Which cognitive processes are reflected by pupil size: a model-based analysis

R.L. Koelewijn, r.l.koelewijn@student.rug.nl,
Supervisor: Dr J.P. Borst

Abstract: Pupil dilation as a reflection of cognitive effort is a known and well researched phenomenon. In this paper the exact cognitive processes underlying the pupil dilation effect are investigated. For this study an eye tracking experiment was conducted where participants performed a complex information processing task. As expected, the pupil dilation increased in the more difficult conditions. Using the ACT-R architecture a cognitive model was built that predicts which cognitive processes are involved in performing the task of the experiment. The architecture is composed of a number of modules which are each associated with a specific function and brain region. The activity patterns of the separate modules of the model were transformed into simulated pupil dilation data. The combination of the modules pupil dilation that best explains the human data gives insight into which cognitive processes are of interest for the pupil dilation response. The linear model that best explained the subjects pupil data consisted of the cognitive processes representing declarative memory, procedural memory and problem representation. The modules that were of greatest influence were the declarative, procedural and imaginal module.

1 Introduction

The human pupil responds to wide variety of influences. Since the early 1960s researchers have concluded that the pupil dilates as a result of some form of mental effort. (Hess and Polt, 1964; Kahneman and Beatty, 1966) Many synonyms are used such as cognitive activity, attentional effort or processing load, but what this mental effort entails exactly remains unknown. (Steinhauer and Hakerem, 1992; Hoeks and Levelt, 1993; Kahneman and Beatty, 1966) Many researchers have found the relation between pupil dilation and mental effort over the years. There is a body of studies on task-evoked pupil dilation and whether it is a good indicator of mental effort. (Van Der Wel and Steenbergen, 2018; Beatty, 1982) Some have theorized over the concept of mental effort but there is no unifying explanation what mental effort is. (Westbrook and Braver, 2015) This study tries to explain what constitutes this notion of mental effort and what causes the pupil to change by using a model based approach.

1.1 Literature review

The first research relating pupillary dilation to mental effort is dated well over a hundred years ago. (Schiff, 1875) showed that the pupil would dilate as a consequence of a variety of non visual stimuli. (Heinrich, 1896) used mental multiplication in his experiment to measure pupil dilation as a result of cognitive processing. This information remained unknown to the wider community up till Hess translated the German study by Heinrich. (Hess, 1975) The paper by Hess triggered the subsequent research on the subject.

(Hess and Polt, 1964) performed an experiment where subjects had to perform four simple multiplications with increasing difficulty. The problems were 7x8, 8x13, 13x14 and 16x23. Their results showed a strong correlation between problem difficulty and pupil dilation. Based on their findings and the setup of the experiment in which they eliminated other factors that influence pupil dilation they concluded that the "pupil response was a direct reflection of neural activity." P.1191 (Hess and Polt, 1964) This was one of the first

descriptions of pupil dilation reflecting mental effort and it inspired more research on the subject.

A literature overview written by (Beatty, 1982) reviewed several studies on task-evoked pupillary responses. (Beatty and Wagoner, 1978) showed that in a visual encoding task with varying complexity the pupil response reflected the processing required to complete the task based on the complexity of the problem. In an experiment by (Kahneman and Beatty, 1966) participants had to perform a short-term memory task in which they had to remember and report increasingly difficult items. The results showed a clear relation between difficulty and pupillary response. This shows that the pupil dilation and mental effort correlation holds over different tasks. In another study (Ahern and Beatty, 1979) showed that the pupil dilation was less for individuals of higher intelligence indicating that they exerted less mental effort. In the experiment two groups with different intelligence test scores performed mental arithmetic problems on three levels of difficulty. They concluded that the pupillary response is a suitable method of measuring intelligence and that the task evoked pupillary response seemed to reflect the mental effort exerted.

A recent review on task evoked pupil dilation as a measure of mental effort was written by (Van Der Wel and Steenbergen, 2018). The reviewed literature showed that more demanding tasks resulted in increases in pupil dilation across different types of tasks. An interesting result was a study by (Van der Meer, Beyer, Horn, Foth, Bornemann, Ries, Kramer, Warmuth, Heekeren, and Wartenbruger, 2010) which contradicted the original study by (Ahern and Beatty, 1979). The recent study showed that individuals of higher intelligence solved analogical problems faster and more accurately but also showed larger pupil dilation effects than average individuals. The reason these two studies contradict each other, so Van Der Meer reasons, is that in contrast to the arithmetic task this task is not over-learned thus making it more responsive to the difference in effort extended by the subjects. (Hogervorst, Brouwer, and Van Erp, 2014) showed that pupil dilation alone is enough to distinguish high cognitive from low cognitive load using classification algorithms. Only electroencephalography outperformed pupil dilation as a measure of cognitive load. A recent study by (Kursawe and Zimmer, 2015) confirmed that increases in set size in visual search tasks lead to increases in pupil diameter. In a recent paper (Westbrook and Braver, 2015) theorized what might constitute mental effort and concluded that it is best regarded as effort-based decision making.

A study by (De Jager, 2018) provided information on what mental effort might be. In the experiment subjects performed an algebra task during which their pupil dilation was recorded. A cognitive model was created and yielded insight as to which cognitive processes might influence pupil dilation. The processes that contributed most to pupil dilation according to his study were declarative memory, procedural memory and visual processing.

1.2 Cognitive Model

A model based approach can be used to determine the cognitive processes involved in the pupil response of a task. The model will perform the same task as human subjects and will give insight into which cognitive processes are of importance to complete the task. The models' activity is then transformed into pupil response data using a method found by (Hoeks and Levelt, 1993) and hence provides a prediction as to how the pupil will respond to the task. (Anderson and et al., 2006; Borst and Anderson, 2016)

ACT-R is the cognitive architecture used to build a model and make predictions about the pupil dilation data in this study. ACT-R is both a theory of human cognition as well as a framework in which models can be built which encompass the theory. (Anderson and et al., 2004) ACT-R simulates human cognition and considers psychological theory and physical human capabilities. The architecture consists of a number of modules each with a specific function and cognition is the result of the cooperation of these modules. (Anderson and et al., 2006)

1.3 Names Task

In the paper by (Van Der Wel and Steenbergen, 2018) three different types of tasks were described: an updating component-, shifting component- or inhibition component of cognitive control. Which

respectively refers to tasks involving: updating working memory, shifting between tasks or overcoming automatic responses. The names task used in this paper is of the first type as subjects have to keep track of incoming information and update their representation of the problem in working memory.

In the names task participants are shown 3 names in a random order plus an instruction. Subjects have to remember the names and possibly alter the order of the names based on the instruction. The instruction is either in the form of numbers, e.g. '13', or in the form of letters, e.g. 'AT', which correspond to numbers that subjects have learned before. The task is explained in more detail in the methods section. The names task was specifically designed to use all 8 modules of ACT-R and to make a distinction between memory retrieval and representational changes. (Anderson, 2007) The latter is useful as retrieval and representational changes are usually of influence on each other. The rather artificial, and complex, names task successfully allows for the separation of retrieval and representational changes between different conditions. Thus a clearer distinction between which modules are relevant for the task and contribute to pupil dilation can be made. Four trial types with varying difficulty should show a difference in pupil response.

1.4 Overview Thesis

This research tries to give insight into which cognitive processes are reflected by pupil dilation and thus what this mental effort mentioned in the literature reflects. An experiment was conducted where subjects had to perform the names task and their pupils were measured. A cognitive model that performs the same task was built in order to investigate which cognitive processes underlie and explain the pupil dilation effect. In the remainder of this paper the methods, cognitive model & architecture and results will be presented.

2 Methods

2.1 Experimental Design

The information processing task used in the experiment is the names task in which participants have to remember three names and possibly alter the sequence based on instructions given. The 3 names were presented to the participants and they were given an instruction in the form of a number- or letter pair. An instruction of the type '13' would mean that the participant had to switch the first and third name presented to them. The participants could also be shown an impossible instruction of the type '14' which meant that no change had to be made as there is no forth name. The instruction could also be a letter pair that corresponds to a number pair which had been learned before. Thus the 4 conditions could require a transformation and/or a retrieval, where transformation entails whether the order of the names has to be altered. Retrieval means whether an instruction has to be retrieved in the case of a letter pair or no retrieval in the case of a number pair. These were the conditions:

1. Condition 1: No retrieval and no transformation (number pair without action, e.g. '14')
2. Condition 2: No retrieval and transformation (number pair with action, e.g. '13')
3. Condition 3: Retrieval and no transformation (letter pair without action, e.g. 'AT'[=14])
4. Condition 4: Retrieval and transformation (letter pair wit action, e.g. 'OR'[=13])

The letter pairs were learned prior to the experiment in the learning phase.

2.2 Studying procedure

Upon arrival the participant was welcomed and given a short introduction to the experiment. They were then asked to read and sign a consent form. During the learning phase the participants learned the letter pair associations that would later be required during the experiment. The list of associations used can be found in the appendix. There were 12 letter pairs to be learned of which 6 required no transformation and 6 did require a transformation. The subjects were given instructions to learn the associations. If these were clear the subjects had 10 minutes to learn the associations, they were also given pen and paper to aid them during the learning. After those ten minutes they were

tested on these associations. Subjects had to perform a test of 4 blocks where in each block every association was asked and repeated until answered correctly. If they answered incorrectly they were given feedback what the correct answer was. This is an extra step that will help them learn the associations until they can recall all of them correctly.

2.3 Experimental Phase

After successful completion of the association test, the instructions for the experiment were given to the subjects and they were instructed to read them carefully. After the instructions were clear to the participant the headrest would be setup and the eye tracker calibrated. After that was all done the subject would start with the practice trials to get a feel for the experiment. If they still had questions or difficulties with the experiment it would be paused or restarted such that they could perform the practice trials again. After each of the 4 blocks of the experiment the participant could take a short rest, the experimenter would only continue the experiment after the participant indicated readiness. The actual experiment, after the 12 practice trials, consisted of 4 blocks of 24 trials each with a total of 108 trials. Each block contained 6 trials of each conditions in a random order so each condition was presented 24 times during the experiment excluding the practice trials.

Each trial started with a fixation cross in the middle of the screen that lasted between 400 and 600 milliseconds, this was randomly chosen for each trial. This fixation was followed by the presentation of the names "Tom, Dick and Fred" in a random order, each name appeared 0.5 seconds on screen and there was no time in between the names. (Figure 2.1) After the names another fixation cross of 500ms was shown and this was followed by the instruction. The instruction was either a number or letter pair and was visible on the screen for a maximum of 10 seconds. During those 10 seconds the subjects could click using a regular mouse to indicate that they were ready to give an answer. If they did not click within the 10 seconds a feedback screen would appear to let them know that they were too late. If the participant did indicate that they were ready to answer by clicking within the 10 seconds then an answer screen appears where they have to click the three names in the order they think is correct. The three names appeared in a random order stacked on top each other on the answer screen. They had 3 seconds to answer otherwise the answer screen would disappear and a feedback screen would appear to let them know that they were too late. The limit of three seconds was chosen such that there would be just enough time to give the answer but no time to continue thinking about their answer. The participant should find an answer within the 10 seconds and only use the 3 seconds of the answer screen for answering. Please refer to figure 2.1 for a visual representation of a trial. At the end of the experiment some information would be required from the participant. Finally the participants were thanked and their questions answered if they had any.

2.4 Participants

For this study 27 subjects participated in the experiment in exchange for 8 euros. Most subjects were in some way associated with the University of Groningen but there were a few exceptions. The data from 2 subjects were removed completely as the eye tracker did not function and no useful data was obtained. 3 subjects were removed due to their accuracy scores being significantly too low and 3 more subjects were removed because of the large number of blinks and saccades during the experiment made the data unreliable.

After inspection of the data and accuracy of the subjects 6 more subjects were removed from the experiment. They either performed worse than random answers or their measured pupil data contained too many blinks and thus yielded no useful data. Ultimately 19 subjects were used to make the analysis. (10 female; 17 right handed; mean age of 22.7; age range 18-31)

2.5 Apparatus & stimuli

The experiment itself as well as the association test used for the learning phase were created using OpenSesame. (Matht, Schreij, and Theeuwes, 2012) The eye-tracker used for the experiment was the Eyelink Portable Duo from SR Research. The eye-tracker measured the eyes with a frequency of 500hz and a nine-point calibration and validation was used for the experiment. A headrest was used

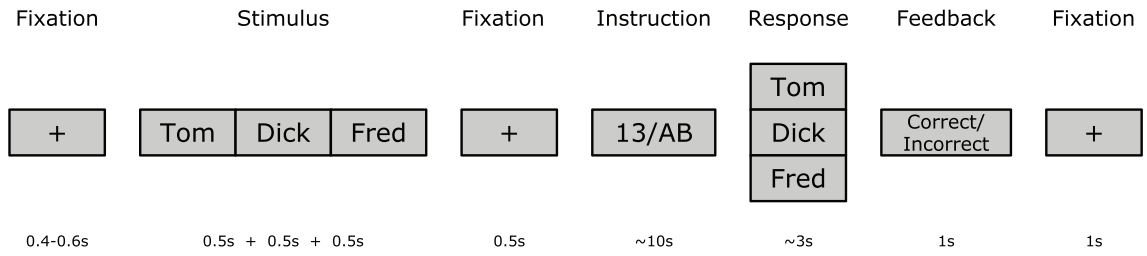

Figure 2.1: Visual representation of a trial

to minimize participants' head movement and improve the accuracy of the pupil measurements. The eye-tracker was in a small room with two desks that were separated using a cabinet. One desk was the experimenters desk and the other was the subjects desk where the headrest and eye-tracker was installed. The researcher could monitor the experiment from his desk as well.

3 Cognitive Model

A cognitive model can be used to simulate subjects performing an experiment while accounting for which cognitive processes are involved. In order to build this cognitive model a cognitive architecture can be used which will make specific assumptions about the functioning of the brain. The function of a cognitive architecture is to "find a specification of the structure of the brain at a level of abstraction that explains how it achieves the function of the mind." (Anderson, 2007) P.7 So a cognitive architecture allows for the creation of cognitive models at a level of abstraction such that it is easy to build a model for a specific task while the framework is in line with the physiological structure and activity patterns of the brain. The cognitive architecture used for this experiment is the "Adaptive Control of Thought - Rational" also known as ACT-R.

3.1 ACT-R

ACT-R is a cognitive architecture that provides a theory of central cognition and the cognitive architecture encompasses this theory. (Anderson and et al., 2004) It allows for the simulation of human cognition within this theoretical framework. The architecture can be used to model a wide variety of tasks and can interact with the real world. The ACT-R environment is based on certain assumptions of cognitive processing and these are implemented in the software. By constructing models in the architecture, the models adhere to the assumptions of ACT-R. A cognitive model will often provide data in the form of accuracy scores and reaction times which should match human data. If the model matches the data then the cognitive model is a plausible explanation of the cognitive processes involved. The structure of ACT-R will be described below followed by a description of the model for the names task.

According to the ACT-R theory "cognition emerges through the interaction of a number of independent modules." (Anderson and et al., 2004) P.314 The architecture consists of 8 modules and buffers each with their own designated function and associated brain regions. Each of these modules perform their operations largely independent of other modules and the modules can interact with the overall cognitive system by placing information in an associated buffer. The modules cannot interact directly with each other, they can only place a small piece of information in their buffer which will be available for the other modules. A central production system can respond to the patterns in these buffers and take actions including sending requests to other buffers. The modules can operate in parallel mostly but suffer from one serial bottleneck which is that they can only hold one piece of information in their buffer, execute only one action

at a time or look at only one item at a time. The separate modules will now be explained.

- **Procedural module** The procedural module orchestrates the interaction between all the modules. The procedural module responds to information available in other modules their buffers and is the only module that can interact with other buffers. The communication among buffers is thus achieved through the procedural module. Actions from this module in ACT-R are defined as production rules which are if-then rules of the following form:
  IF the goal is to read names
  and the visual module is free
  and we found a location of a name
  THEN move attention toward the visual location

  This particular rule would move the visual attention to a name if the visual module is not busy, a visual location has been found and the goal state is 'read names'. Through pattern matching the contents of the buffers and status of other modules, the best matching production rule will fire and perform an action which causes the content of the buffers to change. The module can quickly decide which production rules matches best among many in parallel, but it can only execute one production rule at a time. The modeller creates these production rules itself initially whereby he/she can model how the task at hand should be executed.

- **Declarative module** The declarative modules serves as the long term memory of the architecture. It holds information in the form of chunks which are small pieces of information of a certain form specified by the modeller. This could be a chunk for example:
  (trial13 ISA names name1 Fred name2 Dick name3 Tom)
  Meaning that the chunk is called trial13 which is of the type names and it contains three slots and their corresponding values, namely name1, name2 and name3. The modeller is free to come up with the form of the chunks used by the model. So the declarative module stores these chunks in long term memory and these chunks can be requested through the declarative buffer. However, there can be but one chunk in the buffer at the same time. The model starts with the chunks that have been specified by the modeller and can create and store new facts while performing the experiment using the imaginal module.
  Chunks contain information at the symbolic level, but there also is a sub-symbolic component, namely activation. Chunks can only be retrieved if their activation level is sufficient and this value is influenced by the recency and frequency of usage of a chunk as well as similar chunks. So activation influences the likelihood of retrieval for a certain chunk and when a chunk is not being used its activation will decay. When a request is made to the declarative buffer a chunk matching the request will be searched and if one matches with enough activation it will be retrieved.

- **Imaginal module** This module tends to be used to keep track of the problem state and hence can only store a small amount of information. This module also is the only way for the model to create new chunks and store them in declarative memory.

- **Goal module** The goal module serves to keep track of the state of the model. The goal module usually contains 1 chunk which will keep track of the state of the model as well as other variables that might be important. Usually this chunk never leaves the goal chunk, it is there during the entire experiment and will only be modified. Think of it as an extra layer of control over which production can fire as the productions might be dependent on the value of the state in the goal buffer.

- **Visual module**The visual module is one of the modules that can observe the outside world, just like the aural module. The visual module is responsible for the processing of stimuli presented in the external world. It keeps track of the location of these stimuli and directs attention to these stimuli in order to observe them. So when an unattended visual location has been found, the attention shifts there and the information there will be encoded into ACT-R.

- **Motor module** The motor module is one of the modules that can interact with the external world. The motor module allows for physical actions. It simulates movement from the subjects which are mostly limited to mouse movements like they are used in experiment settings and other hand movements.

3.2 Names Model

The names model performs four trials, one trial per condition. There is no need to perform more trials as the model is always correct and always goes through the same steps for a given condition. The model simulates the names task and performs the experiment analogous to the subjects. The model reads the information and instructions per trial akin to subjects and the model uses the motor module to click the correct answers. Similar to the experiment the model performs trials of all four conditions and its response time depends on the difficulty of the trial. Equivalent to the experiment the trials start with a fixation followed by the names in a random order and the instruction. As soon as the model has formulated its answer it will click, the answer screen will appear and the model answers in the correct order.

Due to the initialization of the model it contains declarative knowledge in the form of chunks that will be required for the model and information that is assumed the subjects would know. It contains two chunk types which are important:
(at14 isa instruction-name letters "at" numbers "14")
(fourteen isa switch instruction "14" num1 1 num2 2 num3 3)
The model has knowledge of what numbers the letters in an instruction represent and it knows what the correct order of the names should be once it has an instruction in the form of numbers. The latter is assumed humans would know so this was added as declarative knowledge. The instruction letter number pair is something the participants have learned before starting the experiment. It is assumed that the model always knows the instruction which reflects the extensive training subjects go through in the learning phase of the experiment. The model can always retrieve the instructions because activation for these chunks was set high enough. The mechanism, however, is implemented in such a way that the model could actually forget or be unable to retrieve the instructions in time, so it is in line with the human performance of the experiment.

At the beginning of each trial the module will wait for the names to show and for each of the three names it will shift attention to the location of the name and encode the name into the imaginal buffer. It does this for all three names and accordingly for the instruction as well, so these will all be stored in declarative memory through the imaginal buffer. For each trial the three names and their position are stored as three separate chunks in memory as it is unlikely that humans remember the three names in one chunk. The names will be retrieved from memory later on based on their position when the model is ready to answer.

After the encoding of the instruction the model behaves differently for the four conditions. In this stage the model tries to figure out in what order the names should be for the answer and after is has realized what the correct order is it will try to retrieve the names in that particular order. It is assumed that for the no transformation and no retrieval condition, e.g. 14, the model does not have to process the instructions any further, it immediately knows that the names as they were presented are the answer, so it moves straight to the answering section of the model.
The third condition is practically the same except that a retrieval is added in between. Seeing as the instruction would be AT, the model has to retrieve from memory that AT encodes 14 and then the trial has been reduced to a trial of condition 1.
For condition 2 there is a transformation, so the model will have to change the order of the names. Upon encoding the instruction for this type of trial the model will make a retrieval request to find a chunk that says what the correct order of the names is. This chunk, kept in the imaginal buffer, will only contain information of the correct order, there are no names associated to it yet. Based on the correct order of names, the three names will be retrieved from memory in the correct order. So for example, if the second and third name have to swapped, then the model will try to retrieve the name with the third position when looking for the

second name to answer and vice versa. The model will retrieve the three names in the order it deems correct and will click the mouse to indicate it is ready to answer. The model will first complete its answer and keep track of this in the imaginal buffer. Participants were also told, and forced by the short answering period, to formulate and answer before actually giving the answer.

Now for condition 4 the process is the exact same as for condition 2 except for the extra retrieval that is made to convert the letters in the instruction to the numbers first. So after the letter instruction has been translated to the number version, the model will try to find the correct order of the names. In condition 1 and 3 the correct order is immediately hard-coded in the imaginal buffer, which keeps track of the problem state, so there is no retrieval request. In condition 2 and 4 however a chunk of type switch is requested as was introduced above and using this the correct order of the names will be stored in the imaginal buffer.

This previous stage of the model was different for all conditions but for the remainder of the model it will be identical for the conditions again. So at this point the model has stored the three names with their respective position in memory and knows the correct order in which it wants to answer in the imaginal buffer.

The imaginal buffer contains the correct order of the names, but not the names themselves. So now using the positions in the imaginal buffer the correct name will be retrieved for each position. Once again the activation for these chunks is high enough to always be correct, but this implementation would allow for the forgetting of the names and their position. Imagine a trial of condition 1 where no names have to be switched then the model first makes a retrieval request for the name with position one. This name will be stored in the imaginal buffer and then the name in the second position will be requested and stored in the imaginal buffer. The same happens for the third name and once the model has a complete answer it clicks to indicate readiness to answer. For conditions with a transformation the exact same production rules execute which works because in that case the imaginal buffer would have the correct order of names. So if the correct order is 1,3,2 then these production will retrieve the first, third and second name to create the answer.

To answer, the model will start reading at the top and move downwards until it has found the first name from its answer, moves the mouse there and click. This will be repeated until all three names have been clicked and then the model is presented with feedback. This part always takes the same amount of time as the names aren't actually shown in a random order in the ACT-R version of the experiment, but the model still searches for the correct names as the participants would. The model is always correct so the feedback is not being read.

To summarize: for all four conditions the model goes through the same process except for the process of reading the instruction to finding the order in which to answer the names. Reading and storing the names and instruction, retrieving the names in the correct order and answering are the same for all 4 conditions. So the difference per condition depends on the extra retrieval request in the conditions where the instructions are in the form of letters. And when no transformation has to be performed the model instantly knows the correct order whereas in the other conditions the model has to retrieve the correct order from memory. So both distinctions between conditions (retrieval and transformation) rely on extra retrieval requests.

A number of parameters were of importance for the functioning of this model. First, because it is assumed that the model can always retrieve the chunks it has stored in declarative memory a negative retrieval threshold was set. Another factor that was of importance is the latency factor which influences the overall speed of retrievals. Since the model relies quite heavily of the retrieval mechanism this value had to be tweaked to match the human data as close as possible.

3.3 Convolution

The model performed the experiment but the resulting data from the ACT-R model is not pupil data, but the activation of the modules. For each module there is an activity pattern where 1 means

that the module is active and 0 means it is not, so it is in essence a string of pulses. Using the activity of the separate modules the pupil data for each module will be simulated and the combination of modules that best explains the the human data will be found. Hoeks en Levelt (1993) discovered an accurate method to convert this attentional input into pupil response data using a convolution.

In figure 3.1 the response to a single pulse of activity at $t = 0$ can be seen. The plot describes the response of the pupil and also demonstrates the slowness of the pupil. The pupils' response to activity follows a gamma distribution with two parameters, namely n and t.max. n describes the shape of the response and t.max describes where the maximum pupil response occurs, so where the peak of the function is. (Hoeks and Levelt, 1993) found two optimal values namely n=10.1 and t.max=930 which proved optimal across a range of tasks and subjects. It is important to note that the pulses have an additive effect on the pupil response so two pulses shortly after each other will be combined and will produce a bigger response. (Figure 3.2)

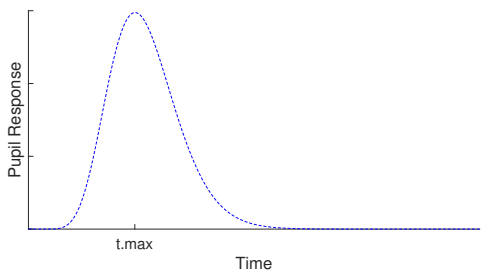

Figure 3.1: The response function of the pupil where at time 0 there is an activity impulse

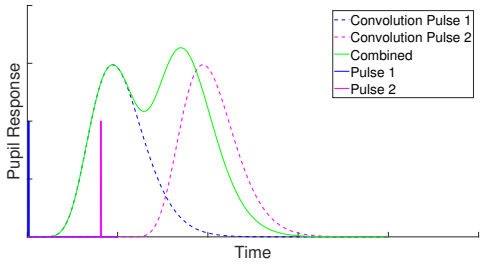

Figure 3.2: Two impulses shortly after each other will result in an additive effect in the convolution output. The convolved output for the separate pulses as well as the combined output can be seen

4 Results

4.1 Experiment results

First the accuracy and reaction times per condition were calculated, these can be seen in figure 4.1 and 4.2. The percentage of correct experimental trials was used to measure accuracy, practice trials were ignored. The names task proved to be quite a challenging task as the accuracy score in condition 4 is only 46% and a number of participants had to be removed due to low accuracy scores. Only correct trials were used in the analysis and trials where the reaction time of the participant was longer than 2 standard deviations from the mean were considered outliers. Trials with too many blinks or saccades yielded no useful data and were removed as well. Based on these measures (correctness, blinks and standard deviation) 634 trials of the 1824 trials from subjects were considered outliers and 1190 were used for the analysis. The analysis was performed on these remaining trials from 19 subjects.

The response times show that condition 1 was considerably easier for participants and condition 4 was significantly harder. The second and third condition show very similar reaction times which seems to indicate that a single transformation or retrieval is of the same complexity but performing both is notably more complicated. A pairwise t-test was executed using the Holm p-value adjustment method and it showed that there was a significant difference between each condition ($p < 0.001$). The

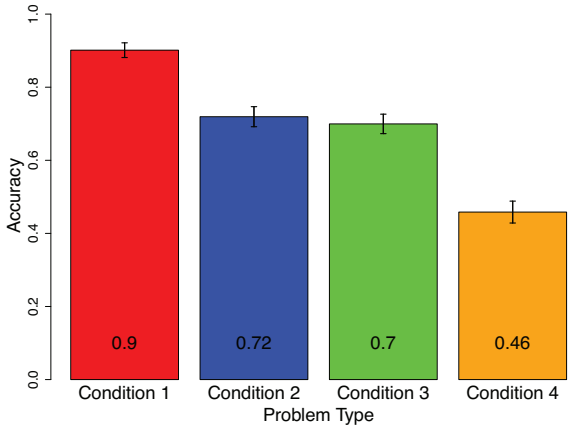

Figure 4.1: The accuracy per condition for the subjects

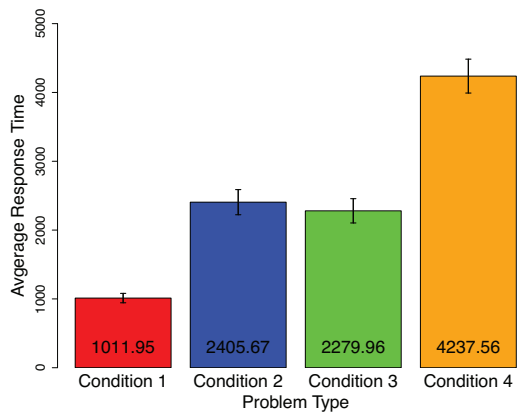

Figure 4.2: The reaction times per condition for the subjects

accuracy scores are in line with this observation and show a similar pattern.

The pupil size was measured at a frequency of 500Hz during the experiment and all the data was down-sampled to 250Hz. Throughout the experiment blinks and saccades occurred during which no measurements could be made. Saccades or blinks were linearly interpolated using 120 ms, or 30 measurements, before and after these gaps in the data.

The pupil responds quite slowly so to ensure that trials do not influence each other and the pupils can go back to their 'resting dilation' a fixation of 1000ms is shown at the end of each trial. During this period the pupil should be able to revert back to its baseline. Each trial begins with a fixation of 400-600ms as well, so there is a total of 1400-1600ms of rest between trials. During this fixation at the beginning of the trial the baseline for the trials is measured. 400ms of pupil dilation is measured in this period before the names are shown because at this stage in the trial there is no way for the participant to know in which condition they are. The measurements of these 400ms are averaged and this value is subtracted from the entire trial to make sure the change in pupil size is because of the trial and not another cause.

As can be seen in figure 4.4 and 4.7 the pupil responds to the experiment and the response depends on the difficulty of the trial. This can best be seen in the response locked plot. In the stimulus locked plot it is evident that the pupil responds to the instruction given in all conditions.

In order to investigate the effect of problem condition on pupil dilation in the experiment an ANOVA was performed. Retrieval and transformation, the factors which separate the conditions, were tested whether they had a significant effect on reaction time and accuracy. For reaction times the results show a significant main effect for transformation $F(1, 18) = 35.74, p < 0.001$ and a significant effect for substitution $F(1, 18) = 43.8, p < 0.001$. The interaction effect between transformation and substitution was not significant $F(1, 18) = 0.032, p = 0.86$ For the accuracy scores of the subjects a significant main effect was also found for transformation $F(1, 18) = 36.82, p < 0.001$. A significant main effect was found for substitution of the accuracy score as well $F(1, 18) = 43.67, p < 0.001$. The interaction effect between substitution and transformation proved not significant $F(1, 18) = 2.25, p = 0.151$.

4.2 Model results

The ACT-R parameters that were adjusted for the experiment were the retrieval threshold and the latency factor. It was assumed that the model was always correct and could always recall its declar-

ative knowledge, so the retrieval threshold was -5 and the latency factor was 0.17 to account for the correct length of retrievals to match the data. In figure 4.3 the reaction times of the participants and the model can be seen. The reaction time from the model approximates the subjects data well for the first three conditions, however for the last condition there is a discrepancy between model and subject data. The overall trend of increasing reaction times is present in the model data however.

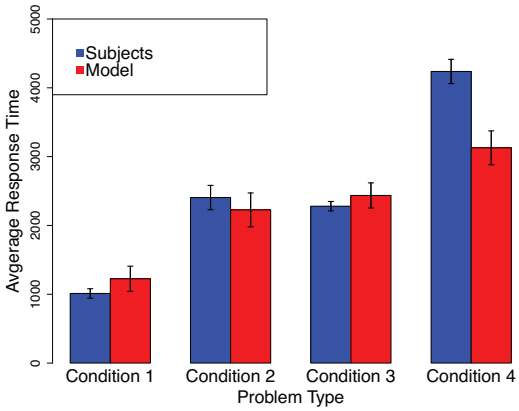

Figure 4.3: The response times of the subjects compared to the model

The model provided data by performing a trial of each condition. In order to make an accurate comparison the models' activity was linearly scaled to match the participants data in length such that the onset and duration of each trial was the same. Thus for each individual trial performed by the participants the model's trial was scaled to match in length meaning that the model has an equal amount of trials as the subjects after scaling. Accordingly the models' activity needed to be converted to pupil dilation data. The activity of each ACT-R module was separately converted into pupil dilation data using the convolution method described by Hoeks & Levelt. A linear model was fitted to see which combination of module's dilation data best matched the subjects pupil dilation data. The linear model provided a coefficient for each module that indicated the contribution of that module to the pupil dilation. The best linear model was used to create a prediction by multiplying the coefficients found with the appropriate module and summing the results for all modules. This resulted in a single data set representing the pupil dilation prediction data based on the cognitive model.

Linear models were created with intercept where this is considered to represent all the processes not captured by ACT-R because the intercept does not directly represent a function of the ACT-R architecture. The linear models were trained on all the data at once where t.max=930 and n=10.1 were used as variables in the convolution. Initially a model was trained using all 5 modules and accordingly models using subsets of modules were trained.

The initial best linear model found used all 5 modules and trained on all data at once. (Table 4.1) The predictions based on this model can be seen in figure 4.5 for stimulus locked predictions and figure 4.8 for response locked predictions. The prediction created does not seem to predict the subjects' data very well, the shape of the data is not similar and the curves should be shaped similarly across conditions. So this results seems to have little predictive power. Due to the prediction not matching the actual data, a different approach was used: linear models were trained on subsets of the ACT-R modules to find the best combination. This resulted in a better linear model that only used the imaginal, declarative and procedural module to predict the pupil dilation data. Please refer to table 4.2 for the coefficients of this linear model and to figure 4.6 for the stimulus locked and figure 4.9 for the response locked predictions. The predictions created with this new method proved to be significantly better than the initial model. In the response locked plot it can be seen that the prediction resembles the shape and timing of the subject pupil dilation data substantially better. So ultimately the linear model based solely on the procedural, declarative and imaginal module proved best. In the remainder of the paper the original model based on all 5 modules will be referred to as LM1 and the model based on the declarative, procedural and imaginal module will be referred to as LM2.

Module	Coefficient	Standard Error	T-Value
Intercept	-0.067397	0.0012452	-54.125
Imaginal	0.048075	0.0012196	39.42
Manual	0.19324	0.0011664	165.67
Production	-0.026549	0.0010915	-24.324
Declarative	0.025532	0.00068031	37.529
Visual	-0.076523	0.0011536	-66.335

Table 4.1: Linear model LM1 fitted with all modules

Module	Coefficient	Standard Error	T-Value
Intercept	-0.035951	0.0011628	-30.916
Imaginal	-0.12016	0.00058333	-205.99
Procedural	0.079878	0.00064822	123.23
Declarative	0.019246	0.00055333	34.782

Table 4.2: Linear model LM2 fitted with imaginal, procedural and declarative modules.

4.3 Stimulus Locked

The predictions as well as the subjects' data can be viewed from two perspectives: stimulus locked and response locked. In the first the data is centered around the stimulus, namely the instruction, and in the response locked plot the pupil dilation leading up to the response can be seen. The last perspective is of interest as pupil dilation tends to reach its maximum just after responding. In figure 4.4 the stimulus locked pupil dilation data for the subjects can be found, in figure 4.5 the predicted pupil dilation data from LM1 and in figure 4.5 the predicted pupil dilation from LM2 in figure 4.6. Pupil dilation for the separate modules can be found in the appendix.

An interesting aspect in the LM1 prediction is the dip around 1000ms in the dilation as there is only a very subtle dip around 100 ms after the stimulus in the original data. This dip is likely caused by the imaginal module: after encoding the names and instructions the imaginal module is not in use for a while until the answer needs to be formulated at which point activity spikes again. The imaginal module's pupil data can be seen in figure A.6. The pupil dilation for the remaining models can also be found in the appendix. The prediction based on LM1 does predict the overall rise in dilation and lack of decay afterwards well. However the pupil dilation starts increasing too late: in the data the first condition peaks at 950ms whereas in the prediction the peak occurs around 2000ms. The lack of decay in the pupil response is explained because participants, and the model, are busy formulating their answer so a strong decay of dilation will not happen until the response has been given.

The prediction based on LM2 is considerably different from LM1 and seems to reflect the actual data slightly better. In the prediction the dilation rises quickly around the 900ms mark for all 4 conditions meaning that pupil reflects the effort exerted by participants after they read the instruction and start working on the problem. Two clear issues arise when comparing the prediction to the data: the dilation prediction for the first condition is too high and for condition 4 the prediction is too low. This is likely the result from the same underlying problem: the prediction relies quite heavily on the procedural module. For condition 1 there are no extra retrieval requests in the ACT-R model in between the production rules so this is the condition in which the production rules succeed each other faster than then other conditions. Because of the additive effect of transforming activity to pupil dilation the dilation will be larger for the first condition. The same mechanism explains why the dilation is the lowest for the fourth condition, the largest number of retrieval requests are made in that condition so the procedural module has the largest 'cool down time'. And because of the reliance on the procedural module by the prediction this will result in lower dilation.

4.4 Response Locked

Response locked plots are interesting because they generally contain the maximum pupil dilation for a trial and this might give insight in which processes are responsible for the peak of pupil dilation. The models' prediction will show which modules contribute to the maximum pupil dilation. Figure 4.7 shows the pupil dilation data for the subjects, figure 4.8 shows the predictions made by the LM1

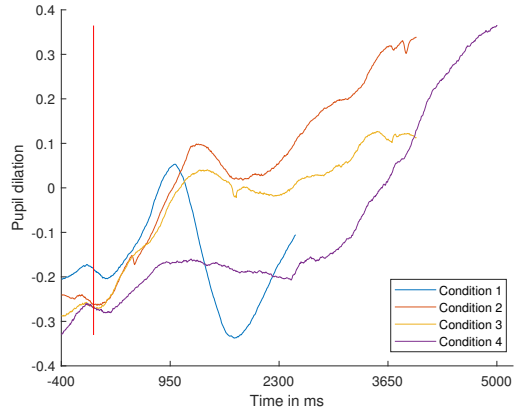

Figure 4.4: Stimulus locked view of pupil dilation data over all subjects

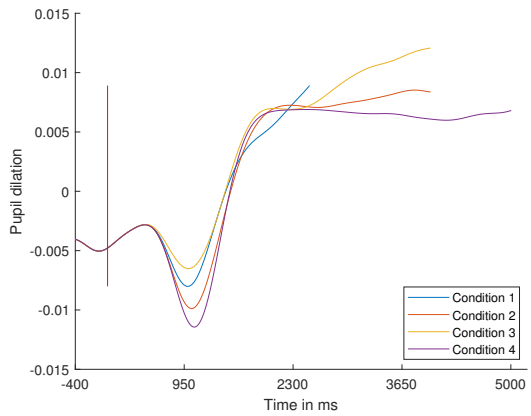

Figure 4.5: Stimulus locked view of LM1 model prediction of pupil dilation

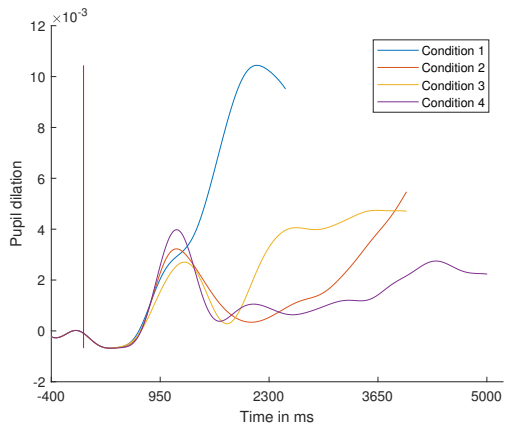

Figure 4.6: Stimulus locked view of LM2 model prediction of pupil dilation

model and figure 4.9 shows the prediction by LM2. In the subject data a pattern where dilation rises until 300ms after they indicate readiness to answer can be seen. Afterwards the pupil dilation decreases quickly.

The prediction from the LM1 model seems to display this trend of rising pupil dilation until an answer has been provided as well, however there is variation in the shape of the data between the conditions whereas this seems quite similar in the actual data. The rise in pupil dilation is likely the result of the manual module as its activity steeply increases around 350ms after response. In the LM1 linear model the manual module has the largest coefficient and this large influence might seem strange at first sight.The motor module is only used at the very end of the trial and during all the problem solving the activity of the manual module is near zero so it is of little influence to the prediction. Then near the end of a trial subjects experience their maximum pupil dilation and the manual module's activity quickly increases at the end of a trial as well. Thus a large coefficient for the manual module has little influence on the bulk of a trial but helps to create the steep incline in pupil dilation near the end of a trial. Another peculiarity is the fact that the procedural module (figure A.3) has one of the smallest coefficients despite its curve closely resembling the actual data.

The prediction from LM2 in figure 4.9 resembles the subject data a lot better than the LM1 prediction. The prediction now shows similar curves for each condition which steeply increase and reach their peak around 350ms after their response. The prediction also shows a quick decay after the peak but the actual data has a stronger decline. Here the same issue can be seen: condition 6 has the largest response and this is likely due to the same reason as in the stimulus locked plots. Condition 3 is another issue: its dilation is lowest even though it should be around the level of condition 2. This can be explained due to linearly scaling the trials: each trial is scaled to match the trial of the participant in length. Overall the model was too fast for condition 3 and thus was scaled down more strongly than the other conditions. This results is a shorter period of activation and because of the additive effect of pupil dilation this results

in less predicted dilation. Another issue to note is that the peaks of the prediction do not align which could be the result of linear scaling as well. Overall the prediction by LM2 resembles the data significantly better and is a more promising result.

5 Discussion

Mental effort is one of the many terms used to describe what is reflected by pupil size in the literature even though it is unknown what this mental effort entails. This study tried to uncover which cognitive processes are reflected by pupil dilation through a model based approach. An experiment was conducted for this study in which participants performed a complex information processing task and a cognitive model was built to simulate this. Based on the models' activation a prediction for the subject pupil dilation could be made and this reflected which cognitive processes were of importance for pupil dilation. The results showed that the best prediction of the subjects data used the procedural, declarative and imaginal module. So mental effort in the names task seems to be best explained by procedural memory, declarative memory and problem representation.

The amount of pupil dilation depends on condition in the data so the difficulty of a trial or the mental effort required is reflected by the pupil. After subjects indicated readiness to answer the dilation decreases quickly indicating that the subjectsstop problem solving at that stage.

A cognitive model was built using ACT-R to investigate which cognitive processes influences pupil dilation. The models' activity was transformed to pupil dilation data using a convolution method defined by (Hoeks and Levelt, 1993). Using linear models the combination of ACT-R modules to best predict the subjects' dilation data was found. Pupil dilation was best explained using the imaginal, procedural and declarative modules and the model based on these models predicted the data best. The models' prediction reflects the general trend and shape of the data but does not provide a perfect prediction. Possible causes and issues will be discussed below.

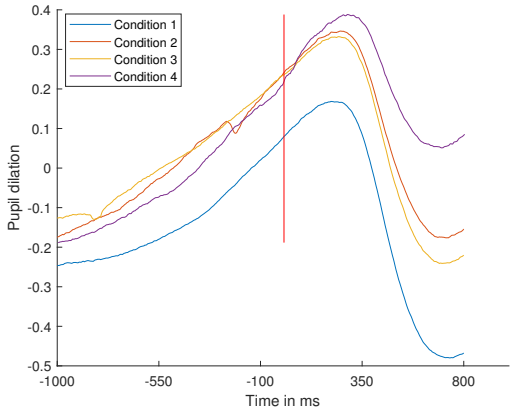

Figure 4.7: Response locked view of pupil dilation data over all subjects

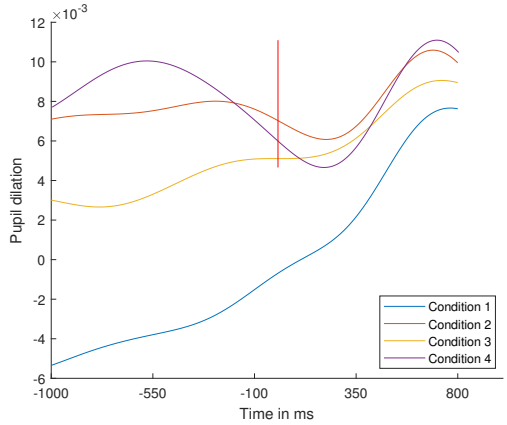

Figure 4.8: Response locked view of LM1 model prediction of pupil dilation

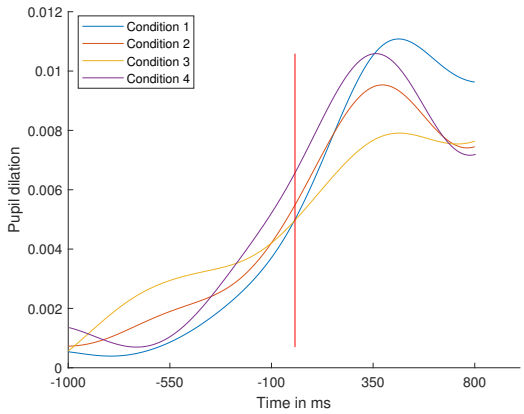

Figure 4.9: Response locked view of LM2 model prediction of pupil dilation

The model found to predict the best fit for the pupil dilation is not perfect and this might be due to a number of reasons. The cognitive model built in this study might not be an accurate reflection of how the subjects execute the task. This is hard to determine but there are indications that the model might be less than perfect. Another possible problem is that there might be factors of influence on pupil dilation that are not captured in ACT-R. There is no attention in ACT-R for example even though this is definitely required in the task. Another process that might be of influence is stress as subjects mentioned they thought the experiment, especially in the fourth condition, was quite stressful.

In general the data gathered during the experiment was far from perfect. This might be due to the difficulty of the task used in the experiment but it could also be due to a flaw in the experimental design. The behavioral data confirms that subjects had difficulties with the experiment, especially the fourth condition proved complex.

An interesting note is that some participants that are included in the analysis actually 'gave up' for the fourth condition. This is noticeable in their plots as the first 3 conditions appear in the correct order, but the fourth condition has the lowest dilation, meaning that they exerted the least mental effort. These participants seemed to have accepted the fact that they did not know trials of the fourth condition.

After completing the experiments subjects mentioned something of interest about the third condition. During the experiment subjects tended to realize that when they saw an instruction of this condition they could go straight to answering. Participants thus seemed to improve for this condition during the experiment but their performance for the other conditions remains similar. This might have unforeseen influences on the results because this is the only condition for which performance increases during the experiment.

In the predictions made by the model based on the imaginal, procedural and declarative modules the dilation for condition 1 is too high and too low for condition 3. This is due to the design of the cognitive model and the process of transforming the models' activity to pupil dilation. The prediction relies heavily on the procedural module and condition 1 is the condition with no extra retrieval requests meaning that the production rules succeed each other faster than in the other conditions. Due to the additive effect of pupil dilation, this results in the highest predicted dilation for condition 1. The model predicted a lower dilation for the third condition due to the scaling of the trials. The model performed trials of the third condition too fast in general meaning that these trials were scaled down more strongly than the other conditions. Due to the additive effect of pupil dilation and the smaller amount of activation because of the scaling this resulted in a curve that is too low. This is in line with the previously mentioned issue that the model is not perfect and reflects that the model maybe should have executed trials of this type differently than it did.

This study uncovered possible cognitive processes influencing pupil dilation, thus reflecting mental effort, through a model based approach. The predictions found by the model do not provide a perfect fit but provide valuable insights into which cognitive processes may be of influence on pupil dilation. The model based approach proved to be a useful method for the study, but improvements need to be made in order to confirm the findings of this study. The task used in the experiment might have been too difficult, ACT-R might not have captured all the factors influencing pupil dilation and the model used in the experiment may need improvement.

Future research can use a model based approach to find cognitive factors influencing pupil dilation. Specifically for the names task the model needs improvement to gain predictive power. In general, research could focus on factors influencing pupil dilation not included in this study such as attention, stress and fatigue. The model based approach used tries to link the model's activity to specific brain areas so an fMRI study could provide more reliable data by measuring brain activity directly.

References

S. Ahern and J. Beatty. Pupillary responses during information processing vary with scholastic apti-

tude test scores. *Science*, 205:1289–1292, 1979.

J.R. Anderson. *How can the human mind occur in the physical universe?* Oxford: Oxford University Press, 2007.

J.R. Anderson and et al. Human symbol manipulation within an integrated cognitive architecture. *Cognitive Science*, 29:313–341, 2004.

J.R. Anderson and et al. Information-processing modules and their relative modality specificity. *Cognitive Psychology*, 54:185–217, 2006.

J. Beatty. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91:267–292, 1982.

J. Beatty and B.L. Wagoner. Pupillometric signs of brain activation vary with level of cognitive processing. *Science*, 199:1216–1218, 1978.

J.P. Borst and J.R. Anderson. A step-by-step tutorial on using the cognitive architecture act-r in combination with fmri data. *Journal of Mathematical Psychology*, 76:94–103, 2016.

M.J. De Jager. Discovering the Cognitive Factors driving Pupil Dilation: A Model-Based Analysis of an Algebra Task, 2018.

W. Heinrich. Die aufmerksamkeit und die funktion der sinnesorgane. *Zeitschrift fr Psychologie und Physiologie der Sinnesorgane*, 9:342–88, 1896.

E.H. Hess. *The tell-tale eye: How your eyes reveal hidden thoughts and emotions.* Van Nostrand Reinhold, New York, 1975.

E.H. Hess and J.M. Polt. Pupil size in relation to mental activity during simple problem-solving. *Science, New Series*, 143:1190–1192, 1964.

B. Hoeks and W.J.M. Levelt. Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research Methods, Instruments, Computers*, 25:16–26, 1993.

M. A. Hogervorst, A. M. Brouwer, and J. B. F. Van Erp. Combining and comparing eeg, peripheral physiology and eyerelated measures for the assessment of mental workload. *Frontiers in Neuroscience*, 8:1–14, 2014.

D. Kahneman and J. Beatty. Pupil diameter and load on memory. *Science, New Series*, 154:1583–1585, 1966.

M.A. Kursawe and H.D. Zimmer. Costs of storing colour and complex shape in visual working memory: Insights from pupil size and slow waves. *Acta Psychologica*, 158:67–77, 2015.

S. Matht, D. Schreij, and J. Theeuwes. Opensesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2):314–324, 2012.

J.M. Schiff. *La pupille considre comme esthsiomtre.* Hachette Livre, 1875.

S.R. Steinhauer and G. Hakerem. The pupillary response in cognitive psychophysiology and schizophrenia. *Annals of the New York Academy of Sciences*, 658:182–204, 1992.

E. Van der Meer, R. Beyer, J. Horn, M. Foth, B. Bornemann, J. Ries, J. Kramer, E. Warmuth, H.R. Heekeren, and I. Wartenbruger. Resource allocation and fluid intelligence: insights from pupillometry. *Psychophysiology*, 47:158–169, 2010.

P. Van Der Wel and H. Steenbergen. Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin Review*, 25: 20052015, 2018.

A. Westbrook and T.S. Braver. Cognitive effort: A neuroeconomic approach. *Cognitive, Affective, Behavioral Neuroscience*, 15:395415, 2015.

A Appendix

Response locked module plots

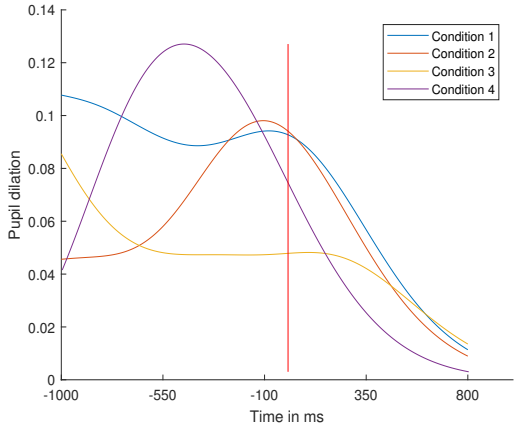

Figure A.1: Response locked activity of imaginal module

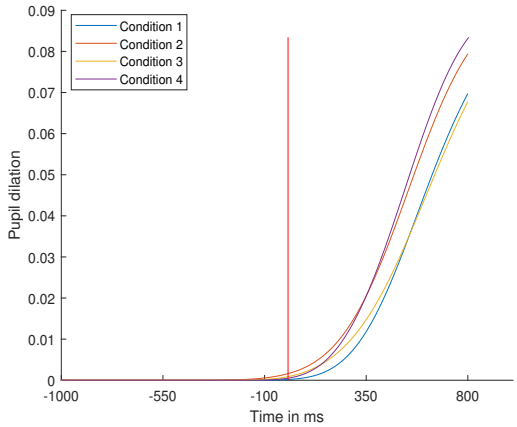

Figure A.2: Response locked activity of manual module

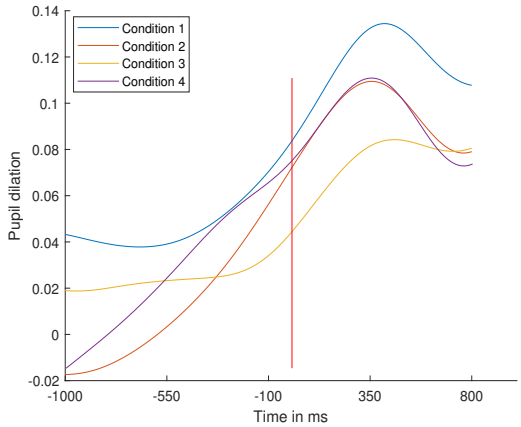

Figure A.3: Response locked activity of procedural module

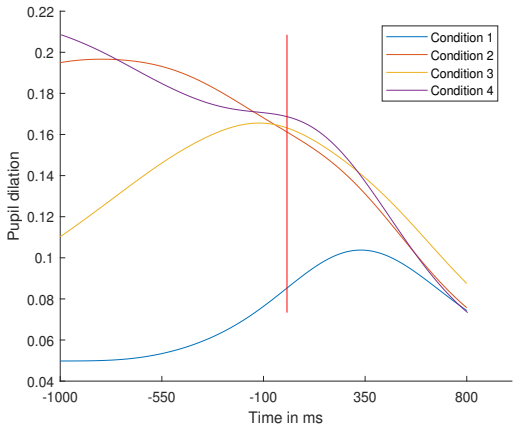

Figure A.4: Response locked activity of declarative module

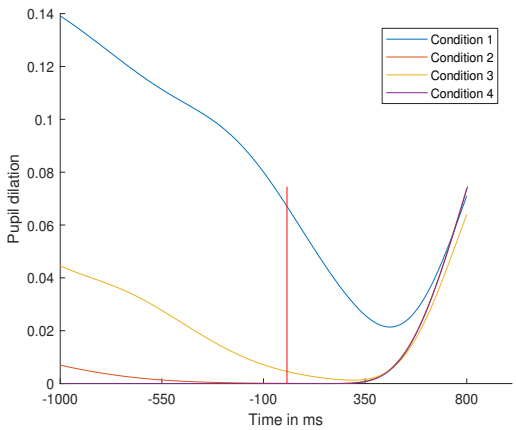

Figure A.5: Response locked activity of visual module

Stimulus locked module plots

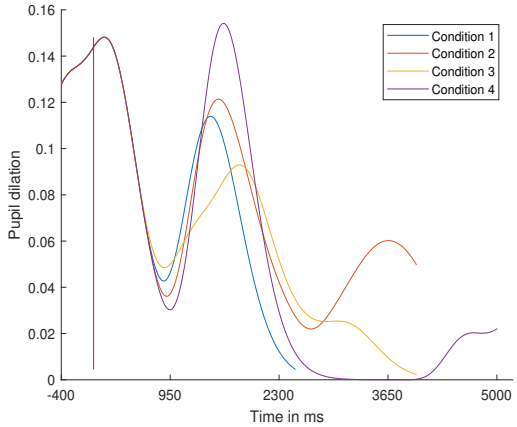

Figure A.6: Stimulus locked activity of imaginal module

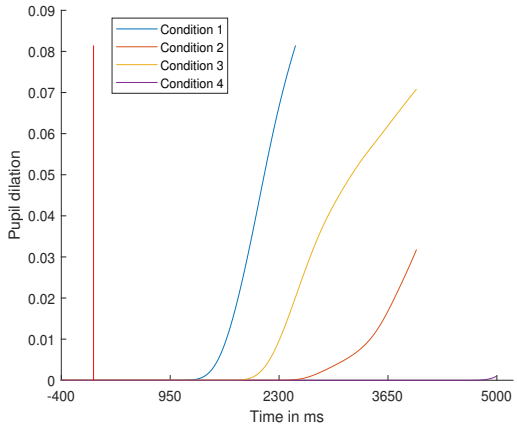

Figure A.7: Stimulus locked activity of manual module

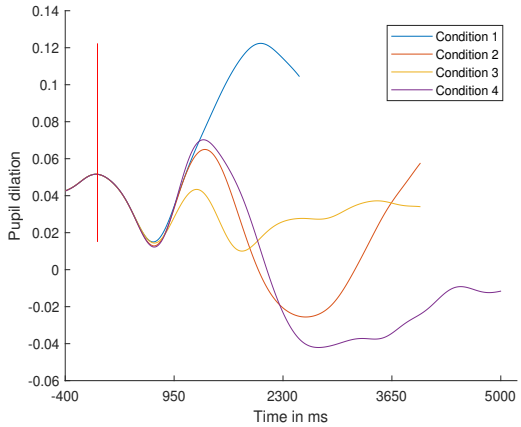

Figure A.8: Stimulus locked activity of procedural module

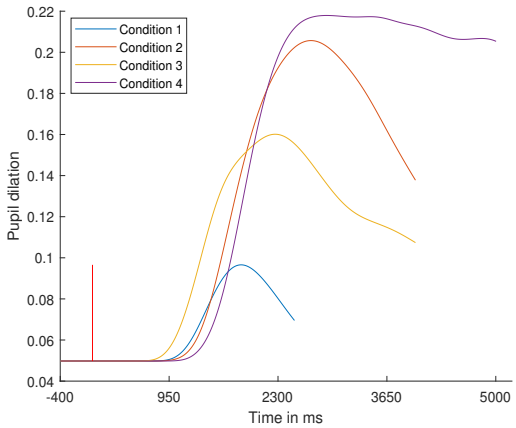

Figure A.9: Stimulus locked activity of declarative module

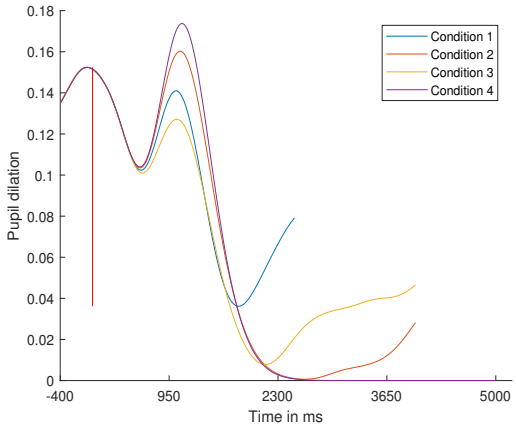

Figure A.10: Stimulus locked activity of visual module