onset.

## 2.5 Pre-processing

From the behavioural testing data, all trials with an RT exceeding 20,000 ms were excluded for analysis. A trial having such a long RT was considered as containing unreliable data. Five such trials were excluded from the analysis. Of the remaining data, the RTs and Error Rates were analysed. Furthermore, all trials with an incorrect response were also excluded from the analysis of RT. Similar to the training data, the RTs and Error Rates of the testing phase were first aggregated within participants, and afterwards between participants.

To pre-process the eye-tracking data, the EDF files generated by the eye-tracker were converted to ASC files. This was done using the EyeLink Data Viewer software packages (SR Research). The ASC file format was then the one used for pre-processing. The eye-tracking data has been further pre-processed using the R framework (R Core Team, 2021), along with the following R packages: dplyr (Wickham, François, Henry, Müller, & Vaughan, 2023), eyelinker (Barthelme, 2021), plyr (Wickham, 2011), & PupilPre (Kyröläinen, Porretta, van Rij, & Järvikivi, 2019).

Hence, the eye-tracking data has initially been cleaned of unnecessary rows, aligned to the Stimulus Onset message, while incorrect trials were removed. Next, the blinks have been removed from the data. Lastly, the data has been baselined using normalization, with a window of -100 ms to 0 ms, and downsampled to 50 Hz. In preparation for the model, the data from all participants has been saved in different files, one for each condition, one containing all data and one with all data, except for new foil trials. The five trials excluded from the behavioural testing data for having RTs higher than 20,000 ms were also excluded from the eye-tracking data, as well as the incorrect trials. More so, for each trial only data from stimulus onset to 2000 ms after the trial's RT was included.

## 2.6 HsMM-GAMM model

The general model (Krause et al., 2023) fits a number of HsMM chains consisting of a set number of states for the system. These states represent the number of processing stages, or event-specific responses elicited by the pupil, that are assumed to occur during associative recognition. While the HsMM chains are fitted, GAMMs are used to approximate the shape of each of the event-specific responses. Each HsMM chain is fitted for a set number of iterations, so as to reach convergence. Out of all of the fitted HsMM chains of the model, the one that reached the highest log-likelihood at the last iteration was selected as the best one for that model, its results being used for analysis.

The ran model architecture assumes that each event-specific response has the exact same shape for all trials, all being gamma Erlang functions. What varies between trials for this model is that each event-specific response is shifted in time between trials. This fixed response shapes model has

been initially run for three to seven states, using five chains and 100 iterations. This architecture was run once for each number of states on the dataset containing all the data, except for new foil trials. Concurrently, four other models were run with the same number of chains and iterations for each number of states, on each of the four individual data sets of the conditions. Thus, five fixed response shapes models were run for each of the five numbers of states - general, Fan 1 targets, Fan 2 targets, Fan 1 re-paired foils, and Fan 2 re-paired foils. The four condition-specific models were run for each number of stages to verify whether these four or the general model were better predictors of the state durations and pupillary response. Figure 2.2 captures the predicted pupil dilation time course of one experimental trial, its event-specific responses, and its original counterpart as predicted by the general fixed response shapes model using five chains and 100 iterations for 6 states. Out of the five general fixed response shapes models, the one which had the mean predicted pupillary responses of each condition and the general response most similar to their original counterparts, was selected.

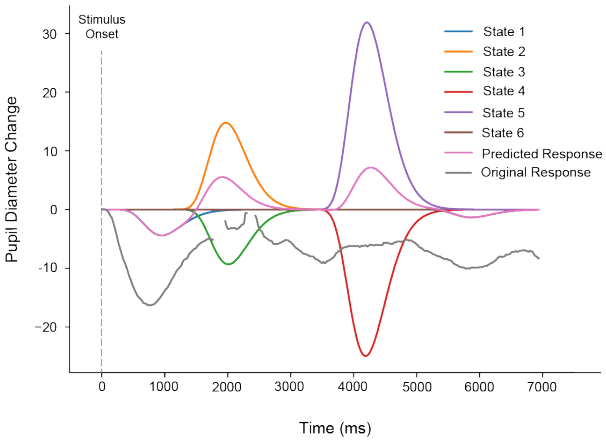


**Figure 2.2: The pupillary response of one trial with its predicted response and each of the event-specific responses as predicted by the general fixed response shapes model using five chains and 100 iterations for 6 states.**

# 3 Results

## 3.1 Behavioural

### 3.1.1 Training

As indicated by Borst et al. (2013), the frequency of the target probes being presented during the three blocks of cued recall of the training phase indicates the learning rate of the targets. The minimum frequency value is one, with each target appearing at least once during each of the training blocks – indicating that the participants learned a pair after a single display. The training data was initially aggregated within participants, and afterwards between participants, to compute the mean frequency per block for each of the four conditions — combinations of Fan and Length. Looking at the left-hand side of Figure 3.1, the mean frequency decreases with each Block for both Fan conditions, as well as both Length conditions. Figure 3.1 (left) also captures a difference in means between the two Fan conditions, ultimately highlighting a clear effect of Block and Fan, replicating Borst et al. (2013).

A repeated measures ANOVA having Fan, Block, and Length as factors was run on the aggregated training data, to verify the significance of the previously observed differences in frequency means. A significant large main effect of Block, $F(2,\ 48) = 52.29$, $MSE = 168.11$, $p < .001$, $\eta_p^2 = .69$, was captured in the training data, the mean frequency significantly decreasing with each block. A significant large main effect of Fan was also captured, $F(1,\ 24) = 52.17$, $MSE = 63.25$, $p < .001$, $\eta_p^2 = .68$, with Fan 2 targets having a significantly higher mean frequency than Fan 1 ones, which decreased across blocks. This decrease is in line with the significant large interaction between Fan and Block, $F(1,\ 24) = 29.55$, $MSE = 27.32$, $p < .001$, $\eta_p^2 = .55$. These three significant effects capture the learning process of the targets during the cued recall task of the training phase.

### 3.1.2 Testing

Similar to the training data, the RTs and Error Rates of the testing phase were first aggregated within participants, and afterwards between participants for analysis.

The centre of Figure 3.1 captures the mean RT for all conditions, with differences between the two

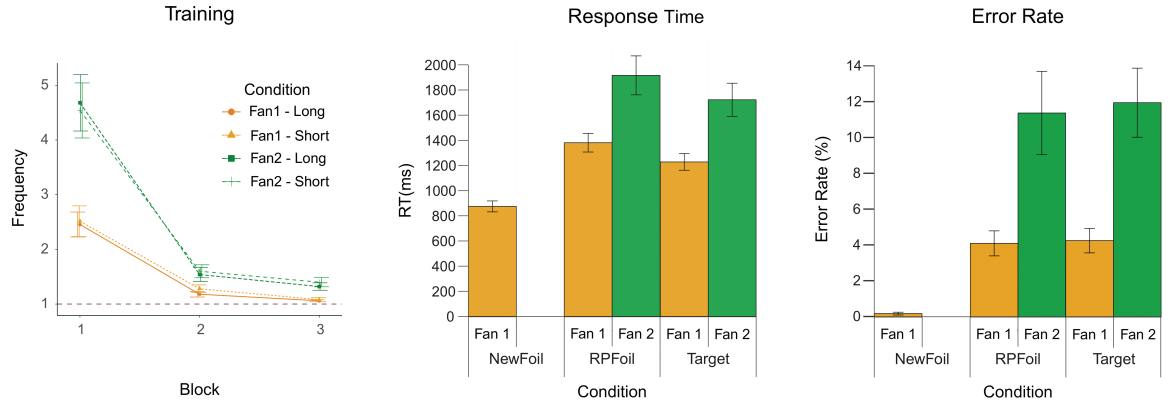


**Figure 3.1: Behavioural results of the training phase on the left, RT of the testing phase in the centre, and the Error Rate of the testing phase on the right. All error bars indicate Standard Error (SE).**

Fan conditions of both targets and re-paired foils. There are also differences between all three Fan 1 conditions and the two Fan 2 conditions, determined by the probe types. A repeated measures ANOVA with Fan and Probe Type as factors was run on the aggregated RT means of targets and re-paired foils to verify the significance of the previously observed differences. A significant main effect of Probe Type over the RT means was found, $F(1, 24) = 18.96$, $MSE = 767353$, $p < .001$, $\eta_p^2 = .44$. Having a large effect size, it is clear that RT for re-paired foils is significantly higher than for targets. The test also captured a main large effect of Fan over RT, $F(1, 24) = 26.64$, $MSE = 679501$, $p < .001$, $\eta_p^2 = .53$, Fan 2 probes being associated with significantly higher RT, compared to Fan 1 probes. This is in line with previous research on associative recognition (e.g. Borst & Anderson, 2015; Borst et al., 2016). However, unlike previous research, the interaction between Fan and Probe Type did not have a significant effect on RT, $F(1, 24) = 0.79$.

A repeated measures ANOVA with Fan and Probe Type as factors was also run on the aggregated Error Rates of targets and re-paired foils. Only a main large significant effect of Fan over Error Rates was captured, $F(1, 24) = 19.80$, $MSE = 1410.70$, $p < .001$, $\eta_p^2 = .45$, Fan 2 probes being linked to significantly higher Error Rates, than those connected to Fan 1 probes. The outcome of this test is also mostly in line with previous research, seeing as the main effect of Probe Type on Error Rate found by Borst et al. (2016) was only marginally significant.The right-hand side of Figure 3.1 clearly captures the highly significant difference between the two Fan conditions on Error Rates.

These behavioural results do still capture re-paired foils taking longer to be identified than targets, and Fan 2 items being more difficult to manage than Fan 1 items. Still, the effect of Fan on RT and Error Rate was proportional between Probe Types, no significant interaction being captured between the two Independent Variables (IVs).

## 3.2 Eyetracking

The pre-processed eye-tracking data were initially aggregated within subjects. From this data, one other aggregation procedure was performed. The data was aggregated within the five conditions of the study, the mean pupillary responses for each condition being captured in Figure 3.2.

Across conditions the qualitative pattern in pupil dilation is similar. To be more exact, they all start from stimulus onset with a small dip in dilation, followed by an increase in diameter that reaches its peak somewhere between 1000 and 2000 ms. After reaching the peak, all time courses show a decrease in pupil size and a tendency to return to baseline. The decline itself is the delayed pupillary response to the mask appearing on the screen, all declines occurring during the mask period. as indicated by Figure 3.2. The following increases, and dips, in pupil sizes seem to be generally similar across con-

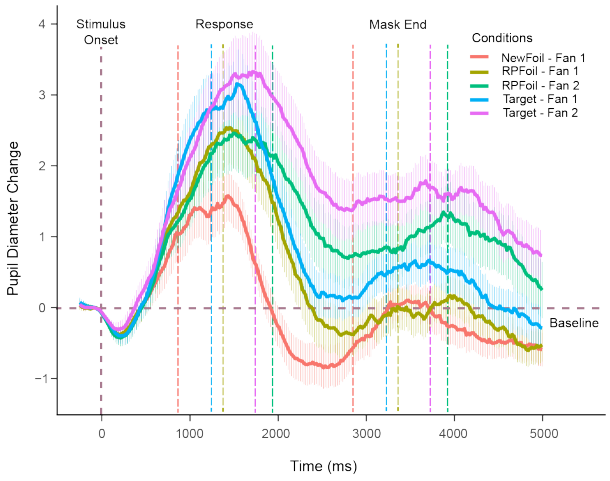


**Figure 3.2: The general mean pupillary response averaged over participants and conditions. All error bars indicate SE.**

ditions. With the main difference between them being their amplitudes, these can be linked to the anticipation of the feedback appearing on the screen. The differences in the amplitudes of the new peaks can be explained by all time courses starting the second ascent after about the same duration of time, very close to the end of the mask period. This duration being associated with a similar decrease in pupil size between conditions, the second ascent of each condition is relatively proportional to the amplitude of their first peak.

Focusing on the differences between the mean pupillary responses of the five conditions as seen in Figure 3.2, it is relatively clear that the mean response of the New Foils (orange) is quite different from the rest. It has the lowest peak out of all responses, which could be associated with the lowest intensity of mental effort required to provide a response to the stimulus out of all conditions. Its descent also starts rather early, compared to the other four, which can be expected, considering that the mean RT was also lowest for this condition.

Looking at the other four responses, some patterns can be observed in the behaviour of their first bumps in relation to the two IVs of the study: Fan and Probe Type. The two mean responses of the targets (pink and blue) have a faster and more abrupt initial ascent, than those two of the re-paired foils (green and olive). Additionally, the amplitude of the initial peaks of target time courses are considerably higher than those of the re-paired foils. This difference in amplitude highlights an overall higher level of intensity of mental effort being associated with providing a response to the targets, than to re-paired foils. Nevertheless, re-paired foils are typically perceived as the more complicated condition ( e.g. Borst et al., 2013). This increased amplitude in target peaks might also stem from cognitive processing stages being completed faster for targets than for re-paired foils, their individual responses being moved closer together and summed to form general response peaks with greater amplitudes. Thus, this observation cements the need to look at event-specific responses, rather than the general pupillary time courses to analyse the intensity of mental effort elicited in each condition.

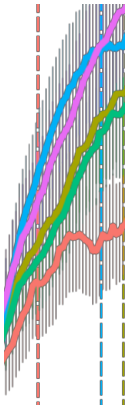

**Figure 3.3: Zoom in on the initial ascent of the condition mean pupillary responses captured in Figure 3.2.**

Turning to the differences between the two Fan conditions, Fan 1 responses (blue and olive) seem to have slightly faster and more abrupt initial ascents than those of Fan 2 responses (pink and green), as captured in Figure 3.3. Concurrently, Fan 1 responses have their descents distinctly earlier and much more abrupt compared to Fan 2 responses. This latter difference pinpoints that responses to Fan 2 pairs require a longer duration of increased mental effort compared to responses to Fan 1 pairs. Ultimately, the motor response provided by the participants mark different moments in the pupillary time courses of each Fan condition. Fan 1 responses appear closer to the beginning or center

of the initial peaks of the pupillary timecourses, while for Fan 2 responses, the start of the decline is marked by the participants providing their motor response to the stimulus.

## 3.3 HsMM-GAMM model

As explained in the Methods, five fixed response shapes models were run for three to seven states on a dataset containing all target and re-paired foil trials. The results are shown in Figure 3.4. It should be noted that all subplots have a turquoise stage at the beginning of each mean stage duration representation, which is associated with the duration elapsed between stimulus onset and the beginning of the first stage. Figure 3.5 captures the mean pupillary response shapes predicted by each of the models, as well as the original pupillary responses.

Looking at each of the five models in Figure 3.5, it is clear that the five- and six-stages ones seem to have the mean predictions closest to the original responses. The first two models have their predictions display behaviour that is far too general, making the predicted condition responses simply look like either variations of one Erlang gamma function — the three stages model —, or the convolution of two such functions — the four stage one. The seven-stage model, on the other hand, captures behaviour that is not even there, with an additional early smaller peak in the response at around 750 ms. Between the five and six stages models, the latter seems closer to the original responses, mirroring the elongated duration of the Fan 2 peaks. These differences can also be seen in Figure 3.4. The first two models have far too general of a stage duration distribution, the second state, starting far too late for all means. The general behaviour of their stage distribution stems from the models attributing the differences between conditions to random stages, or even incorporating more than one actual processing stage in one state. The seven-stage model displays an estimated stage duration structure that is, as expected, far too intricate. The start of the second stage now occurring a little too early, not even 100 ms after stimuli onset. In previous models of associative recognition, the second processing stage starts at least 200 ms after stimulus onset (Borst & Anderson, 2015; Borst et al., 2016). Finally, the five- and six-stage models appear to display the most plausible mean estimated state du-

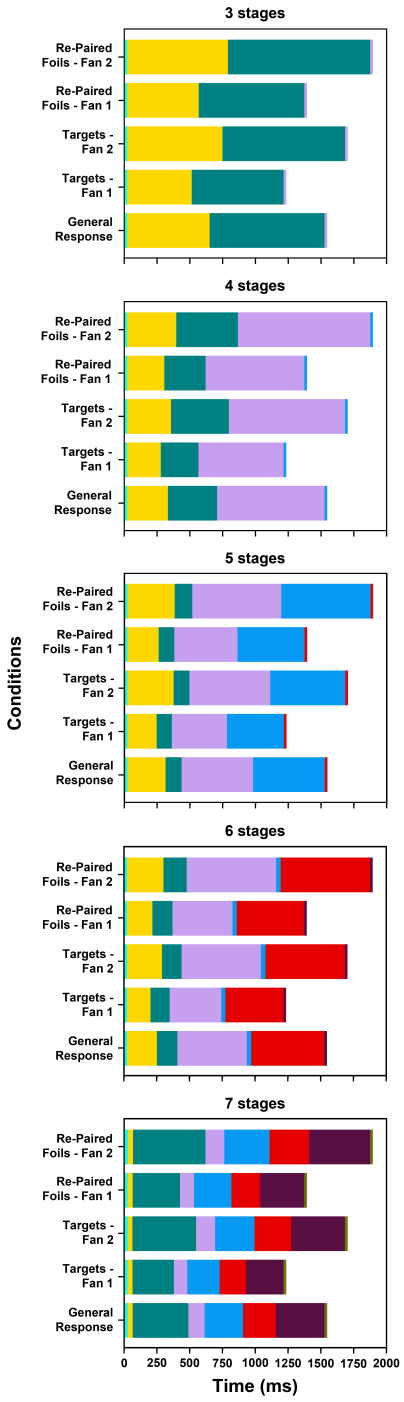


**Figure 3.4: The average duration of each state in the five fixed response shapes general models. These averages are presented for the general responses, as well as for each condition.**

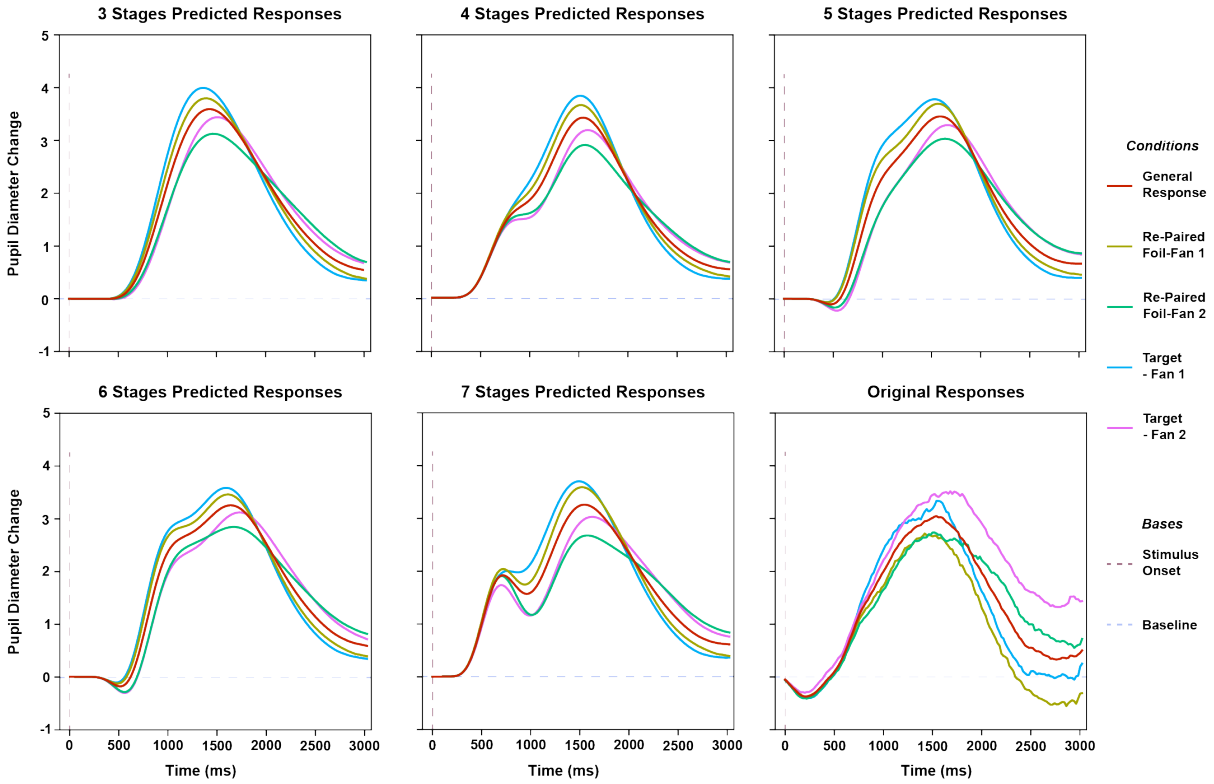


**Figure 3.5:** The mean pupillary responses as predicted by each of the five general fixed response shapes models. Each figure contains the mean response for each condition, as well as the general mean response over all trials. The figure on the bottom right-hand side contains the original mean pupillary responses, as obtained from the participants' data.

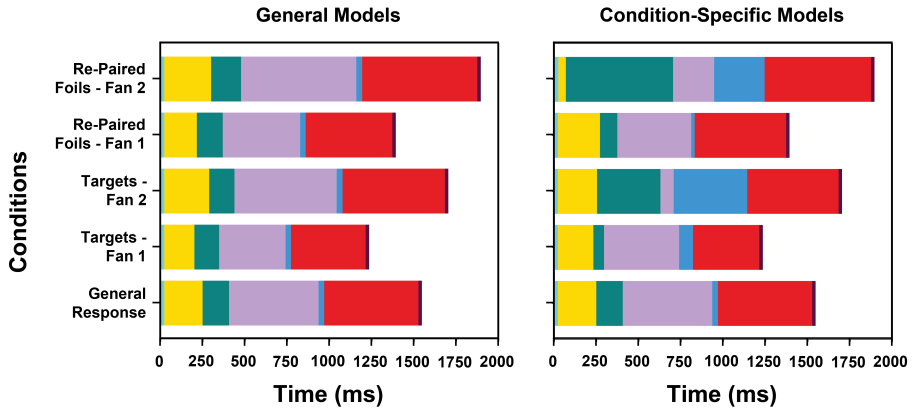


**Figure 3.6:** The average duration of each state in the six-stages fixed response shapes general models (left). These averages are presented for the general responses, as well as for each condition. The average duration of each state as captured by the six-stages fixed response shapes condition-specific models (right). The General Response bar is the same in both plots, being taken from the general model.

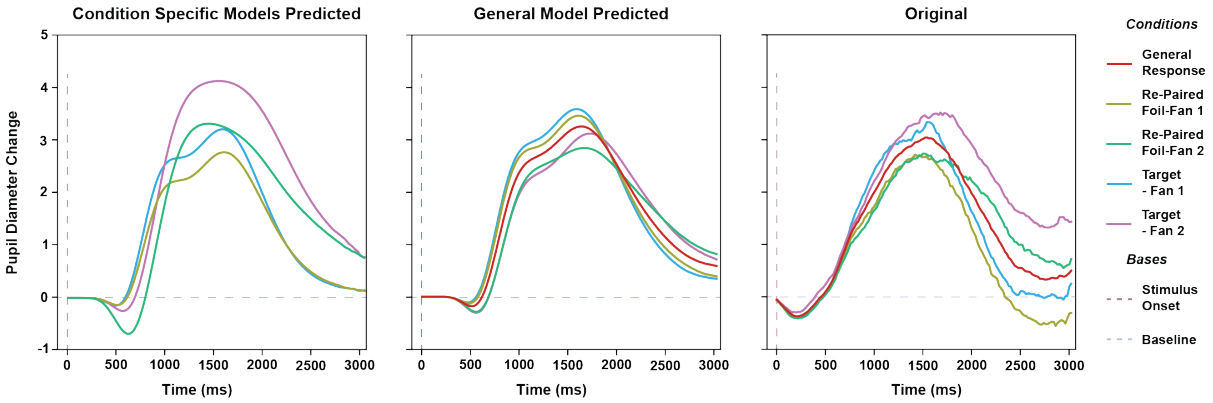


**Figure 3.7: The mean pupillary responses as predicted by the condition-specific fixed response shapes models for 6 stages (left). The mean pupillary responses as predicted by the 6 stages general fixed response shapes model (centre). The original mean pupillary responses, as obtained from the participants' data (right).**

| Stages | Effect | Repeated Measures ANOVA | | | |
|---|---|---|---|---|---|
| | | $F(1, 24)$ | $MSE$ | $p$ | $\eta_p^2$ |
| 1 | Fan | 15.78 | 502 | $< .001$ | .40 |
| 2 | Fan | 7.02 | 11.37 | .014 | .23 |
| | Probe Type | 7.64 | 17.75 | .011 | .24 |
| | Interaction between Fan and Probe Type | 6.29 | 5.46 | .019 | .21 |
| 3 | Fan | 29.25 | 3498 | $< .001$ | .55 |
| | Probe Type | 14.31 | 33.60 | $< .001$ | .37 |
| 4 | Fan | 3.57 | 0.56 | .071 | .13 |
| 5 | Fan | 25.36 | 1891.30 | $< .001$ | .51 |
| | Probe Type | 27.62 | 332.80 | $< .001$ | .54 |

**Table 3.1: The results of the repeated measures ANOVAs performed over the duration of the states of the general six-stage fixed response shapes model.**

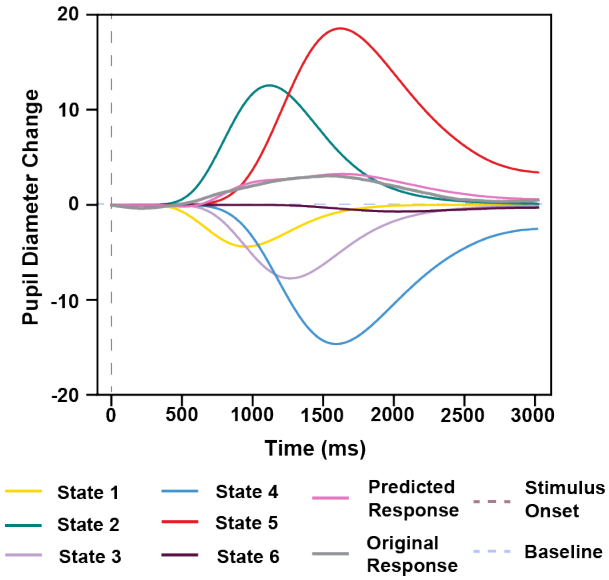


**Figure 3.8: The mean response elicited by each state in the six-stage general fixed response shape model, the predicted mean response and the original mean response.**

ration distributions. Between the two, the six-stage one seems, once again, to be the closest to previous associative recognition models, the five-stage one having the second stage — most probably associated with familiarity in both models — start relatively late around 300 ms.

Keeping that in mind, it seems the six-stage general model has the closest mean predictions to the original data, as seen in Figure 3.5. The mean predictions by condition of this model were also compared to those of the four fixed response shapes 6-stage condition-specific models. The comparison between the mean estimated states duration is captured in Figure 3.6. It is clear that the mean estimates of the condition-specific models do not capture any clear patterns between conditions, in the way the general model means do. This is not only the case for the six-stage condition-specific models but for all condition-specific ones, regardless of the number of stages. A figure capturing the state duration distribution of all general and condition-specific models can be found in the appendix (Figure .2). Turning to Figure 3.7, it can be seen that the condition-specific mean pupillary time courses capture the increased amplitude of Fan 2 peaks, compared to that of Fan 1 peaks, which the general model does not catch. However, this increase in Fan 2 peak amplitudes is quite exaggerated in the condition-specific models, both predictions having their peaks well above their corresponding original responses. The Fan 1 mean predictions are relatively similar between the condition-specific models and the general model, with the former capturing the difference between the two amplitudes slightly better than the latter. Nevertheless, the 6-stages general model seems to provide a far better overall estimation of the original results, than the condition-specific models. The mean pupillary responses from all condition-specific models are shown in a figure in the appendix(Figure .3).

Having selected the 6-stages general model as having the closest predictions to the original model, the significance of the differences between the duration of the states in each condition has been analysed. Six repeated measures ANOVAs have been run, having Fan and Probe Type as factors, one for the duration of each stage. The statistical results of these analyses can be found in Table 3.1. A significant main effect of Fan over the duration of the first stage was captured, Fan 2 pairs having a significantly longer duration of the first stage, compared to Fan 1 pairs. Over the duration of the second stage, a significant main effect of Fan, one of Probe Type, and a significant interaction between the two was found. The analysis of the third stage revealed a significant main effect of Fan, and a significant main effect of Probe Type. Additionally, a marginally significant main effect of Fan on the duration of the fourth state was found. Lastly, a significant main effect of Fan and one of Probe Type were found over the duration of the fifth state. No other significant effects were captured between the duration of the states in each condition.

The mean pupillary response elicited by each state as predicted by the six-stage general fixed response shapes model is highlighted in Figure 3.8. It can be seen that some states elicit responses with higher amplitudes than others, these states being, in theory, linked to a higher intensity of mental effort.

# 4 Discussion

## 4.1 General discussion

The present study aimed to implement a HsMM-GAMM model (Krause et al., 2023) for pupil deconvolution to analyse the underlying cognitive stage structure of associative recognition as revealed by pupil dilation. If the stage structure of associative recognition, and the temporal distribution of these stages, have been previously studied using neuroimaging techniques (e.g. Anderson et al., 2016; Borst et al., 2016; Sohn et al., 2005), the intensity of mental effort elicited by different stages has not yet been brought into focus.

Pupil size increases as a function of mental effort, different events eliciting different responses from the pupil (Hoeks & Levelt, 1993). As such, pupil size is believed to shed light on the level of intensity of mental effort evoked by the cognitive processing stages underlying associative recognition. Accordingly, in order to analyse the level of intensity of mental effort elicited by each stage, pupil deconvolution employing HsMM-GAMMs (Krause et al., 2023) was used to decompose the pupillary time courses. Additionally, associative fan and probe type were manipulated to provide further insight into how they affect the intensity level of mental effort elicited by different stages.

Before focusing on what the eye tracking analysis may reveal, one remark shall be made regarding the behavioural results. While these were all generally in line with previous studies, all RT means for the testing phase were higher than those recorded in the past (Borst et al., 2016, 2013). All of the means were about 200 ms higher than those recorded by Borst et al. (2016). This may prove to capture a certain delay in the temporal distribution of all processing stages, compared to the temporal distribution captured in previous associative recognition models. This delay may be a result of the inclusion of the mask as a trial component, which ultimately elongated the duration of the trials themselves as well. Concurrently, another explanation for the delay might have been the participants' anticipation of long-word pairs as probes, which were included in the training, but not in the testing phase. The possibility of encountering long-word pairs may have led the participants to be more cautious before responding, causing the overall delay in RTs.

## 4.2 Model

Five fixed response shapes general models were run to identify which number of stages (from three to seven) is most probable to describe the analysed cognitive task through the intensity of mental effort elicited by each stage. Twenty other condition-specific models were also run — four for each number of stages — to this end. The condition-specific models were found to lack any connection between the ones with the same number of stages in the patterns displayed by their stage structures. This effect may have been a result of the condition-specific models picking up on patterns or effects that were not actually there. As such, the aforementioned patterns were computed by the models to only fit the current model, and not the corresponding ones, with the same number of stages. Thus, the general models are believed to capture a more probable cognitive stage structure representation.

Among these fixed response shapes general models, the one modelling six stages is the one considered to be most probable. Its generated mean pupillary responses by condition, and mean general response, were the ones that looked most similar to their original counterparts. Once selected, the six-stage general model also revealed to have its stage duration distribution most similar to that found in previous models (Anderson, 2007; Borst & Anderson, 2015; Borst et al., 2016), cementing their validity.

A decomposition of the stage structure of the general response of the six-stage model can be seen on the left-hand side of Figure 4.1. It must be mentioned that the stages as captured by the HsMM-GAMM may last longer than captured in Figure 4.1. The model captures the stages as states that end once the next state begins. In actual fact, multiple stages are bound to occur simultaneously, following a dual-processing model of recognition (e.g. Rugg and Curran, 2007), as captured in previous studies (Borst & Anderson, 2015; Borst et al., 2016). Accordingly, the HsMM-GAMM model representation as seen on the left side of Figure 4.1, soundly captures the beginning of each cognitive processing stage, without properly marking their end.

Focusing further on the model's stage duration distribution, it seems to be most similar to that

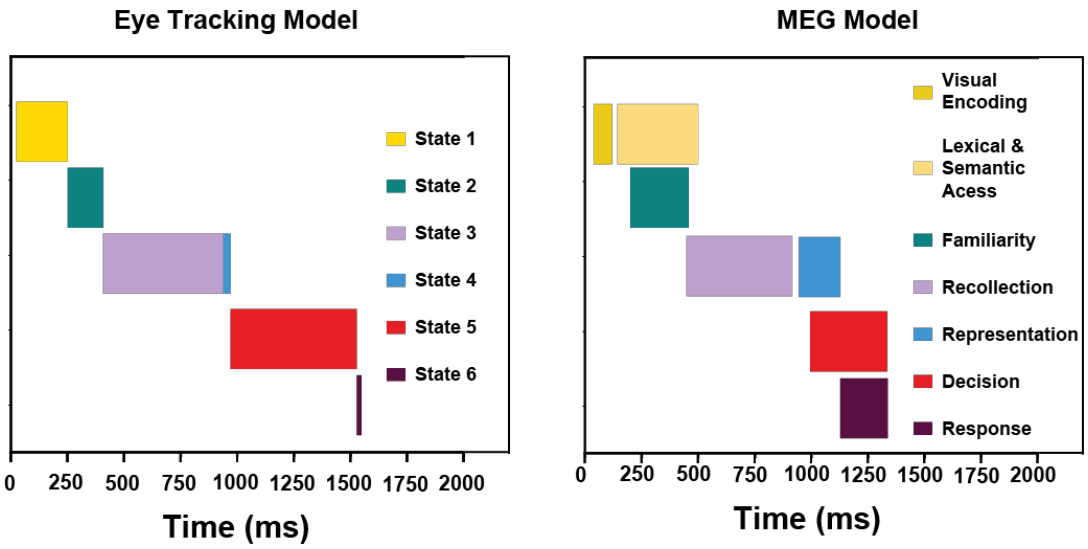


**Figure 4.1: The average duration of each state in the 6 stages fixed response shapes general model for the general response (left). The MEG associative recognition model proposed by Borst et al. (2016) (right).**

revealed by the MEG model (Borst et al., 2016), as seen on the right-hand side of Figure 4.1. Keeping in mind the overall delay in the duration of the eye-tracking model, its first stage seems to correspond with both the Visual Encoding and Lexical & Semantic Access stages of the MEG model. This can be seen as an overall Encoding stage of the eye-tracking model. The second stage of the eye-tracking model finds its correspondence in the Familiarity stage of the MEG model, while the third and fourth stages are clearly connected to Recollection and Representation. Thereafter, the last two stages of the eye-tracking model can be linked to the Decision and Response stages of the MEG model respectively.

While the two models can be associated, it is clear that there are still some discrepancies between the two. Aside from the overall delay in stage temporal distribution, these differences ought to arise since the eye-tracking one models the stage duration as a function of mental effort, rather than cortical activation, as modelled by the MEG one. This difference is most apparent when looking at the Decision and Response stages. Both Decision stages start at around 950-1000 ms, yet, the MEG one ends at about 1300 ms, while the eye-tracking one may last until 1500-1550 ms. Thus, the eye-tracking model shows the Decision stage to last longer than the MEG one, since the former models its stages as a function of the intensity of mental effort. On the other hand, the Response stage starts at about 1150 ms and ends at 1300 ms for the MEG model, while for the eye-tracking one, it starts at about 1500 ms and lasts until 1550 ms. The difference between the duration of the two stages in both models underlines the difference in the level of intensity of mental effort elicited by the two. As such, the Decision stage is linked to a high level of intensity of mental effort, while the Response stage relates to quite a low level one. The low level of intensity of mental effort elicited by the last stage of the eye-tracking model may further pinpoint that once the decision has been taken, enacting the motor response connected to it does not require much mental effort.

Looking at the differences between the duration of each state in each condition of the eye-tracking model, it is noticeable that all but two stages are significantly different between conditions, based on Fan or Probe Type, or even both (Table 3.1). Those two states are the fourth and sixth one, associated with the Representation and Response stages respectively. The fourth one did capture a marginally significant main effect of Fan, which reminds of previous findings. Borst et al. (2016) underlined

their found significant main effect of Fan to impact the cortical activation strength associated with the Representation stage based on the strength of the memory. Yet, it seems Fan does not quite significantly impact the duration of this stage, as captured by the HsMM-GAMM model, in relation to the intensity level of mental effort, with both Fan 1 and Fan 2 recalled pairs generally requiring the same level for representation. Following up, the Response stage seems to require the same low-intensity level of mental effort regardless of Fan or Probe Type.

## 4.3 Associative Fan

Fan was thought to impact the level of mental effort required by the Familiarity and Recollection stages, intensity level of mental effort increasing with the associative fan. However, the two Fan conditions have a significant effect on four stages of the eye-tracking model, as found using HsMM-GAMM — the Encoding, Familiarity, Recollection and Decision stages.

The effect of Fan on the first state of the eye-tracking model, associated with an Encoding stage, might appear surprising, considering the Visual Encoding and Lexical & Semantic Access stages were found to be influenced by the word length of the pair by Borst et al. (2016). This effect might actually be explained by the later start of the Familiarity stage for Fan 2 pairs, compared to Fan 1 pairs, instead of a longer-lasting Encoding stage for Fan 2 pairs.

The second state, associated with the Familiarity stage, is also impacted by the Fan conditions, with Fan 2 items underlining longer second states, compared to Fan 1 items. This effect correlates with the early-fan effect as highlighted by Borst et al. (2016), Fan 2 items being easier to recognise, yet with more pairs being activated, which might explain an increase in the duration of the stage as modelled by mental effort. More so, the significance of the interaction between Fan and Probe Type indicates that the Fan effect on the Familiarity stage was larger for re-paired foils than for targets. This significant interaction highlights that the early Fan effect lasts longer for re-paired foils than for targets.

The difference between conditions in the third state of the model, linked to the Recollection stage, is also shaped by the difference in the associative fan. The higher the fan, the longer the duration of the third state, that is, the longer the recollection of the associative information. With Fan 2 items belonging in more pairs, these take longer to be recollected. This effect is in line with the previously pinpointed late fan effect (e.g. Borst et al., 2016), which seems to have the same effect on the duration of the Recollection stage both when it is modelled as a function of cortical activation, and one of the intensity level of mental effort. This effect is also in line with the formulated hypothesis, stating that the duration of the recollection stage increases with the Fan.

The last state where the Fan has a significant effect is the fifth one in the eye-tracking model, the Decision stage lasting longer for Fan 2 pairs than for Fan 1 pairs. While previous studies have not found Fan to impact this cognitive processing stage (e.g. Borst et al., 2013; Borst et al., 2016), the fifth state of the HsMM-GAMM model is significantly impacted by it. Considering that the states of the model are not directly equivalent to processing stages, it is quite probable that this apparent effect of Fan on the fifth state is tied to the previously found Fan effect on the Representation stage (Borst et al., 2016). If this effect was only found to be marginal on the fourth state of the HsMM-GAMM model in this study, it could be that the strength of the memories, as impacted by the Fan, to be captured in the Decision stage of the eye-tracking model, rather than the Representation one.

## 4.4 Probe Type

Probe Type has been hypothesised to have an effect on the Recollection and Decision stages. Nevertheless, the analysis of the difference between conditions of the eye-tracking model found Probe Type to have an effect on one more additional stage, along with the two previously mentioned, the Familiarity stage. This stage was found to be influenced by Probe Type only when comparing new foils to targets/re-paired foils in the past (Borst & Anderson, 2015). However, new foil trial data was not included in this analysis, this stage being now seemingly influenced by Probe Type when comparing targets to re-paired foils.

The second stage of the current HsMM-GAMM model lasts longer for re-paired foils, compared to targets. This is not in line with previous findings,

where the Familiarity stage was only impacted by the associative Fan (e.g. Borst et al., 2013; Borst et al., 2016). Even so, it must be reiterated that the end of the second state does not necessarily mark the end of the Familiarity stage in the eye-tracking model. What it surely does, is mark the beginning of the Recollection stage. Therefore, this significant effect, might, in actuality, show that the Recollection stage starts earlier for targets than for re-paired foils. The former is easier to recognise, as target pairs have been shown more often, especially during training, when re-paired foils were not.

Accordingly, the third state of the model captured a Probe Type effect between the different conditions, with the third state lasting longer for re-paired foils, compared with targets. This further highlights that the Recollection stage, as modelled by the intensity level of mental effort, lasts longer for re-paired foils than for targets. Yet, the qualitative analysis of the original pupillary responses found targets to have higher amplitudes for the initial peaks. This may ultimately indicate that targets last less to be recalled, even if they take a higher intensity level of mental effort to be recalled than re-paired foils. Thereafter, the difference in duration is, in fact, in accordance with previous findings (Borst et al., 2013).

The effect of Probe Type on the duration of the fifth state of the HsMM-GAMM model, associated with the Decision stage, resembles the pattern found for the Recollection stage. Thus, re-paired foils display a longer-lasting duration of the Decision stage as modelled by mental effort, with a lower amplitude of the peaks in the original pupillary responses, compared to targets. Similar to previous findings (Borst et al., 2016, 2013), the decision for targets does take less time, yet it requires more mental effort.

## 4.5 HsMM-GAMM for Pupil Deconvolution

HsMM-GAMM using fixed response shapes for the event-specific responses composing predicted pupillary responses proves to show promise as a method for pupil deconvolution. The observed effects over the states of the selected model are generally in line with previous studies on the cognitive processing stages underlining associative recognition. Supporting the MEG model of Borst et al. (2016) as being the most plausible, the selected HsMM-GAMM sheds light on far more than what previous models of associative recognition have found, the relation of the duration of the stages to the intensity level of mental effort required to perform the selected cognitive task.

Nonetheless, the HsMM-GAMM approach is not without any fault. Most importantly, the utilized approach only takes into account different event-specific responses differing in their temporal distribution between subjects and trials. This may prove to be an insufficient modelling approach, considering a significant main effect of Fan was found between conditions in almost all states. Thereafter, this method of deconvolution may not be specific enough to correctly assign the intensity level of mental effort to specific stages in its current architecture. To this end, a more extensive HsMM-GAMM approach ought to consider the shape of the same event-specific response to also vary between subjects and trials. Even so, for a more in-depth and well-rounded analysis of the impact of Fan and Probe Type on the intensity level of mental effort elicited by different stages, a model assuming that only for certain states the event-specific responses have different shapes between conditions should be implemented.

Additionally, the HsMM-GAMM approach currently models some events as having negative amplitudes. This would, therefore, indicate that certain states, associated with stages such as Recollection as seen in Figure 3.8, elicit a response with a negative intensity level of mental effort. Moreover, these negative responses seem to have amplitudes as great as the positive ones, both being quite larger than the amplitudes found in the original pupillary response. Thereafter, these positive and negative responses are seemingly modelled so as to somewhat cancel each other out in order to form a predicted response similar to the original one. This behaviour of the model does not seem to resemble a veracious approach to pupil deconvolution, highlighting a need for a change in the implementation of the approach.

# References

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford Uni-

versity Press.

Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, *128*(2), 186.

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.

Barthelme, S. (2021). eyelinker: Import asc files from eyelink eye trackers [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=eyelinker` (R package version 0.2.1)

Boring, E. G. (1929). *A history of experimental psychology.* Genesis Publishing Pvt Ltd.

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., Ghuman, A. S., & Anderson, J. R. (2016). Tracking cognitive processing stages with meg: A spatio-temporal model of associative recognition in the brain. *NeuroImage*, *141*, 416–430.

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. *Journal of cognitive neuroscience*, *25*(12), 2151–2166.

Coltheart, M. (1981). The mrc psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, *33*(4), 497–505.

Donders, F. C. (1969). On the speed of mental processes (W. Koster, Trans.). *Acta psychologica*, *30*, 412–431. (Original work published 1868)

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*(3611), 1190–1192.

Hoeks, B., & Levelt, W. J. (1993). Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research methods, instruments, & computers*, *25*(1), 16–26.

Krause, J., Borst, J. P., & van Rij, J. (2023, June 15 - 16). *Investigating early frequency effects in lexical decision making using pupil deconvolution.* [Conference Presentation], 43rd TABU Dag, Groningen, The Netherlands.

Kyröläinen, A.-J., Porretta, V., van Rij, J., & Järvikivi, J. (2019). *PupilPre: Tools for preprocessing pupil size data.* Retrieved from `https://CRAN.R-project.org/package=PupilPre` (Version 0.6.2, updated 2020-03-08)

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). Opensesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods*, *44*(2), 314–324.

R Core Team. (2021). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in cognitive sciences*, *11*(6), 251–257.

Sohn, M.-H., Goode, A., Stenger, V. A., Jung, K.-J., Carter, C. S., & Anderson, J. R. (2005). An information-processing model of three cortical regions: Evidence in episodic memory retrieval. *NeuroImage*, *25*(1), 21–33.

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American scientist*, *57*(4), 421–457.

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, *40*(1), 1–29. Retrieved from `https://www.jstatsoft.org/v40/i01/`

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). dplyr: A grammar of data manipulation [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=dplyr` (R package version 1.1.1)

Wierda, S. M., van Rijn, H., Taatgen, N. A., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences*, *109*(22), 8456–8460.

Yu, S.-Z. (2010). Hidden semi-markov models. *Artificial intelligence*, *174*(2), 215–243.

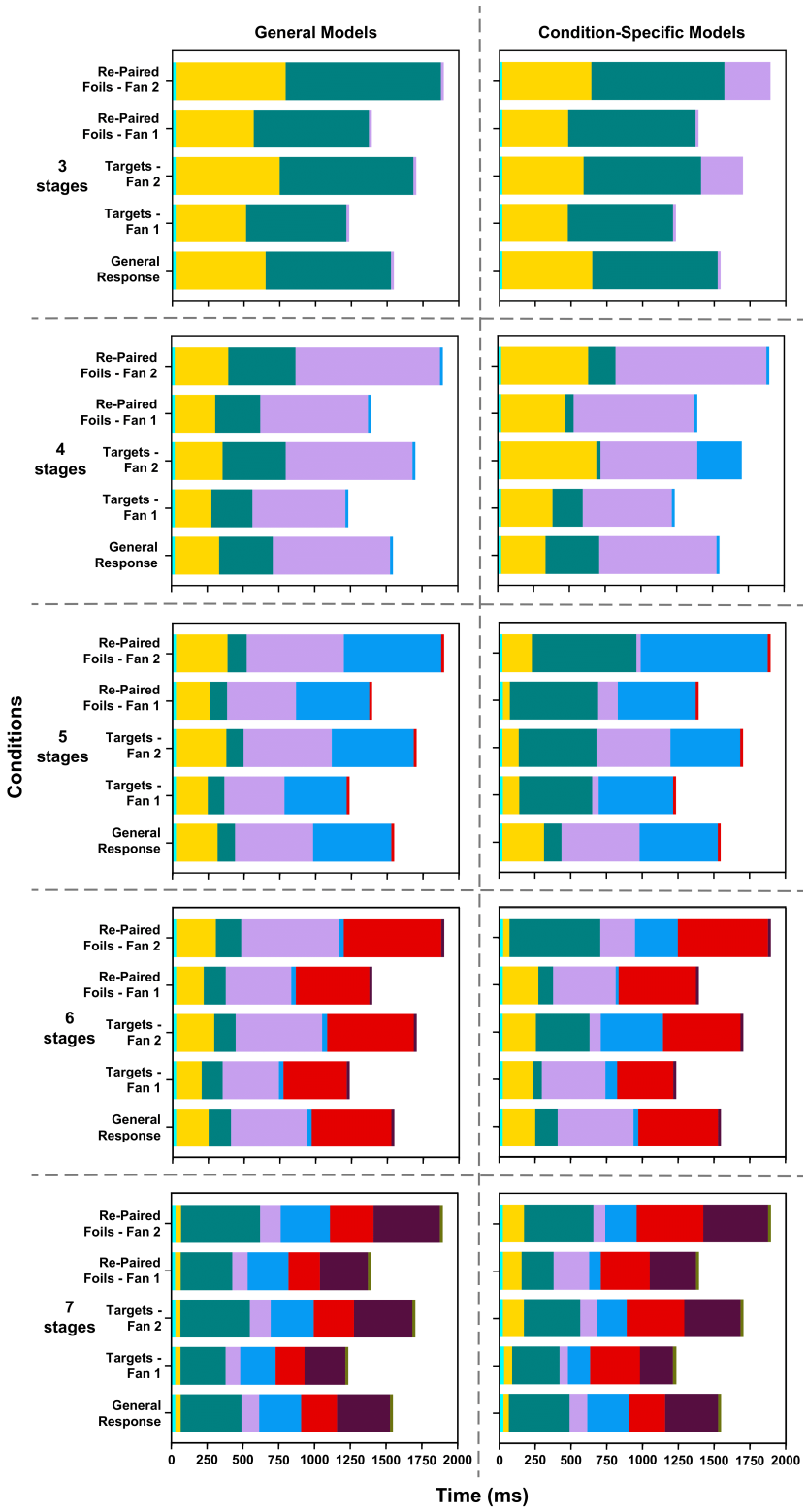


**Figure .2:** The average duration of each state in the five fixed response shapes general models. These averages are presented for the general responses, as well as for each condition. (left) The average duration of each state in the fixed response shapes condition-specific models. All condition-specific models with the same number of stages are grouped in the same plot. (right) The General Response bar is the same in plots with the same number of stages, being taken from the corresponding general model.

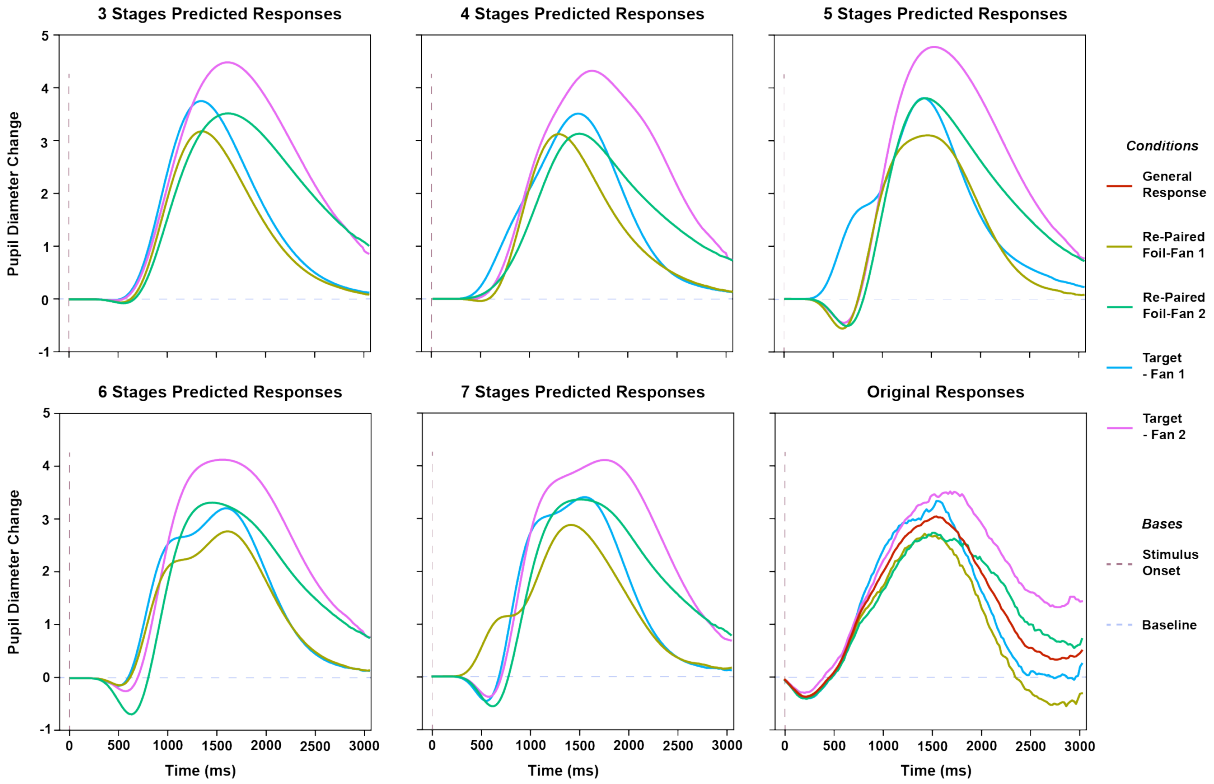


**Figure .3: The mean pupillary responses as predicted by each of the condition-specific fixed response shapes models. Each figure contains the mean response for each condition-specific model with the same number of stages. The figure on the bottom right-hand side contains the original mean pupillary responses, as obtained from the participants' data.**