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

Tracking pupil size to determine mental processing in associative recognition

Bachelor's Project Thesis

Eoghan McDonagh, s4316851, e.t.mcdonagh@student.rug.nl,
Supervisors: Dr J. Borst & J. Krause

In the past number of decades, several models have been put forth to explain the cognitive mechanisms behind associative recognition. These models have grown more intricate over time, with brain imaging techniques such as EEG and MEG providing a clearer picture of the nature and duration of the individual stages that comprise the associative recognition process. Our study aimed to identify those same stages using pupillary data, which we collected and then analysed through a combination of hidden semi-Markov models and generalized additive mixed models. We also hoped to investigate the usefulness of pupillary data as a cognitive analysis tool, particularly in terms of its ability to identify discrete processing stages. The resulting models were broadly comparable to those produced by earlier studies, thereby supporting the applicability of pupillary data in this and similar fields.

1 Introduction

Associative recognition is the act of recalling the relationship between multiple items, in addition to the items themselves. Several theories as to the associative recognition process have been posited (Anderson & Bower, 1973; Anderson, 2007), and in more recent years such theories have been refined and scrutinized through the use of various neuroimaging techniques (Borst & Anderson, 2015; Borst et al., 2016). The goal of this particular study is to examine associative recognition in terms of pupil size, and in so doing assess the applicability of pupillary data as an analytical tool.

1.1 Associative Recognition

There have been numerous attempts to identify the various processing stages that comprise associative recognition. Early studies relied on behavioural (Anderson, 1983) and fMRI data (Sohn et al., 2005), culminating in an ACT-R model capable of approaching associative recognition tasks in much the same way as humans (Anderson, 2007). This model offered a general understanding of the associative recognition process and its component stages, though its level of temporal detail was limited by the data on which it was based.

The ACT-R model suggested four stages in total, beginning with "encoding". This occurs during the perception of an item and refers to it being stored as information in the brain. Encoding is followed by "associative retrieval", during which a person searches their memory for an existing item that matches the newly perceived one. This stage is typically the longest of the four. Afterwards the relatively brief, self-explanatory "decision" and "response" stages take place.

More recently, EEG data - with its finer temporal resolution - was used to further elaborate upon the ACT-R model (Borst & Anderson, 2015). This resulted in the identification of an additional stage, "familiarity", which coincides with the end and beginning of encoding and associative retrieval, respectively. This overlap stands in contrast to the serial stages proposed by the ACT-R model, providing strong evidence in favour of the dual-process theory of associative recognition (Yonelinas, 2002).

In brief, the dual-process theory suggests that associative recognition consists of two memory operations: "familiarity" is when an item is assessed on the basis of having previously been seen, whereas the longer "recollection" is when the association between two items is gauged. These two approaches to recognition appear to involve some degree of overlap, with recollection beginning before the familiarity stage has fully concluded. Single-process theory, on the other hand, suggests that familiarity and recollection are part of one overarching stage (Gillund & Shiffrin, 1984).

A more recent relevant study, and the one from which our own research most heavily draws inspiration, involved the use of MEG data (Borst et al., 2016). The temporal and spatial detail provided by MEG data presented an exciting opportunity to verify the preceding ACT-R and EEG models. The MEG model of associative recognition is indeed the most detailed of the three, suggesting a total of six stages. As with previous models it indicates that encoding occurs first, and that the decision and response stages occur last. It also points to three distinct stages within the interim, namely "familiarity", "recollection" and "representation". These stages take place one after the other, although the MEG model offers a clear understanding of the extent to which they overlap. The "lexical and semantic access" part of the encoding stage, for instance, occurs at roughly the same time as familiarity and at the beginning of recollection. The majority of the representation stage, to give another example, takes place in tandem with the decision stage.

Ideally, the pupillary responses gathered in the course of our own research would paint a similarly detailed picture of the associative recognition process. Our results should also offer some indication as to the usefulness of pupillary data in deconstructing this and other cognitive functions.

1.2 Pupil Dilation

For decades, pupil size has been widely accepted as an indicator of cognitive activity (Hess & Polt, 1964). Our pupils increase in size (dilate) during periods of concentration or arousal and decrease in size (constrict) when we relax (Hoeks & Levelt, 1993). Our intent, therefore, was to have participants perform an associative recognition task and track their pupil size throughout. We predicted that the pupillary changes would point to distinct stages within the associative recognition process which, in turn, would allow for further verification of the most widely accepted associative recognition models.

Another aspect of pupil dilation, and one which presents wholly distinct research possibilities when compared to the previously discussed methods of data collection, is intensity. An analysis of the brain's activity with MEG can reveal a great deal about the role of specific brain regions, as well as the duration of their activation. Pupillary data is inferior in both regards, providing no insight as to the brain's behaviour (and, as such, no insight as to the exact nature of individual stages) and failing to account for the possibility of dual-processing (i.e.: overlapping stages). The degree to which the pupil's size increases, however, is indicative of cognitive effort (Kahneman, 1973), thereby allowing for a more comprehensive view of the associative recognition process.

1.3 Current Study

In summary, our objective was to have participants perform an associative recognition task and track their pupillary responses. The task required participants to learn and recall word pairs, and consisted of a training phase (during which the word pairs are learned) and a testing phase (during which their memory is tested and the relevant data is collected).

Once a sufficient amount of data had been collected it underwent pre-processing, after which it was used to train a number of hidden semi-Markov mixed models (HsMMs) and generalized additive mixed models (GAMMs). We used the former to identify associative recognition stages, while the latter mapped each pupillary response to those stages (Wood, 2017). These models allowed us to intuitively analyse our results as well as draw comparisons between ours and previous research. This, in turn, provided some insight as to the usefulness of pupillary data in this and similar research fields.

Finally, it is important to make a distinction between the above analysis method - conceived and developed by Joshua Krause - and previous attempts at pupil de-convolution. Hoeks & Levelt (1993) argued that pupil dilation is a function of "attentional pulses", which are themselves associated with cognitive events. They reasoned that such a function could be calculated, thereby allowing a given pupillary response to be de-convolved in terms of its component events. Wierda et al. (2012) later expanded upon this additive approach, presenting a method by which longer intervals of pupillary activity could be examined and de-convolved. Krause's method marks a significant departure from those and similar de-convolution techniques: through the use of the aforementioned models, pupillary responses can be broken down into a number of processing stages. These stages can then be examined in terms of their duration and general pupillary response, thus offering a greater level of detail than could be obtained through more conventional processes.

2 Methods

The experiment was comprised of a training and testing phase. During the training phase participants learned 32 word pairs, consisting of 16 short word pairs and 16 long word pairs. During the testing phase participants were asked to distinguish between target and non-target word pairs, with the latter encompassing incorrect pairings of individual target words ("re-paired foils") and entirely new word pairs ("new foils"). Long word pairs did not appear during testing. Training and testing occurred on the same day, though participants were encouraged to take a break in between phases.

In addition to the probe type of each word pair we manipulated and examined the effects of fan, or the number of pairs that a particular target word appeared in, on experimental performance. The position of the words within each pair remained consistent throughout.

2.1 Participants

The experiment was conducted with 25 participants (13 male, 11 female and 1 undisclosed, with ages ranging from 20 to 26 and with a mean age of 21.48), as well as a pilot participant whose data was excluded from our final results. A 26$^{\text{th}}$ participant attempted the experiment while wearing colour contact lenses, which caused the eye-tracking apparatus to require frequent adjustment and prevented them from finishing the experiment's testing phase. Participants were henceforth asked not to wear colour contact lenses.

Each participant was a student of the University of Groningen and gave their consent by filling out a form just prior to the experiment. Participants were paid €16 for taking part.

2.2 Materials

Each participant was assigned a unique pool of 136 word pairs and tested with word pairs drawn from that pool. The pools themselves were derived from the 464 word list used in Borst et al. (2013). This list was comprised solely of nouns with word frequencies between 2 and 100 occurrences per million, and with minimum imageability ratings of 300. Half of the words on the list were between 4 and 5 letters in length, and it was with these short words that we constructed the required pools for each of our 25 participants.

During the training phase participants were presented with 16 target word pairs, as well as 16 long word pairs. These long words were chosen at random from the long word half of the aforementioned list and ranged between 7 and 8 letters in length. All 16 long word pairs were the same for each participant. As participants were not assessed on their recognition of the long word pairs, their inclusion during the testing phase was primarily intended to reproduce the level of learning difficulty in Borst et al. (2013) and Borst et al. (2016). We also predicted that a mix of short and long word pairs would increase task engagement.

Each 136 word pair pool was used to create 13 blocks for the experiment's testing phase. These blocks, in turn, contained 40 word pairs: the 16 target and 16 re-paired foil pairs were consistent across all 13 blocks, but each one featured 8 different new, non-target word pairs drawn from the pool.

2.3 Procedure

Training and testing were conducted on the same day. The training phase itself consisted of two stages, the first of which involved participants being shown each of their 16 target word pairs and the 16 long word pairs. Participants were instructed to learn each word pair that appeared. They were also advised to make a connection between the two words so as to better remember them. Word pairs appeared on screen for 5000ms and were each followed by a 500ms blank screen, with the latter preventing participants from being overwhelmed by stimuli and allowing them a reasonable opportunity to commit each item to memory. After every pair had been presented, participants were notified that the second stage of training would begin when they were ready. During this stage, participants were shown the first word from one of the previously learned pairs. They were instructed to type the word (or *words*, in the case of the fan 2 pairs) that they had seen the prompt word appear with. In the event of an incorrect response, participants were shown the correct word(s) for 2500ms before being presented with the next prompt. After all the prompt words had appeared once, participants were again shown the words to which they had responded incorrectly and, again, provided with the correct word(s) for 2500ms if incorrect in their answer. This process was repeated until participants had responded correctly to each word. Participants were made to complete this second training stage three times before moving on to testing.

The testing phase required participants to perform a recognition task with 13 different blocks of 40 word pairs each, as well as an initial practice round featuring a smaller block of 10 word pairs (comprised of 4 target, 4 re-paired foil and 2 non-target pairs). Participants were shown each word pair - each preceded by a 500ms fixation cross - and instructed to respond quickly and accurately. Participants were asked for their dominant hand prior to the beginning of the experiment; right-handed participants were told to press 'J' if they recognised a given word pair as a target and 'K' if not, with left-handed participants being told to press 'F' if they recognised a word pair and 'D' if not.

Each response was followed by a 2000ms hash mask (with the length of the mask matching the length of the stimuli) to allow their pupil dilation to return to (or approach) base-line levels. Finally they were shown feedback in the form of "Correct" or "Incorrect" for 1000ms and a fixation dot for another 2000ms.

Participants were permitted to take a break between each block. During testing they remained seated and with their heads placed on a headrest, with the positions of the headrest and experimental display remaining consistent between every iteration of the experiment.

2.4 Data Collection and Preprocessing

Performance and eye-tracking data were collected from each participant via OpenSesame (Mathôt et al., 2012; Peirce, 2007, 2009): the number of occurrences of each prompt word (per block) was recorded during the training phase, and participants had their response times and accuracy recorded during the testing phase. The testing phase also saw the application of the Eyelink Portable Duo, with which we recorded the diameter of their right pupils over the course of the experiment. This data was recorded at a rate of 500Hz. The angle of the Eyelink apparatus was adjusted for each participant so that their pupils were in focus, although the parameters for the accompanying Eyelink software were kept constant; these parameters included calibration pacing interval, which was set to 1000. The position of the headrest also went unchanged, though participants were encouraged to adjust its height so that they were seated comfortably for the experiment's duration.

The Eyelink was calibrated to a sufficient level before the beginning of the testing phase and again whenever the Eyelink lost track of the participant's pupil, usually after they had left and re-entered its field of vision. The calibration process consisted of two repetitions of the same procedure, the first being purely to re-calibrate the Eyelink and the second to verify that it could effectively track the pupil under those new calibration settings: participants were required to look at nine points on the screen, one after another, with the Eyelink automatically detecting when participants had looked at a particular point and then transitioning to the next. After a successful calibration participants were asked to confirm their readiness, whereupon the experiment was resumed by the experimenter.

The eye-tracking data required some modification before it could be properly analysed: we removed any fixations outside the range of the experimental screen's resolution (namely 1920x1080), as well as any blinks that occurred during recording. This was first attempted manually, with each response to each individual trial being checked for blinks after initial automatic cleaning. Blinks regularly went undetected during the automatic cleaning procedure, thereby requiring several hours of manual verification and removal per participant. It was decided to adjust the scope of the automatic cleanup and set the padding to 125ms, vastly improving the effectiveness of the automatic cleaning process - for clarification, "padding" refers to the range of removed data on either side of a detected blink. After a participant's pupillary data had been cleaned it underwent a relatively brief manual check, with further adjustments being made to the data as necessary. Next the data was baselined to 200ms so as to compare pupil sizes for and between each participant. Finally the data was downsampled from a rate of 500Hz to 50Hz.

2.5 Analysis

The Markov-switching Spline Models toolbox, developed by Joshua Krause, provided the foundation for our data analysis. We used semi Markov-switching de-convolving generalized additive mixed models to deconstruct the pupil dilation time-course (i.e.: change in pupil dilation over time) for each trial into a specified number of stages. The first part of this deconstruction process involved HsMMs, which identified the individual stages within each time-course. We then used splines to estimate the overall response for each stage, and had GAMMs penalize those response estimates on the basis of their having been overfitted.

We examined the stages proposed by each of our models, as well as their predicted pupillary responses, in order to assess their suitability. We experimented with a range of stage numbers (3 to 7, inclusive) with the remainder of the models' other parameters being kept largely uniform; foremost among these parameters is the number of chains, which was set to 5 early in our analysis process.

Chains refer to the number of separate runs made by a single model. A model with a chain variable of two, for example, will make two attempts to de-convolve the provided pupillary response data into the specified number of stages. These attempts (and their corresponding predictions) are each penalized based on the likelihood of overfitting having occurred. The least penalized, best fitting prediction is then chosen as the "best chain" and the rest are effectively discarded.

We used the pupillary response data from all 25 participants in training our models. That being said, our initial models drew on data from a single condition (e.g.: fan 1 targets, fan 2 re-paired foils) and were trained with responses from fewer than 25 participants. This allowed us to train and compare them relatively quickly. Later models were trained using all of our accrued pupillary data, with the exception of responses belonging to the new foil condition. This is simply because no recognition occurs during this particular condition.

3 Results

3.1 Behavioural

3.1.1 Training

With respect to the training phase we recorded the average number of repetitions of each word pair before it was fully learned - or rather, before participants matched the prompt word with its correct pairing(s). This number varied between blocks and between conditions, being highest during the first block and especially high in the case of fan 2 word pairs (see the leftmost panel of Fig. 3.1).

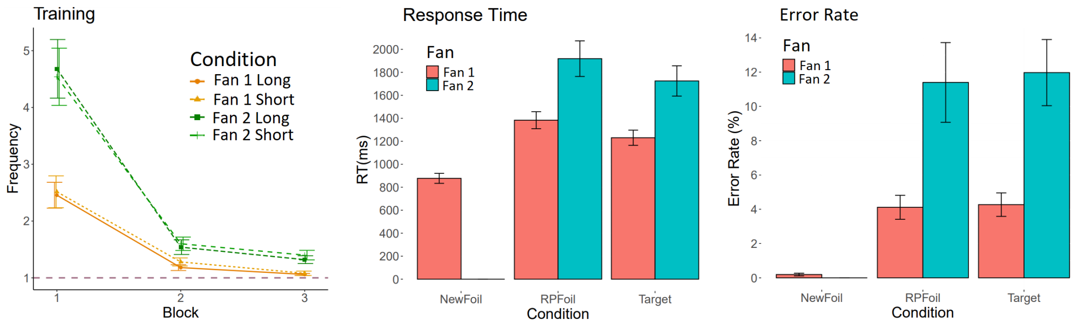
3.1.2 Testing

During the testing phase we recorded the response times (Fig. 3.1, middle panel) and error rates (Fig. 3.1, rightmost panel) of each participant, in addition to their pupillary data. Note that response times of more than 2 seconds were excluded from our final dataset, our reasoning being that only responses given under that threshold represent genuine recognition. The response times of incorrect responses were excluded on this same basis.

New foils were responded to the most quickly, on average. Across the remaining conditions, fan 2 word pairs elicited noticeably higher response times than their fan 1 counterparts. It should be noted, however, that response times were higher for both fan types in the re-paired foil condition when compared to the target condition. Having performed a one-way ANOVA, both fan ($F(1) = 1197.1$, $p<0.01$) and probe type ($F(3)=136.2$, $p<0.01$) were found to have a statistically significant effect on response times. These results would suggest that the fan of a particular word pair had more of an impact on the speed of its processing than the probe type, and that re-paired foils required more effort to identify than target word pairs.

Similar conclusions can be derived from the average error rates. New foils were responded to the most accurately. Fan 1 word pairs across both of the other conditions were responded to less accurately, with fan 2 word pairs across both conditions being

Figure 3.1: Behavioural results, from left to right: training, testing phase response times and testing phase error rates



responded to least accurately of all. For both fan types, targets elicited slightly more errors on average than re-paired foils. As with response time, these results indicate that new foils are the least demanding condition and that fan has a larger impact on accuracy than word pair type. Notably, targets were identified less accurately than re-paired foils across both fan 1 and fan 2. A one-way ANOVA showed that both fan ($F(1)$=405.75, p<0.01) and probe type ($F(3)$=16.66, p<0.01) had a statistically significant effect on error rate.

3.2 Pupillary

Pupil diameter was measured over the course of each trial. Once processed (i.e.: once blinks and other artifacts were removed) and baselined, each pupillary response was grouped together according to its condition. This allowed us to train a model on one specific word pair type, as well as calculate the average pupillary response for each condition. As evidenced by Fig. 3.2, the overall pattern of the pupil's diameter change is consistent across conditions. The key difference between them is the degree to which they cause the diameter to change, with fan 2 target word pairs eliciting the greatest increases and new foils eliciting the lowest. These results are consistent with the notion that pupil size is an indicator of cognitive effort, and suggest that our pupillary data is generally reliable.

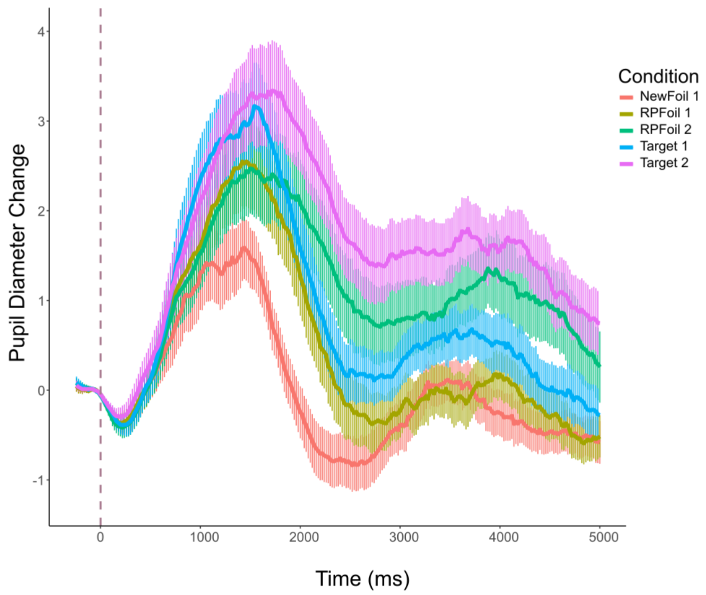
3.3 Model

Our data analysis produced five comparable models, each having been trained using the entirety of our processed pupillary response data and set to a different number of stages. These models deconvolved each pupillary response, thereby allowing us to calculate the average, general pupillary response - and the individual stages comprising said response - according to each model (see Fig. 3.4 for an example). We also extracted the predicted pupillary response for each condition according to each model (Fig. 3.5). This proved especially useful in gauging the accuracy of a given model, as the issues of overfitting and underfitting became relatively clear when comparing a model's predictions to the pupillary response for each condition according to the collected data. Finally we obtained the average duration of each stage within each of our models (Fig. 3.3). Although lacking start and end points - thereby preventing us from investigating stage overlap - these durations were exceptionally useful in comparing ours to existing models.

All five of our models include an initial stage to account for stimulus onset (see the light blue, leftmost stage in each bar in Fig. 3.3). This stage falls outside the scope of the associative recognition process and can therefore be ignored. Our models also share a brief final stage.

The predicted pupillary responses for each condition according to the three stage model are significantly less detailed when compared to our data, notably lacking any initial dip in diameter. The four stage model is similarly imprecise, with its average stage durations proving particularly revealing: its first two stages are of roughly the same duration and followed by a very long third stage. This is a stark contrast to the ACT-R model (also comprised of four stages) which posited that the second stage, associative retrieval, is by far the most

Figure 3.2: Average pupillary response (change in pupil diameter) over time, for each condition



time-consuming.

The predicted pupillary responses of the five stage model do include an initial decline in diameter and are generally more detailed than those of the preceding two models. It further bears some similarity to the five stage EEG model. Its stages (excluding the brief final stage) are generally comparable in length, and the same can be said of the stages outlined by the EEG model.

The predicted responses of the six stage model are again more detailed, though only slightly so. Of greater interest are the average stage durations, which correspond somewhat to those of the six stage MEG model. The first and second stages, those being the encoding and familiarity stages in the context of the MEG model, are relatively short. They are followed by the comparatively long third stage, whose MEG counterpart is the similarly long "recollection". According to the MEG model the final three stages (namely representation, decision and response) are of about the same duration. Our six stage model instead indicates that the fourth and sixth are extremely brief, with the fifth having a comparable average duration to that of the third.

Lastly, the seven stage model is something of an outlier. Its average durations suggest that a brief stage occurs immediately following stimulus onset. It also lacks the equally brief fourth stage which appears in the six stage model, instead indicating that the fourth, fifth and sixth stages are of comparable length. In general, the seven stage model disrupts the pattern established by the preceding four: the three to six stage models appear to capture the same broad stage duration structure, with each model separating the stages of its predecessors into smaller, more detailed parts. The seven stage model, by comparison, presents a series of stages that bear little resemblance to even the six stage model's. Its predicted pupillary responses (bottom right panel, Fig. 3.5) are also quite distinct in their behaviour.

4 Discussion

The goals of our study were to examine the associative recognition process, and in so doing assess the relevance of pupillary data. This assessment necessitates that our findings - and particularly our models - be compared to the relevant, pre-existing research.

The results of our experiment were as expected, and largely align with the findings of similar studies (Borst et al., 2016): fan 2 word pairs generally required more effort than their fan 1 counterparts,

Figure 3.3: Average stage durations for our five models, per condition

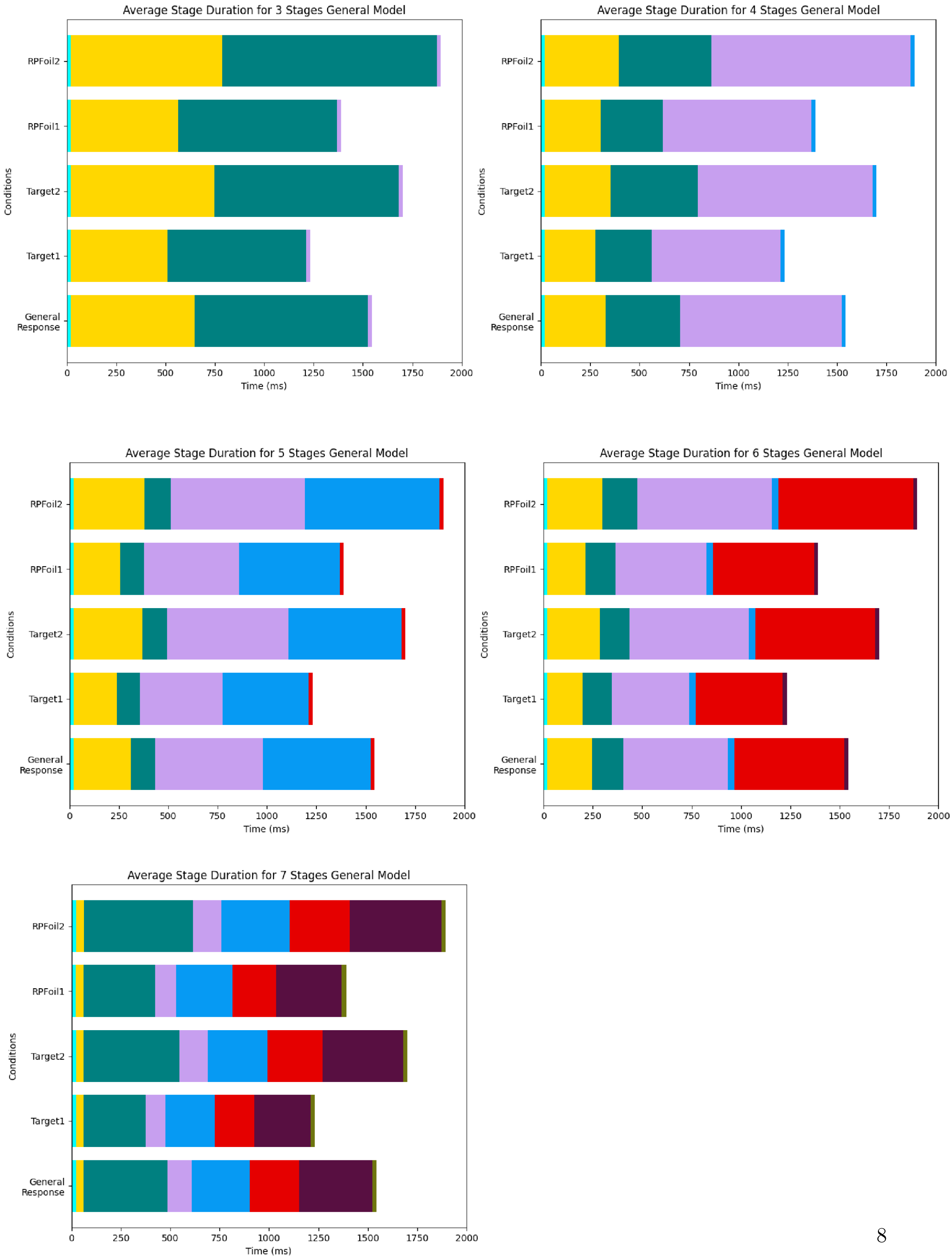


Figure 3.4: Predicted pupillary response for each individual stage, according to the six stage model

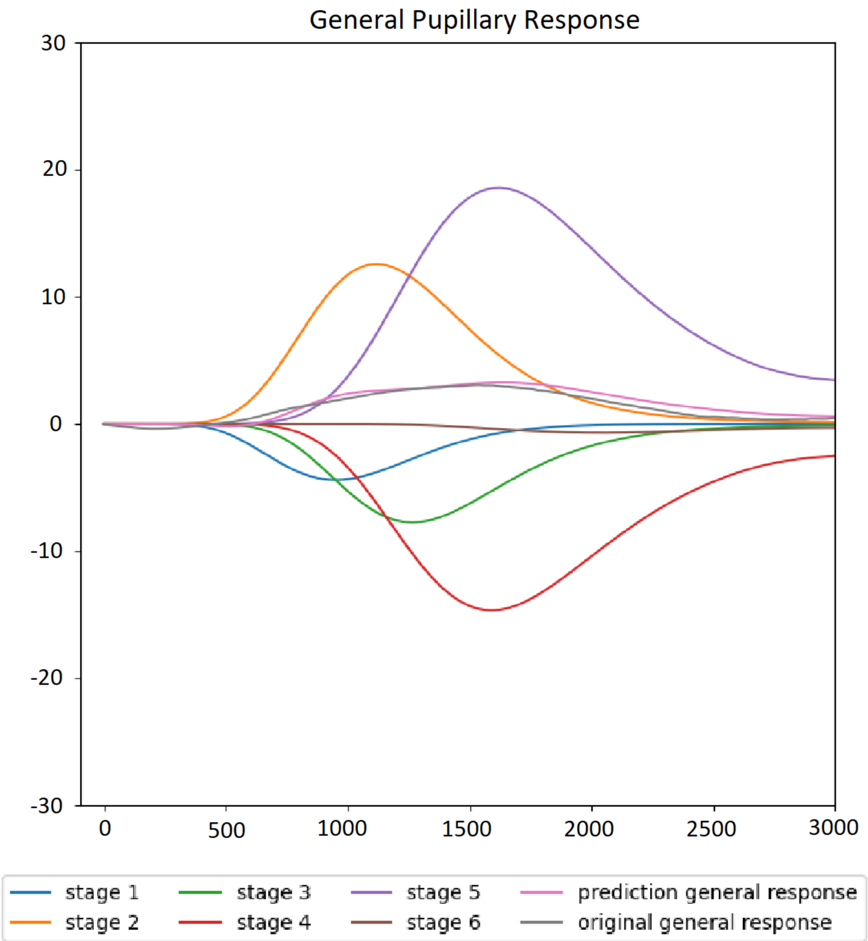
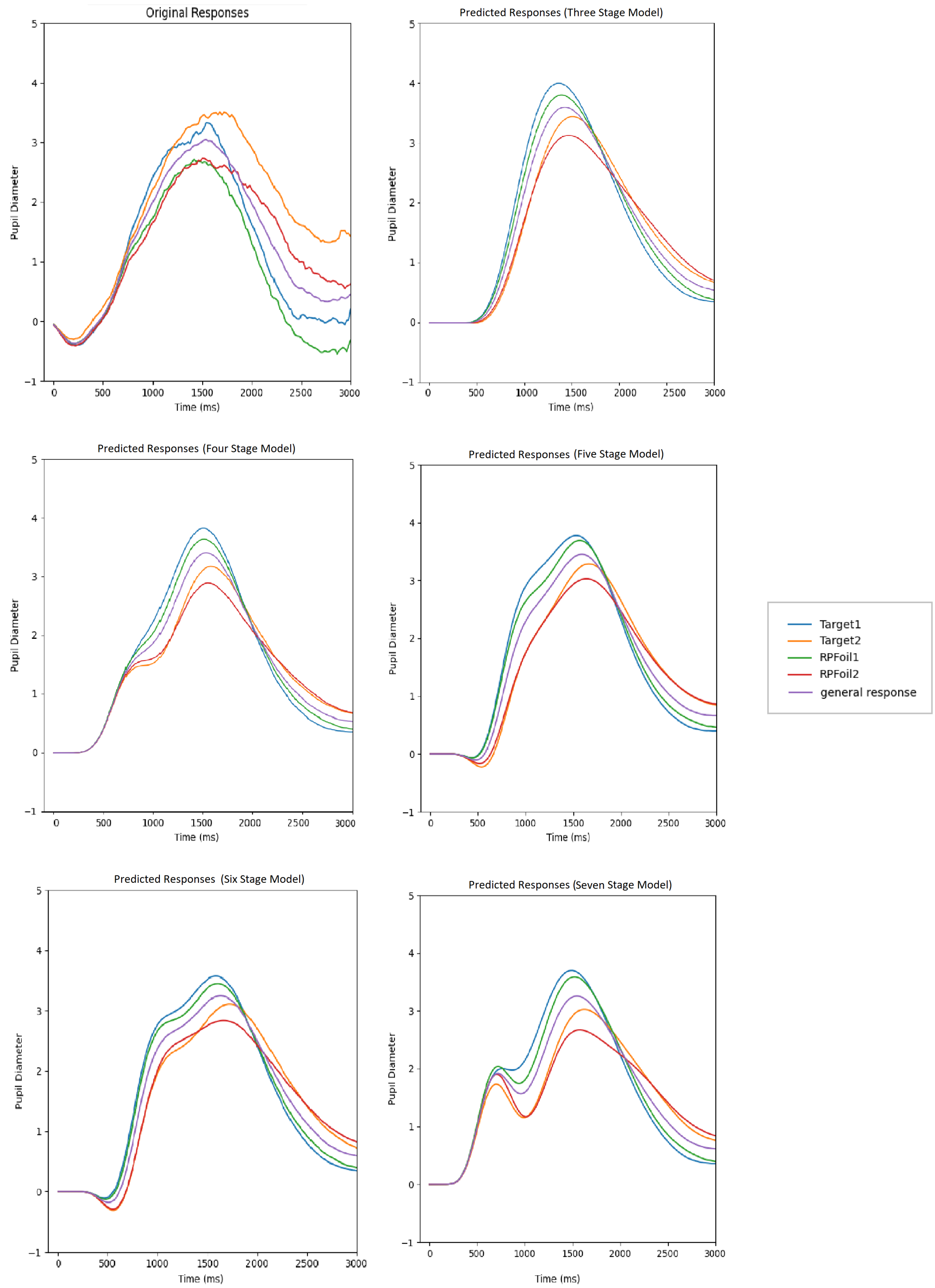


Figure 3.5: Pupillary responses for each condition, according to (from left to right, starting on the top row): data, three stage model, four stage model, five stage model, six stage model, seven stage model



both in terms of encoding and recollection. New foils were almost always identified as such. On average, re-paired foils required slightly more time to respond to than targets while the latter were more frequently misidentified than the former.

As outlined earlier in this paper, pupillary data offers nothing with respect to the precise cognitive function of a particular stage. For that reason our models are best viewed through the lens of previous research. It is apparent, when comparing our final five models to the MEG model of associative recognition, that many of them are insufficiently detailed. Our three, four and five stage models, like the earlier ACT-R model, merge stages and thus oversimplify many aspects of the associative recognition process. Our seven stage model exhibits a tendency to overfit - likely a result of its having to find more stages than are present. This overfitting is best evidenced by the predicted responses of the seven stage model, some of which feature a decrease in diameter at 1000ms not visible in the actual data. Our six stage model, perhaps predictably, is the most well-behaved and aligns most closely with the aforementioned MEG model.

The similarities between the two extend beyond their having an identical number of stages: although our own model fails to account for the possibility of overlapping stages, it does provide the order in which they begin as well as their average duration. With this information we can draw comparisons between the stages in the six stage model and those of the MEG model which, despite the occasional overlap, also begin at different times in the associative recognition process.

Although there are obvious differences between the two, both our model and the MEG model suggest that recollection and decision-making are among the longest stages, and that the visual encoding (and lexical and semantic access) stage is longer than the ensuing familiarity stage. Furthermore, the durations and overall behaviour of our six stage model's stages are as one would expect - provided that those stages are indeed similar in nature to those occurring in the MEG model. The response stage is exceedingly brief (due to the low mental effort involved) while the visual encoding and recollection stages are highly sensitive to the type of word pair being looked at, increasing quite noticeably when dealing with fan 2 items.

In summary, the results of our six stage model align somewhat with the prevailing MEG model. This, in conjunction with the predictable behaviour of our model's stages, would suggest that pupillary data is a suitable indicator of the internal processes that occur during associative recognition. Our conclusion, therefore, is that pupillary data presents a viable alternative (or accompaniment) to more widespread brain-analysis tools. We imagine that future research will better exploit the unique advantages of pupillary data, providing an elevated degree of insight as to mental effort in addition to the temporal and spatial detail captured by the likes of EEG, fMRI and MEG.

References

Anderson, J. R. (1983). *The architecture of cognition.* Harvard University Press.

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

Anderson, J. R., & Bower, G. H. (1973). Human associative memory.

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.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological review*, *91*(1), 1.

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.

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Citeseer.

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

Peirce, J. W. (2007). Psychopy—psychophysics software in python. *Journal of neuroscience methods*, *162*(1-2), 8–13.

Peirce, J. W. (2009). Generating stimuli for neuroscience using psychopy. *Frontiers in neuroinformatics*, *2*, 343.

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.

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.

Wood, S. N. (2017). *Generalized additive models: an introduction with r.* CRC press.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of memory and language*, *46*(3), 441–517.