Using eye tracker data to assess human empathy and attention

Marta Ewa Dubas

Thesis submitted for the degree of
Master in Applied Computer and Information Technology (ACIT)
30 credits

Department of Computer Science
Faculty of Technology, Art and Design

OSLO METROPOLITAN UNIVERSITY

Spring 2022

Using eye tracker data to assess human empathy and attention

Marta Ewa Dubas

© 2022 Marta Ewa Dubas

Using eye tracker data to assess human empathy and attention

http://www.oslomet.no/

Printed: Oslo Metropolitan University

Abstract

In this thesis we explore up to which extent eye pupil diameter, measured through an eye tracking device, reflects levels of empathy among individuals during a specific task performance. The analysis is conducted based on the eye tracker data collected during an experiment designed by Bhurtel et al [1]. Empathy is a complex phenomena defined by various definitions and it is challenging to measure it. Here, we analyze the hypothesis that empathy correlates with individual's pupil size. The size of the eye pupil is the indicator of attention (positive correlation), which means that the existing positive correlation between the empathy and pupil size could indicate an existing correlation between attention and empathy. The results show that from the analyzed data we could not discover correlation between empathy scores and pupil size, hence it was not possible to investigate further the relation between empathy and attention. Moreover, a critical discussion on the experimental design is presented. In particular, we describe an alternative experiment to better assess if empathy is correlated with attention. The new experiment could help in collecting better data and trigger an emotional reaction among the participants, which is necessary to invoke a higher level of empathy.

The research for this thesis as well as its writing was entirely done between January and May 2022.

Acknowledgments

I would like to thank my research supervisors, Pedro Rego Lencastre e Silva and Pedro Lind, for their guidance throughout this project. Thank you very much for your dedication, involvement, support, knowledge, ideas and patience.

Contents

Abstract **i**

Acknowledgments **ii**

1 Introduction **1**
1.1 Motivation 1
1.2 Problem Statement 2
1.3 Structure of the thesis 3

2 Background and state of the art **4**
2.1 Eye movements 4
2.2 Pupil diameter 6
2.3 Eye tracking technology 7
2.4 Main theoretical background for time series analysis 8
2.5 Statistical hypothesis testing 10
2.6 Use of eye tracking in research 12
2.6.1 Eye tracker data in attention research 12
2.6.2 Eye tracker data in empathy research 16

3 Data and methodology **18**
3.1 The experiment by Bhurtel et al: design, protocol and ethical approval 18
3.2 Data sets to be analyzed 20
3.3 Pre-processing of the data sets 21
3.4 Methodology and analysis framework 24

4 From pupil size to empathy scores **27**
4.1 Typical diameters and uncertainties for each participant . . . 27
4.2 Power analysis and hypothesis testing 31
4.3 Summary 33

5 Conclusions and future directions **34**

Appendices **38**

A Full results for test group **39**

B Full results for control group **52**

C Empathy score questionnaire **65**

D Link to GitHub code used in analysis of the data sets **68**

List of Figures

2.1 Anatomy of the eye. Figure taken from http://www.eyesightresearch.org/. 4
2.2 Muscles responsible for the changes of the pupil size. Figure taken from Fig.2 in Ref. [19]. 6
2.3 *Search coil*, a type of an eye-attached eye tracker 7
2.4 Example placement of the electrodes in an experiment using EOG. Figure taken from Fig.1 in Ref. [22]. 8
2.5 Tobii eye tracker, attachable to the screen. Figure taken from https://m.media-amazon.com/images/. 8
2.6 Normal distribution with mean and standard deviation. Figure taken from Fig.1 in Ref. [24]. 9
2.7 Example plot of non-stationary strongly seasonal data with an increasing trend. Figure taken from Fig.2.3 in https://otexts.com/. 10
2.8 Example of an outlier in time-series data. Figure taken from https://towardsdatascience.com/. 10
2.9 Research areas where eye tracking is applied. Figure taken from Fig.1 in Ref. [28]. 12

3.1 Design of the experiment. 19
3.2 Left and right pupil size changes for one participant during one attempt. 22
3.3 Scatter plot of the left and right pupil diameter for one participant one during attempt. 23
3.4 Scatter plot of the left and right pupil diameter for one participant one during attempt. Grouped by event type. . . 24
3.5 Illustration of a slice of the series for each attempt of each participant, including the label (fixation, saccade, unclassified) and mean of pupils diameter. 24

4.1 Average pupil size with standard deviation for participant 1. For all the results see Appendix A and B 28
4.2 Test group: scatter plot of the average pupil size vs empathy variations: a) empathy before, b) empathy after and c) difference empathy after-before before and after the intervention. 29

4.3 Control group: scatter plot of the average pupil size vs empathy variations: a) empathy before, b) empathy after c) difference empathy after-before and d) average empathy group before and after the intervention. 30

List of Tables

3.1 Final result matrix used in the analysis. 25

4.1 Correlation coefficient for pupil diameter and empathy scores in test and control group. 28
4.2 Mean, standard deviation and number of samples in test and control group. 31
4.3 Cohen's descriptors for magnitude. 32
4.4 Calculated pooled standard deviation and effect. 32
4.5 Empathy scores before and after the intervention in test and control group. 32

Chapter 1

Introduction

1.1 Motivation

"Where words are restrained, the eyes often talk a great deal" [2]. This sentence with its authorship attributed to Samuel Richardson had only had a metaphorical meaning in over 200 years. It was about to change in the middle of 20^{th} century when an important invention altered the way we are able to look at the eyes. Although the attempts to build an eye tracking device started as early as at the end of 19^{th} century, it was not before 1950s when the first less intrusive eye trackers were built [3]. The eye tracking research was accelerated by the invention of the electrooculogram - a new technique for measurement of nerve impulse in the eye, first named and described by Elwin Marg in 1951 [4]. The development in eye tracking techniques together with the development of the more powerful and accessible computers, opened the door for the new areas of the research. The eye tracking research blossomed in the 1970s and 1980s and is still evolving.

The data recorded by a modern eye tracking device can be used to evaluate brain functions [5], explain cognition and behavior [6], measure spatial attention [7] and recognize feelings such as empathy [8].

Empathy is one of the brain capabilities which can be estimated by analysis of an eye tracker data. Empathy is essential in building successful relationships. The word empathy comes from the Greek word *empatheia* which means *"feeling into"* and was used to describe the feelings a spectator can experience by looking at a piece of art [9]. Today's meaning of the word can be described as *"our capacity to grasp and understand the mental and emotional lives of others"* [10]. In some social contexts it can play an extremely important role, for instance in a doctor-patient or caregiver-patient relationship. The empathy is a foundation to build the trust, which is necessary for the patient to tell their story so the doctor or the caregiver can understand the patient's problem and by the same provide her/him the best possible care [11].

There are many documented experiments focusing on the human capabilities to learn empathy, many of which are rooted in the medical context and focusing on increasing the empathy towards the patient.

Results of some of the research indicate the higher level of empathy as a consequence of the experiment [12], [13]. In other experiments the level of empathy among the participants decreases, or the increase is very small contrary to the expectations, like in the experiment of Bhurtel et al [1]. Bhurtel's experiment is described in detail in Chapter 3.1. The results of this research raise questions about the reason for the unexpected decrease of the empathy among the participants. Can the decreasing empathy be caused by the decreasing attention of the participants? How do attention and empathy relate with each other?

The meaning of the term *attention* is widely discussed and throughout the years of the research is has gained multiple explanations. According to the traditional views, attention is a set of cognitive mechanisms maximising the efficiency of the humans' ability to process and store the information [14].

1.2 Problem Statement

As well as empathy, attention cannot be directly measured. Both capabilities can be estimated based on quantitative indicators. Eye tracker data can provide some good candidates for such indicators to assess both the empathy and attention. If it is possible to prove a relation between the empathy and attention, it could help to assess empathy in an indirect way. In this thesis eye tracker data from the previous research [1] is used to analyze the possible correlation between the lack of attention and decreasing empathy. In the experiment [1], participants were asked to answer a questionnaire, answer the experiment questions and answer the questionnaire again without having a break. This could possibly lead to the lower ability to concentrate and higher level of impatience. Also, contrarily to people with disabilities, participants were not trying to fulfil their actual needs, nor were they expecting any reward. The question builds on the suspicion the participants became tired and over-saturated with the experiment's tasks. The central research question of this thesis is

> **Are empathy and attention in individuals related with each other?**

Our starting hypothesis is

> **Empathy is positively correlated with eye pupil diameter.**

This hypothesis will be tested by using tools from statistical learning and data science. The research question can be detailed in the following sub-questions:

- How significant is this relationship?
- Which statistical tools for data analysis can better uncover the relationship between empathy and pupil diameter?
- What can a data scientist learn from this data?

To address these questions we will also make use of the following assumptions, which follow on from empirical knowledge (see Chapter 2):

- Pupil diameter is a measure of attention.
- Fixation duration is a measure of attention.

The attempt to answer the research question is based on the analysis of the data from the previous research [1].

1.3 Structure of the thesis

The rest of this thesis is organized as follows:

- *Chapter 2* introduces the background and summarizes state of the art. In this chapter we present the technical details and terminology regarding eye movements, eye tracker types and the characteristics of the data recorded by the eye tracking devices. Also in this chapter there is a review of the eye tracking related research focusing on studies concerning attention and empathy.
- *Chapter 3* describes the data sets to be analyzed and methodology, highlighting the previous research and the data [1] used as foundation for this thesis. Later, it presents the analysis of the data from the previous research in the context of finding the correlation between the empathy and attention.
- The results of the data analysis are presented in the *Chapter 4*.
- Finally, *Chapter 5* presents the discussion of the results, conclusion and suggests directions for the future work.

The research for this thesis as well as its writing was entirely done between January and May 2022.

Chapter 2

Background and state of the art

This chapter explains the terminology related to eye movements and eye tracking. It presents the most common types of the eye trackers with their basic technical characteristics. Statistical tools used in analysis of eye tracker data and the current state of the eye tracker based research related to the attention and empathy are also presented in this chapter. In the context of this thesis, an eye movement is understood as any shift of the position of the eye in its orbit. Pupil dilation concerns the muscle movements inside the eye, without affecting its position and therefore is not considered as an eye movement. This understanding is common in the field.

2.1 Eye movements

Human eye is a complex optical apparatus. The first documented description of the eye as an optical instrument comes from the end of 16^{th} century. Johannes Kepler, best known for the laws of planetary motion, was the first one to describe how the eye works. He did not recognize all of the parts of the process correctly, but he was not far from the today's description [15].

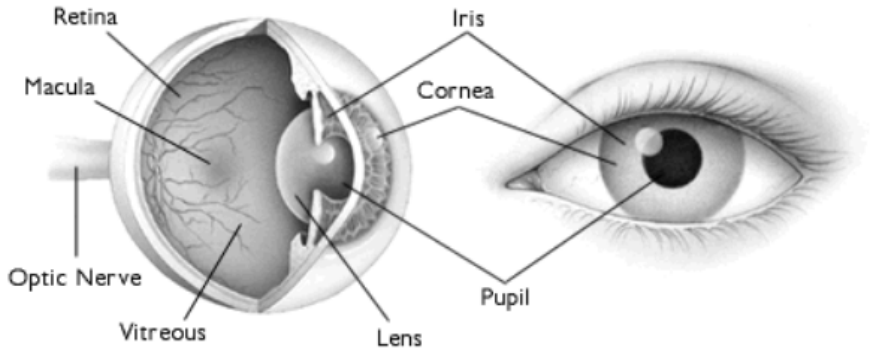


Figure 2.1: Anatomy of the eye. Figure taken from http://www.eyesightresearch.org/.

The light reflected by the objects enters the eye through the pupil. The pupil and iris are covered by cornea - the part of the eye that reflects the light. The lens flips the image 180 degrees horizontally and then the image is projected on the retina - the back of the eyeball. On the retina there are cones and rods - cells responsible for translating the light into electrical signal and sending them further for the processing. On the bottom of the retina, there is a special area called fovea. This area has a high concentration of the cones and has very high perceptively. Fovea is responsible for the sharp vision which means that the eye's position must be adjusted so the light can focus on fovea in order create a sharp image of what the eye looks at. The repositioning of fovea is the foundation of sharp vision and eye movements play a great role in this process. The anatomy of the eye is presented in Figure 2.1.

The eye's position is controlled by three pairs of muscles responsible for horizontal, vertical and torsional (roll) movements. Taxonomy of the eye movements and the definitions may differ depending on the sources. According to Andrew T. Duchowski [16], there are five basic type movements the eye makes when repositioning the fovea: saccades, smooth pursuit, vergence, vestibular and nystagmus.

Saccades are the fast movements happening in the process of repositioning the fovea (moving the eye to the next fixation location).

Smooth pursuits are the movements happening when the eye tracks a moving target, for instance a flying bird or the cars on the motorway driving in the same direction as the observer.

Vergence movements are disconjugate (not paired) movements responsible for the focus on a distant object.

Nystagmus movements are conjugate (paired) eye movements that can be divided into three types: *optokinetic, vestibular* and *physiological* (also called *tremor*). The two first ones are responsible for compensation of the retinal movement of the target. The role of the last one is not known [16], [3].

Fixations are an important part of the eye behavior registered by the eye tracker. There are many understandings of *fixation* and the definition of the term became a subject of a research itself [17]. In this thesis, a fixation is understood as a period when the eyes are still. This definition is technically imprecise since during the fixation the eyes still make miniature movements: *microsaccades*, *drift* and *tremor* to stabilize the retina in relation to an object. However, in the context of this thesis the understanding of the fixation as a state when the eye remains still is sufficient. Having explained the fixation, it is possible to explain *saccade* as a rapid eye motion from one fixation to another [3].

The eye movements can be used as a tool in mapping of the brain functioning. These movements have been a subject of the research

in the context of cognition (e.g. visual attention, working memory, task switching), and give insight into disorders like attention deficit hyperactivity disorder (ADHD) and autism.

2.2 Pupil diameter

While the eye is the living optical device, the pupil is the gate controlling how much of the reflected light can enter the device to be further processed on the retina. When looking into someone else's eyes from a short distance, the observer can see own reflection which could remind of a small doll - *pupilla* in Latin, and this is where the word *pupil* derives from [18].

By changing its diameter the pupil controls the amount of light projecting the image on the retina. The more light, the higher degree of diffusion and what comes with it - blurriness of the image. Lower amount of light allows higher visual acuity (sharpness). Human pupil size changes involuntarily and simultaneously in both eyes, even if only one eye is exposed to the changing light stimuli. The pupil dilation is controlled by two contrarily working muscles: *dilator pupillae* and *sphincter (constrictor) pupillae* shown in Figure 2.2.

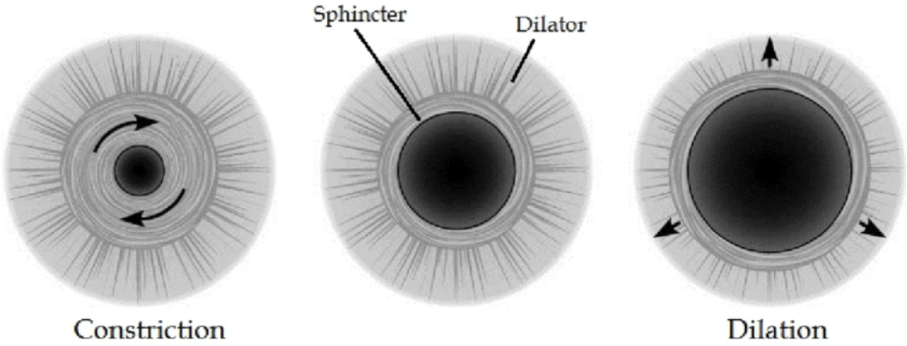


Figure 2.2: Muscles responsible for the changes of the pupil size. Figure taken from Fig.2 in Ref. [19].

The average pupil size in adults in comfortable, bright light conditions is between 2 and 4 mm in diameter [20, 18]. In dark conditions, the pupil can resize to 8 mm with average of 7 mm. Pupil dilation is not only dependent on the luminance factor. The pupil size changes also depend on the emotions and mental states. These changes have much smaller deviation than the ones caused by the changing light intensity. The measurements of the pupil size (pupillometry), as well as the eye movements can be an indicator for cognitive processes of the brain like memorizing or attention. Pupillometry is used to examine patients with a suspicion of brain or other neurological injuries. The measurements of the pupil size are currently used ind psychological and neurological research, and also in marketing and User Interface (UI) design.

2.3 Eye tracking technology

The first known tools designed to track the eye movements in the late 19^{th} century could be mistaken for torture instruments. They were uncomfortable and required a direct attachment of a part of the machinery to the participant's eye. The development of the less intrusive techniques based on the measurement of the reflection of the light in cornea in the second part of the 20^{th} sparked off the development of the eye trackers commonly used nowadays. The modern eye tracking devices can be divided in three basic categories based on their technical features.

Eye-attached tracking techniques are one of the most precise, but at the same time the most intrusive and uncomfortable for the participants. This technique does not provide the point of regard (area the participant is viewing) of measurement. A search coil is an example of such a device (Figure 2.3).

Electric potential measurement technique called Electro-OculoGraphy (EOG) relies on the recording the differences of the electric potential on the skin around the eye socket.

Optical technique is based on the measurement of the corneal reflection and is the most common method used in the eye tracking research today. Opposite to the techniques mentioned before, the video eye trackers provide the point of regard measurement computed in real-time. The tool can be either mounted on the table, screen or the participant's head. The tracker measures the reflection from the cornea (called *Purkinje reflection* or *Purkinje image*) in relation to the location of the center of the pupil. Tobii eye tracker is an example of a device mounted on a computer screen (Figure 2.5)

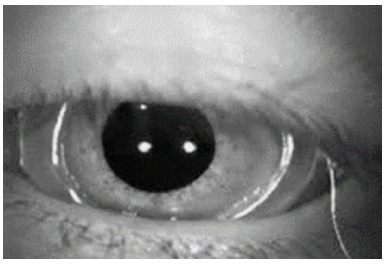


Figure 2.3: *Search coil*, a type of an eye-attached eye tracker
Figure taken from Fig.8.6 in Ref. [21].

Currently many different variants of the optical video-based eye trackers are the most common in the research. The video trackers can track either one or both eyes. They can be either stationary or mobile: the stationary trackers require the participant to remain still while the mobile systems consist of a monitor and a wearable head device and allow the participant to move freely.

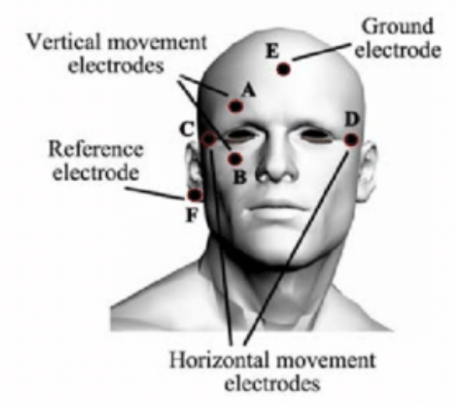


Figure 2.4: Example placement of the electrodes in an experiment using EOG. Figure taken from Fig.1 in Ref. [22].

.

Figure 2.5: Tobii eye tracker, attachable to the screen. Figure taken from https://m.media-amazon.com/images/.

2.4 Main theoretical background for time series analysis

Modern eye trackers can measure and record not only the eye movements (saccades and fixations), but also pupil dilation, point of gaze, and blinking. The observations are recorded sequentially in time at regular intervals and represented by numbers. A collection of observations taken sequentially in time is called a time series [23] and eye tracker data can be analysed as time series data. Time series analysis is a set of methods used to extract meaningful statistical information and characteristics of the time series data. Basic sample statistics can be used in a descriptive statistical analysis of the time series data:

Mean also called an arithmetical average, is a sum of all the values in the analyzed sample divided by the number of observations (Figure 2.6). It represents the typical value of the sample but not always since the mean is very easily affected by the outliers (observation with abnormally high or low value compared to the other values).

Variance is the average of the squared differences from the mean. It gives

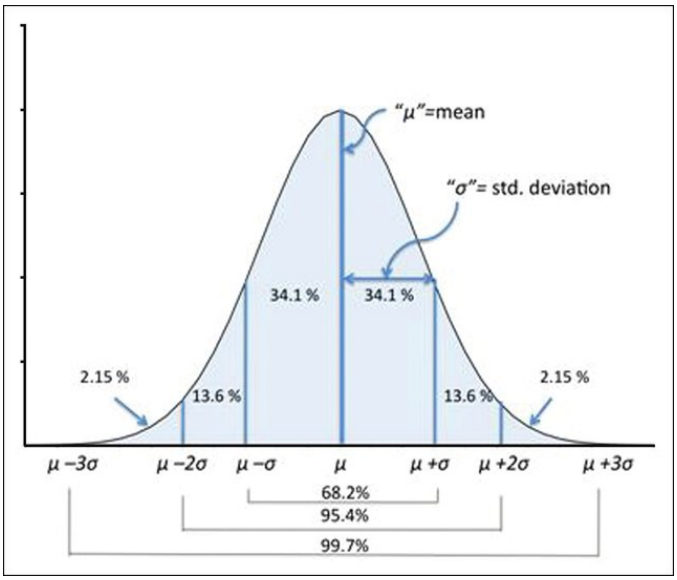


Figure 2.6: Normal distribution with mean and standard deviation. Figure taken from Fig.1 in Ref. [24].

the information about the degree to which each point differs from the mean.

Standard deviation is a square root of the variance. It reflects how spread the numbers in the sample are (see Figure 2.6).

In addition to the statistical tools, time series data has characteristics which can be identified and described by the analysis of the data:

Trend is a measurement which tells if the values of the observation increases or decreases over time. Trends can be local or global, linear or non-linear. Global linear trends can be used for forecasting (Figure 2.7).

Seasonality is a regularly repeating pattern related to the calendar time periods (seasons of the year, quarters, months, days of the week) (Figure 2.7).

Frequency is the number of observations before the seasonal pattern repeats.

Stationarity gives the information if the statistical properties of the data change over time. If a time series data set is stationary it means the mean and variance of the data set are constant. If they are not constant, the data is not stationary.

Outlires are data points which do not follow the trend and has a much lower or higher value than the rest of the observations (Figure 2.8).

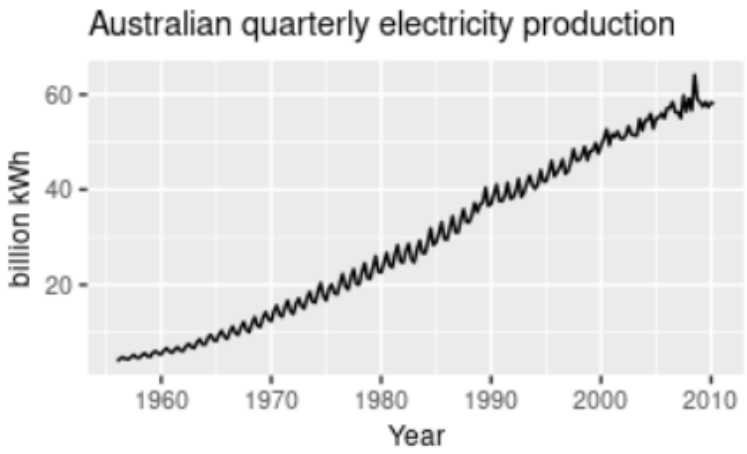


Figure 2.7: Example plot of non-stationary strongly seasonal data with an increasing trend. Figure taken from Fig.2.3 in https://otexts.com/.

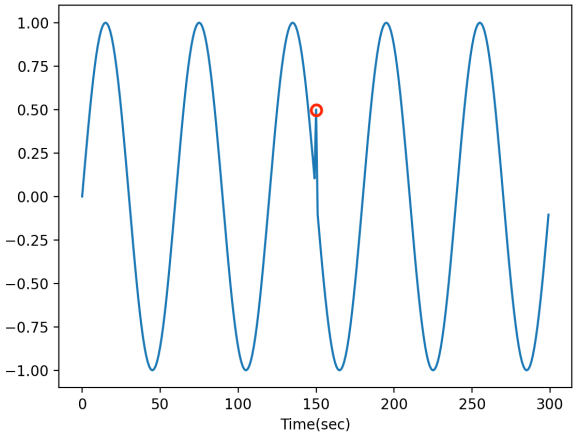


Figure 2.8: Example of an outlier in time-series data. Figure taken from https://towardsdatascience.com/.

2.5 Statistical hypothesis testing

The results of any data analysis are based on the observations from the sample. The sample statistics are characteristics of the sample, not the entire population. The sample, as a part of the population, is similar to it, but is not exactly the same. Sampling error is the difference between the sample statistics and the population parameter. In order to make valid conclusions regarding the population parameter based on the sample statistics it is necessary to use inferential statistics techniques. These techniques check the probability the results are random and ensure unbiased results. Hypothesis testing is one of the techniques that helps to exclude the sampling error from the results. Its main function is to establish if a statement about a population is correct or not. The judgement is based on the sample data and gives the probability the statement is right or wrong

[25, 26].

Hypothesis testing assesses the effect of the topic of the research. In hypothesis testing proposed by Neyman and Pearson [27] two clear, mutually exclusive hypotheses about the studied population are considered: *null hypothesis* and *alternative hypothesis.*

Null hypothesis (H_0) is the first hypothesis and usually states that there is no effect. It can be thought of as a default theory requiring a strong evidence to be rejected. For example, in an experiment with a test group and control group, the null hypothesis cannot be rejected if the results from the both groups are very similar.

Test hypothesis (H_1), also called *alternative hypothesis* or *research hypothesis*, is the alternative hypothesis for the null hypothesis that the research will test. It means that if the null hypothesis is rejected, the test hypothesis must be accepted. Usually the test hypothesis states that there is no-zero effect. In case of comparing the results form the test and control group, the test hypothesis can be accepted if the difference between the results from the groups is different than zero.

The process of accepting or rejecting the research hypothesis is based on the *decision rule*. There are two possible types of mistakes that can be made during the process:

1. *Type I error* (α-error) occurs when the null hypothesis is falsely rejected (null hypothesis rejected while it is true).

2. *Type II error* (β-error) occurs when the null hypothesis is falsely accepted (null hypothesis accepted while it is false and the research hypothesis is true).

Effect size is the difference between the sample value and null hypothesis value. For instance, for the control and test group the effect size could be the difference of the mean value for each of the groups. The effect size is given by:

$$d = \frac{|\mu_{\mathsf{t}} - \mu_{\mathsf{c}}|}{\tilde{\sigma}}\,, \tag{2.1}$$

where μ_{test} and μ_{control} are the average values (scores) of the test and control groups respectively and $\tilde{\sigma}$ is the pooled standard deviation given by

$$\tilde{\sigma}^2 = \frac{(N_t - 1)\sigma_{\mathsf{t}}^2 + (N_c - 1)\sigma_{\mathsf{c}}^2}{N_t + N_c - 2}\,, \tag{2.2}$$

where σ_{t} and σ_{c} are the standard deviations of the scores of the test and control groups respectively.

Significance level represented by the symbol α is the probability of *type I error*.

Power equals 1-β where β is the probability of the *type II error*.

Hypothesis with one parameter of the population is called a simple hypothesis. If the hypothesis contains multiple parameters, it is called a composite hypothesis.

2.6 Use of eye tracking in research

As described earlier in Chapter 1, the beginning of eye tracking research dates back to the end of the 19th century. The development of new technologies utilized in eye tracking devices made them less intrusive and more comfortable for the participants. At the same time, the growing number of the producers makes the eye tracking devices easily available, also for the research purposes. The application of eye tracking research is characterized by a wide diversity. There are many examples of use of the eye tracker in neuroscience, psychology, industrial engineering, advertising and computer (see Figure 2.9).

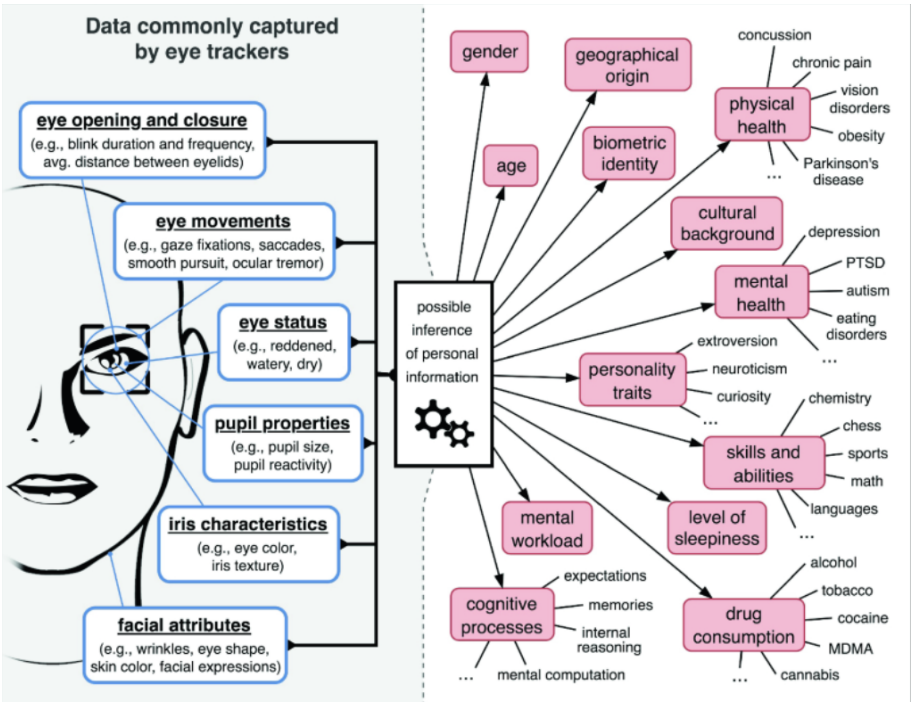


Figure 2.9: Research areas where eye tracking is applied. Figure taken from Fig.1 in Ref. [28].

The most relevant fields of eye tracking related research in the context of this thesis are attention and empathy.

2.6.1 Eye tracker data in attention research

There is a well-known sentence stating that *"everyone knows what attention is"* [29]. Many years later a complete complete opposite point of view is presented in another research [14]. Regardless the disagreement on the definition, attention is one of the most studied topics in the context of

eye tracker data. It is possible to distinguish multiple types of attention: *Focused, sustained, selective, divided* and *switched* are only some of the defined attention types appearing in various taxonomies [30, 31]. In eye tracking research, it is common to record the visual focus and use it for estimation of the attention level. Thanks to their versatility, eye tracking devices can be applied in various surroundings and contexts, making possible participant's natural interaction with the stimuli. Both eye movements [31, 32, 33] and changes in the pupil size [34] can be used as measurements of the participant's attention. Attention is closely connected to memory load where the fixation duration and pupillometry can also be used for estimation of the attention level [7]. Studies of attention are especially important in the research focused on the Attention-Deficit/Hyperactivity Disorder (ADHD). Eye tracking has been used for emotion recognition studies among children with ADHD where both the eye movements [35] and pupillometry [36] are used for assessment attention level. Saccades and fixations are used in the research as measurements of attention.

In their research on visual attention to environmental print among young children, Neumann et al [37] used number of fixations and fixation duration as a measurement of attention to the fixation area. The research focused on determining what preschool children (3-5 years) look at in an environmental print (signs, labels, and logos). The main idea was to find out if the contextual cues can draw children's attention away from the print. The experiment was supposed to estimate the value of the environmental print in early reading development. The researchers focused on three questions:

- How much the prereading children fixate on the words in environmental print?
- Do the fixations vary depending on the type of the print?
- Do the fixations depend on the letter and word knowledge?

Eye tracker was used to record the eye movements while the participants were looking at the photographs of the well-known logos, both with and without the environmental context. In the analysis of the data, the researchers used two types of measurements for (i) attention to any area of the presented item and (ii) attention to the words only. To measure the general attention (i) the following measurements were used:

- total number of fixations on the item was measured (sum of individual fixations on the item)
- total fixation duration on the item (sum of individual fixation duration on the item)

To measure the attention to the words (ii) different measures were used:

- time to first fixation (time from the presentation to the first fixation to the word)

- percentage of total fixations (proportion of total fixations on the word to the total fixations on the item)
- percentage of total fixation durations (proportion of total duration fixations on the word to the total duration of fixations on the item)

The results of the experiment showed that the prereading children pay attention to the words embedded in the environmental print and the prints with larger, centrally placed words may facilitate the print learning.

In another research, Türkan et al [38] focused on change detection performance by comparing the fixation duration among the participants with ADHD and the control group with typical development. The measurements used in the research were: fixation latency, fixation duration and fixation count. The purpose of the study was to find out if the children (aged 8-11 years) with ADHD have more difficulties in detecting changes due to their voluntary eye movement control and attention shifting mechanisms. The results from the group with ADHD were compared with the results from the control group of typically developing children. The researchers attempted to answer the following questions:

- Do the children with ADHD perform worse in the detection task than the children with typical development?
- Does the performance depend on the type of change in the task?
- Can the differences in the performance be explained by the different scanning strategies (eye movement patterns)?

The eye tracker recorded the eye movements of the participants while they were doing the change detection task. In the task, participants looked at the pairs of photographs where one photograph of each pair was slightly changed (color, location, presence/absence change). In the analysis of the data, the following measures of interest were used: (i) accuracy percentage of the change detection task, (ii) average detection time of the change detection task and (iii) eye movement measure during the change detection task. The last measure is based on the eye tracker data and consists of the following measurements:

- fixation latency
- duration of the first fixation
- fixation count on the changed area
- fixation duration on the changed area

The results of the experiment show the participants with ADHD had lower change detection performance than the participants without any diagnosis. The participants with typical development made longer fixations on the changed area and their first fixation was also longer than among the children with ADHD.

Research by Smallwood et al [39] has focus on pupil diameter and its changes in response to an external stimuli. The study test the pupil diameter changes in the context of decoupling hypothesis. The decoupling hypothesis states that the mind can reversibly decouple attention from sensory information. According to this hypothesis, mental activity can be divided in online processing of sensory information and offline mode where the cognition which depends on the contents of memory, not the actual stimuli. This division of thought processing allows to create internal trains of thought without being prevented by external events. In their study Smallwood et al address the question whether the changes of the pupil diameter are consistent with the decoupling hypothesis. They assume the pupil diameter can be used as a measure to assess cognitive activity during decoupled thought.They based their assumption on the following findings from the previous research:

- Pupil diameter changes in response to an external stimuli [40]
- Pupil diameter increases during long term memory retrieval [41]
- Pupil diameter is linked to known control processes in the brain [42]

In their research, Smallwood et al examined if the changes of pupil diameter manifest the two different modes of activity defined by decoupling hypothesis (online and offline). They defined following 5 predictions:

- Pupil diameter will increase as events in the task are encoded.
- Pupil diameter will not increase when events in the task are presented.
- Pupil diameter will show high baseline activity which is uncoupled from the task event.
- High baseline pupil diameter prior to probes will be indicative of slow correct responses and/or failure to encode task events.
- Baseline pupil diameter will show a nonlinear or step-wise relationship to continuous measures of external attention.

The team developed 2 tasks to test the defined predictions: a Working Memory task and a Choice Reaction task. The result showed the pupil diameter increased indicating the processing of task stimuli in the online cognition. During the periods characterized by the offline thought activity pupil diameter either not changed or showed abnormal changes. The results of the experiment confirmed the decoupling hypothesis suggesting that the capacity of spontaneous cognitive activity depends on minimizing the disruptive stimuli from the external world.

Kang et al [43] investigate pupil dilation as a consequence of attentional effort. In their research, the team investigated further the results from the research by Smallwood [39] and focused on the question what the pupillary changes can tell about the real-time human information processing. The experiment consisted of two types of tasks:

- Choice Reaction Time (CR) tasks and
- Working Memory (WM) tasks

During the experiment, pupil diameter was collected from 22 participants by eye tracking device. The data from 16 participants passed the quality control cut-offs and was analyzed further. The luminescent conditions remained constant during the tasks. According to the researchers, the results prove the pupillometry can be used to index moment-by-moment fluctuations of attention to changes in information.

Based on the findings from the mentioned studies it is possible to conclude that (i) the increasing number of saccades correlates with decreasing level of attention, (ii) the longer fixation periods and pupil dilation correlate with the higher level of attention.

2.6.2 Eye tracker data in empathy research

In psychology, the eye tracker data contributes in the studies focusing on the emotion recognition and visual perception. The idea of using the eye tracker to detect the participant's emotions is relatively new and there is not much research in this field yet [8]. Pupillometry plays a vital role in assessing the emotions: pupil diameter is larger when the participant is experiencing positive emotions. At the same time, there are many factors which can affect the pupillary reaction at the same time and it can be challenging to determine what caused the dilation in a specific situation. The eye movements and their speed have also been used in emotion recognition [44]. Pupil position [45], fixation duration [46] and distance between sclera and iris [47] have also been used in the research focusing on the emotion recognition.

Ping Liu et al [48] studied the impact of the empathy trait on the processing of the emotional information. In their research, they investigated the intentional bias to happy faces among high- and low-empathy participants. The attention level was assessed by using the Interpersonal Reactivity Index (Chinese version). The attention was measured by using the dot-probe task and eye tracker. The following measures were used:

- reaction time (judging the position of the probing dot)
- time to the first fixation (recorded by the eye tracker)
- total fixation duration on the area of interest

The participants were presented a selection of positive emotional words and neutral words. They were also presented a selection of happy and neutral faces and the reaction times were measured by the eye tracker. Later, the results were combined with the assessed empathy levels. Based on the eye tracker data, the high-empathy participants had longer total fixation duration on the happy faces. This indicates that the higher attention was positively correlated with the empathy level.

A different approach to estimate empathy was presented by Siri Leknes et al [49]. In their study, the team investigated the impact of the hormone *oxytocin* on the emotional processing. Specifically, the study focused on investigations of the effect of oxytocin on the evaluation of happy and angry facial expressions, and relation of the results to the participants sensitivity to others' emotions. Each participant completed the task twice: once in a session with placebo and once after the application of intranasal oxytocin. The pupil diameter of the participants was recorded while a set of faces with different emotional expressions was presented to them on a computer screen. The background section of each image was adjusted so the change of the luminance would not affect the pupil size. During the task, the following measurements were recorded:

- perceived mood of the presented face (behavioral measurement)
- mood of the participant (behavioral measurement)
- diameter of the participant's left eye pupil (pupillometry)

The facial expressions were divided into categories (happiness, friendliness, attractiveness, sadness) and each of the categories was represented by implicit and explicit version. The study investigated if there is a correlation between the sensitivity to other's emotions, the use of the oxytocin and the pupil diameter. The results showed that the oxytocin had higher enhancement on pupil dilation among the participants with the higher sensitivity to other's emotions.

Chapter 3

Data and methodology

3.1 The experiment by Bhurtel et al: design, protocol and ethical approval

As mentioned in the Chapter 1, this thesis analysis data collected for another empathy-related study by S. Bhurtel et al [1]. In their research, Bhurtel et al proposed an experimental protocol to promote empathy towards non-verbal people. The protocol is based on the training of communication with an eye tracking device. The research promotes the idea of increasing the empathy among the participants by asking them to use the same eye tracker based setup as used by people with lock-in syndrome and non-verbal people with severe form of paralysis (e.g. quadriplegia).

The aim of the study was to increase the level of empathy among healthy participants towards people not being able to communicate verbally or with gestures.

The experiment measured the empathy level among the participant before and after using the eye tracker setup for communication. A total of 44 participants were divided in control and test of 22 persons each. The participants were master and bachelor students in the age between 20 and 40. The participants were informed about the procedure and purpose of the study. Each participant had to sign a consent and fill out a questionnaire regarding the demographic data. The personal data was collected and processed according to the requirements of the Norwegian Centre for Research Data (NSD), approved by this entity with an application reference number 119986. In the experiment, the following methods were used to collect the data:

- eye tracking device Tobii Pro X3-120
- eye tracker based communication interface E-tran
- questionnaire of Cognitive and Affective Empathy (QCAE) [50]

The Tobii eye tracking device was used together with the communication interface. In E-tran, the user can choose letters and symbols on digital boards shown on a screen. The selection is made by the eye movements. In

the tasks, the participants were asked to create 8 questions using the e-tran interface.

The questionnaire contained 31 standard QCAE statements and 9 statements added for this particular experiment. Participants were asked to answer how much they agree or disagree with the presented statement on the scale from 1 to 4. The sum of the points would give the empathy score of the participant where the higher result indicates higher empathy level. For each of the 44 participants, two empathy scores were registered represented by an integer number: one score summing up the points from the 31 standard statements and one score summing up the points from all the 40 statements (31 standard and 9 added). For each participant 4 scores were registered: standard before and after (31 statements) and extended before and after (40 statements). In the further analysis the extended results were taken into consideration. All the questionnaire questions are included in the Appendix C.

In the experiment, each participant was asked to:

1. Assess empathy before using the interface (answer the QCAE)
2. Complete the experimental task:
   2.1. Complete the tasks with the eye tracker interface (test group)
   2.2. Look at a selection of abstract images (control group)
3. Assess empathy after using the interface (answer the QCAE)

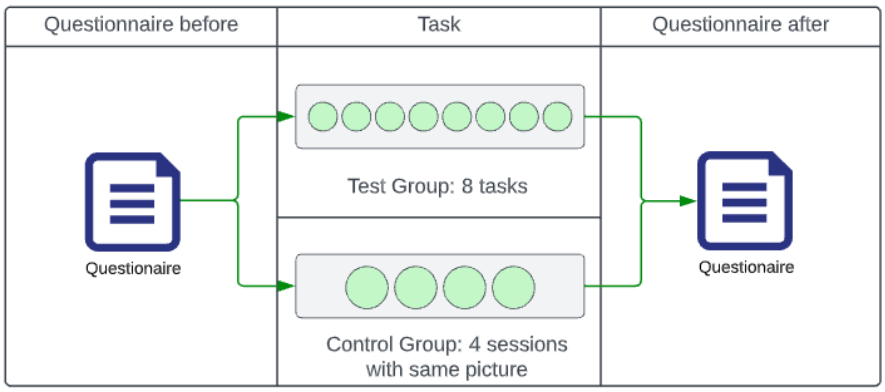


Figure 3.1: Design of the experiment.

The level of empathy measured by the questionnaire before and after using the interface should show if the empathy among the participants increased after using the experimental framework. The tested hypothesis assumed the effect should be seen after the experiment. It means the level of empathy among the participants in the test group was expected to increase after answering the E-tran tasks. The empathy in the test group has increased, however, not as much as expected. Among some participants the empathy score has decreased after the intervention. In the control group

the empathy decreased after the experiment. The analysis of the results revealed the following:

- Average time used on the experimental task: 25 minutes.
- Positive correlation between the time spent and empathy score (difference between empathy measured after and before) for the participants who spent more than 10 minutes on the task. After around 15 minutes the score begins to decrease.
- No difference in time spent on answering the QCAE before and after points towards full spectrum of empathy scores.
- Difference in time spent on answering the QCAE before and after points towards decrease of the empathy.
- Average time used on the control group task: 6 min.
- No difference in the empathy score in the control group.
- Shorter time used on fulfilling the experimental task by the control group participants.

The decrease of empathy in some of the participants is contrary to common intuition and hypothesis testing. To explain the results, the authors discuss the possible flaws in the design of the experiment and the questionnaire used to assess the empathy level. According to the authors, the lack of effect and decreasing empathy score might be caused by the fact the participant got bored or irritated. Most of the participants used less time when answering the QCAE after the practical task and expressed tiredness after fulfilling the experiment. The second QCAE was applied directly after the practical E-tran task - something that could negatively affect the patience and attention of the participants.

After the experiment was accomplished, Bhurtel et al continued to gather the data. The data sets analysed in this thesis contain the answers from 60 participants (Bhurtel's experiment was based on 44 result sets.). The further analysis focuses on the data registered for the test group. The data set from the test group consists of 30 sets of: - questionnaire-based empathy scores before and after the experiment tasks and - eye tracker data registered during each of the 8 test tasks.

3.2 Data sets to be analyzed

The eye tracker recorded the eye movements and changes of the pupil size of each participant during the experimental task. Such data sets were not analyzed by Bhurtel et al. According to the research presented in chapter 2, it is possible to asses attention based on eye movements and pupil diameter registered by eye tracker device. Following the authors suggestion regarding the lack of attention causing the low empathy score,

we will investigate if there is a correlation between pupil dilation measured by the eye tracker and the decreasing empathy levels.

The eye-movement data and pupil size changes were registered by Tobii Pro X3-120 eye tracking device. The eye tracker register the coordinates of the eyes with the sampling rate of 120 Hz. Pupil diameter is registered with a lower rate of 40 Hz [51]. The data is saved as time series, with the timestamp of each registration. The following properties relevant for the further analysis of the eye are registered by the eye tracking device during the task:

Recording timestamp (data type: integer) shows the number of elapsed seconds since the previous event registration. Starts on 0 in each user-file.

Participant name (data type: string): registered name of the participant, saved in format "Participant000x" where x is the number 1-60.

Event (data type: enumeration, nullable): registered event or null when no event registered. Gives information about changes during the recording. Possible values: *RecordingStart, RecordingEnd, Eye tracker Calibration start, Eye tracker Calibration end, ImageStimulusStart, ImageStimulusEnd, MouseEvent*.

Pupil diameter left (data type: float): registered diameter of the left pupil.

Pupil diameter right (data type: float): registered diameter of the right pupil.

Eye movement type (data type: enumeration, nullable): type of the eye movement registered at the time. Possible values: *Fixation, Saccade, Unclassified, EyesNotFound, null*.

Data for all the 8 tasks were saved in the one file per participant. Column "Event" contains the information about the start and the end of the measurement (values: *ImageStimulusStart* for start and *ImageStimulusEnd* for end of the registration of the task).

For the code used to analyze the data please see the GitHub reference link in Appendix D.

3.3 Pre-processing of the data sets

Event type: *null* and *EyesNotFound*

Event types registered by the eye tracker are: *Fixation, Saccade, Unclassified, EyesNotFound, null*. The event types significant for the analysis are *Fixation* and *Saccade*. All the other events can be classified as one type. Data in the Event column was transformed so that all the null-values and *EyesNotFound* where changed to the event type *Unclassified*.

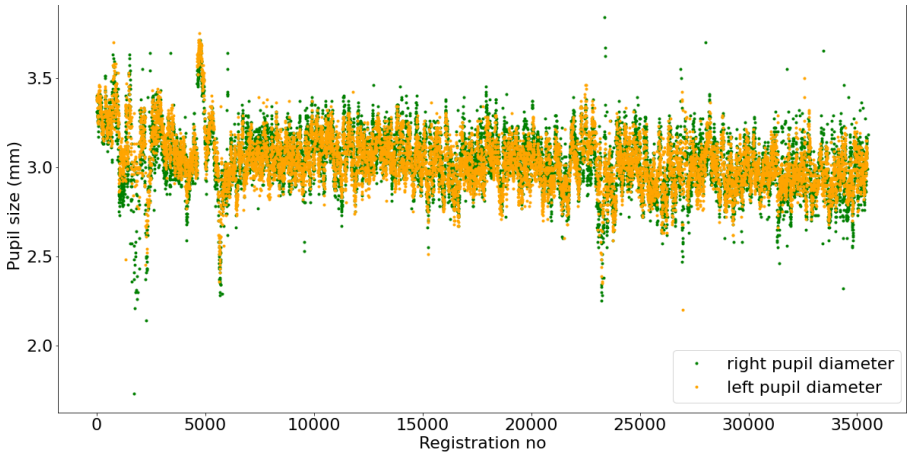


Figure 3.2: Left and right pupil size changes for one participant during one attempt.

Participant's number of attempts

The experiment in the test group consisted of 8 consecutive tasks. Of the 30 participants in the test group, 24 participants had registered 8 time-series (one for each experiment task). These were considered in the further analysis. Participants number 3, 5, 7, 9, 11 and 33 were rejected due to too many registered attempts (respectively: 16, 24, 32, 40, 49 and 9 attempts). In the control group the control task consisted of 4 attempts. Only the participants with 4 registered attempts were taken into consideration. Participants number 2, 4, 6, 8 and 32 were rejected due to too few or too many registered attempts (respectively: 3, 5, 9, 13 and 17 attempts).

Average diameter from both eyes

Due to the sampling frequency, 2/3 of the data points for eye pupil registration are missing for both pupils. In addition, a smaller number of the data points is missing for either left or right pupil due to other unknown variables.

Human eye pupils are generally equal in size and any changes of the pupil size are consensual [20]. Based on this assumption it should be possible to continue the analysis of data with mean value of the both pupils, instead of analyzing left and right pupil separately. For the data points where value for one of the pupils is missing, the size of the other pupil could be used as average. Figure 3.2 shows a typical example of the changes of one participant pupil size during one of the experiment tasks. The values registered for left and right eye pupil are mostly overlapping on the graph which indicates they are quantitatively similar.

A scatter plot of the left and right pupil is visualized in the Figure 3.3. The correlation coefficient between the left and right pupil size for the measurement visualized in Figure 3.3 equals 0.69, and is calculated from the formula:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}.} \tag{3.1}$$

Correlation coefficient for grouped event types is as follows: a) fixations: 0.74 b) saccades: 0.52 c) unclassified: 0.61. Figure 3.4 shows scatter plots of the left and right pupil size grouped by the event type label. Correlation coefficient can have a value between -1 and 1, where values closest to 1 (resp. -1) mean strong positive (resp. negative) correlation. Correlation coefficient with value of 0.69 can be interpreted as a significant positive correlation. Based on this finding, in the further data analysis the mean value from both pupils was used, and if the value from one eye was missing, the existing value was used as mean.

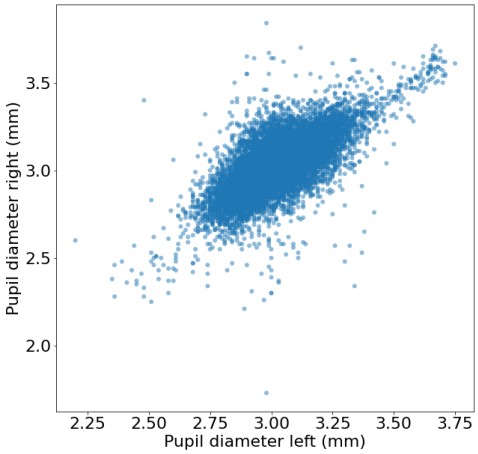
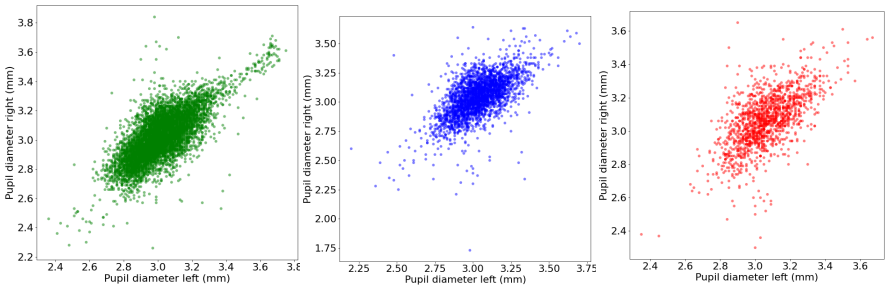


Figure 3.3: Scatter plot of the left and right pupil diameter for one participant one during attempt.

After the pre-processing the data registered for each participant form the test group consists of eight time series with the following information registered over time:

- average pupil size in millimeter
- mutually exclusive registered eye event type: saccacde (label = 3), fixation (label = 2) or unclassified (label = 1)

The typical attempt data is visualised in Figure 3.5. This data was the basis for the further analysis presented in the following sections.



(a) Fixations (b) Saccades (c) Unclassified

Figure 3.4: Scatter plot of the left and right pupil diameter for one participant one during attempt. Grouped by event type.

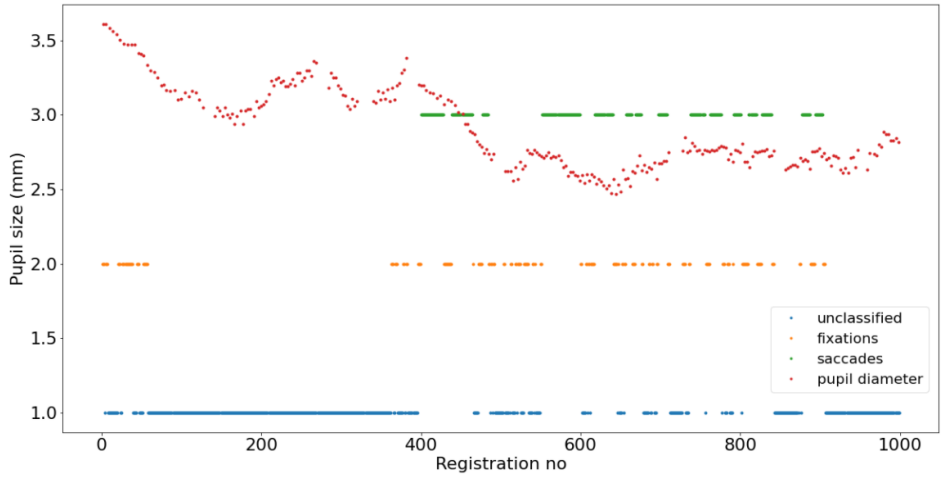


Figure 3.5: Illustration of a slice of the series for each attempt of each participant, including the label (fixation, saccade, unclassified) and mean of pupils diameter.

3.4 Methodology and analysis framework

Our analysis comprehends the following steps:

- Each participant has 8 registered attempts, with a pupil diameter in each attempt registered in time t.
- From each attempt data it is possible to derive the number and further fraction of the events of the type *saccade* f_S and *fixation* f_F, participant average pupil size $D(t)$ and uncertainty given by standard deviation of the pupil size in time.
- The empathy score from before and after the intervention are also was also registered for each participant. The scores were added to the table, along with the empathy variations (empathy after - empathy before).

The final result matrix is visualised in the Table 3.1.

participant_no	attempt_no	rmean	std_deviation	empathy_before	empathy_after	empathy_diff	fraction_fix	fraction_sacc	total_mean	total_std_dev
1	1	3.0175	0.1318	134.0	137	3.0	0.5905	0.2544	3.0241	0.0070
1	2	3.0180	0.1351	134.0	137	3.0	0.5916	0.2568	3.0241	0.0070
1	3	3.0347	0.1552	134.0	137	3.0	0.5871	0.2574	3.0241	0.0070
1	4	3.0270	0.1472	134.0	137	3.0	0.5795	0.2583	3.0241	0.0070
1	5	3.0290	0.1479	134.0	137	3.0	0.5874	0.2586	3.0241	0.0070
1	6	3.0142	0.1360	134.0	137	3.0	0.5802	0.2560	3.0241	0.0070
1	7	3.0237	0.1420	134.0	137	3.0	0.5820	0.2575	3.0241	0.0070
1	8	3.0290	0.1471	134.0	137	3.0	0.5854	0.2600	3.0241	0.0070
13	1	2.8579	0.2368	105.0	117	12.0	0.5263	0.1272	2.7933	0.0369
13	2	2.8157	0.2126	105.0	117	12.0	0.5267	0.1293	2.7933	0.0369
13	3	2.7647	0.2000	105.0	117	12.0	0.5091	0.1342	2.7933	0.0369
13	4	2.7719	0.2283	105.0	117	12.0	0.4720	0.1315	2.7933	0.0369
13	5	2.7624	0.2169	105.0	117	12.0	0.4842	0.1337	2.7933	0.0369
13	6	2.7617	0.2148	105.0	117	12.0	0.4821	0.1340	2.7933	0.0369
13	7	2.7804	0.2280	105.0	117	12.0	0.4829	0.1296	2.7933	0.0369
13	8	2.8320	0.2277	105.0	117	12.0	0.5102	0.1271	2.7933	0.0369
15	1	2.9626	0.2917	122.0	137	15.0	0.4367	0.1533	2.9543	0.0392
15	2	2.9558	0.2902	122.0	137	15.0	0.4281	0.1530	2.9543	0.0392
15	3	2.9883	0.3031	122.0	137	15.0	0.4101	0.1459	2.9543	0.0392
15	4	2.9836	0.2803	122.0	137	15.0	0.3672	0.1490	2.9543	0.0392
15	5	2.9162	0.2782	122.0	137	15.0	0.4147	0.1566	2.9543	0.0392
15	6	2.8981	0.2540	122.0	137	15.0	0.4142	0.1568	2.9543	0.0392
15	7	2.9209	0.2838	122.0	137	15.0	0.4107	0.1567	2.9543	0.0392
15	8	3.0089	0.3183	122.0	137	15.0	0.4163	0.1446	2.9543	0.0392
...	...	...	...	...	...	...	...	...	...	...

Table 3.1: Final result matrix used in the analysis.

The matrix was used in the following chapter to investigate the possible correlations between the empathy scores and participants' pupil diameter. The significance of the results was tested in power analysis.

Chapter 4

From pupil size to empathy scores

It is know from the previous research that the pupil diameter size is positively correlated to attention [39, 40, 41, 42]. This chapter focuses on investigating a possible correlation between the pupil diameter size and empathy scores gathered by Bhurtel et al in their experiment [1]. If empathy is also positively correlated with pupil diameter size, it would be possible to use it as a new measurement for empathy level assessment. In further research, it could be also possible to investigate if the attention and empathy are correlated to each other. Current methods used for assessing the empathy level are not objective - they are based on a relative opinion and answers of the subject. The pupil diameter could give an insight in the empathy level based on a reaction not possible to control by the subject, and by the same more reliable.

4.1 Typical diameters and uncertainties for each participant

Pupil diameter was registered for each participant in each of the 8 time series representing the intervention tasks. Standard deviation of the pupil compared to the mean is small - the values of the pupil size in each attempt are closely clustered to the mean (see Figure 4.1). For all the participants see Appendix A. Consequently, we can use one single value per each participant, averaging over 8 attempts.

The correlation between two variables is represented on a scatter plot and by calculating and interpreting the correlation coefficient. Figures 4.2 show the scatter plot between the average pupil diameter and empathy scores: a) before the intervention, b) after the intervention, c) difference in measured empathy between after and before, and d) average empathy score in the test group.

There is no clear, visible pattern in the presented Figure 4.2. For clarity, the same figures are plotted for the control group (Figure 4.3).

Based on the diagrams, there is no correlation between the empathy

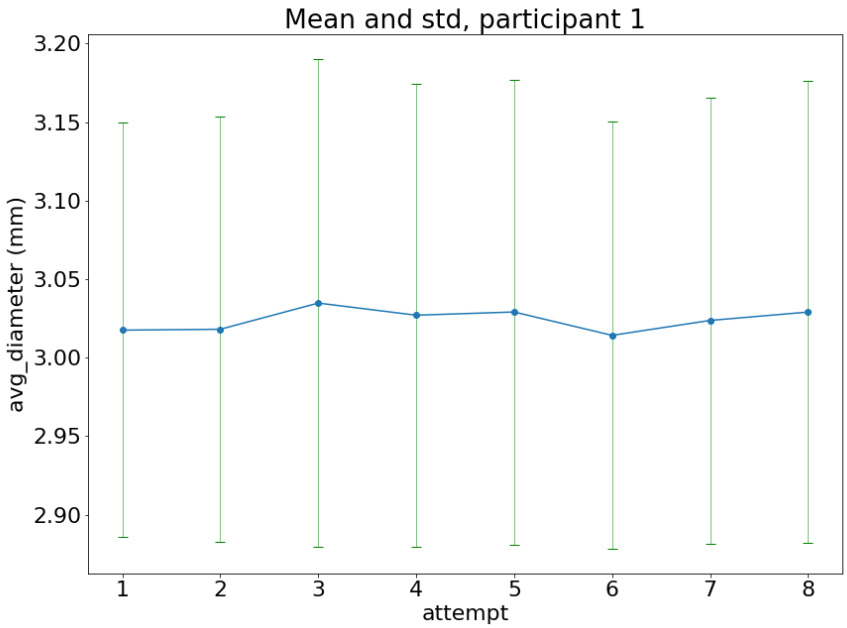


Figure 4.1: Average pupil size with standard deviation for participant 1. For all the results see Appendix A and B

scores and registered pupil diameter. The corresponding correlation variables are presented in Table 4.1. Based on values of the calculated correlation variables it is not possible to prove an existence of a relationship between the empathy scores and pupil diameter size.

	Correlation coefficient			
	Avg pupil diameter vs Empathy score before	Avg pupil diameter vs Empathy score after	Avg pupil diameter vs Empathy score difference (after - before)	Avg pupil diameter vs Avg empathy score
Test group	-0.04010371	-0.09531171	-0.04778947	-0.08030432
Control group	0.0600348	-0.10974145	-0.12247727	-0.06459935

Table 4.1: Correlation coefficient for pupil diameter and empathy scores in test and control group.

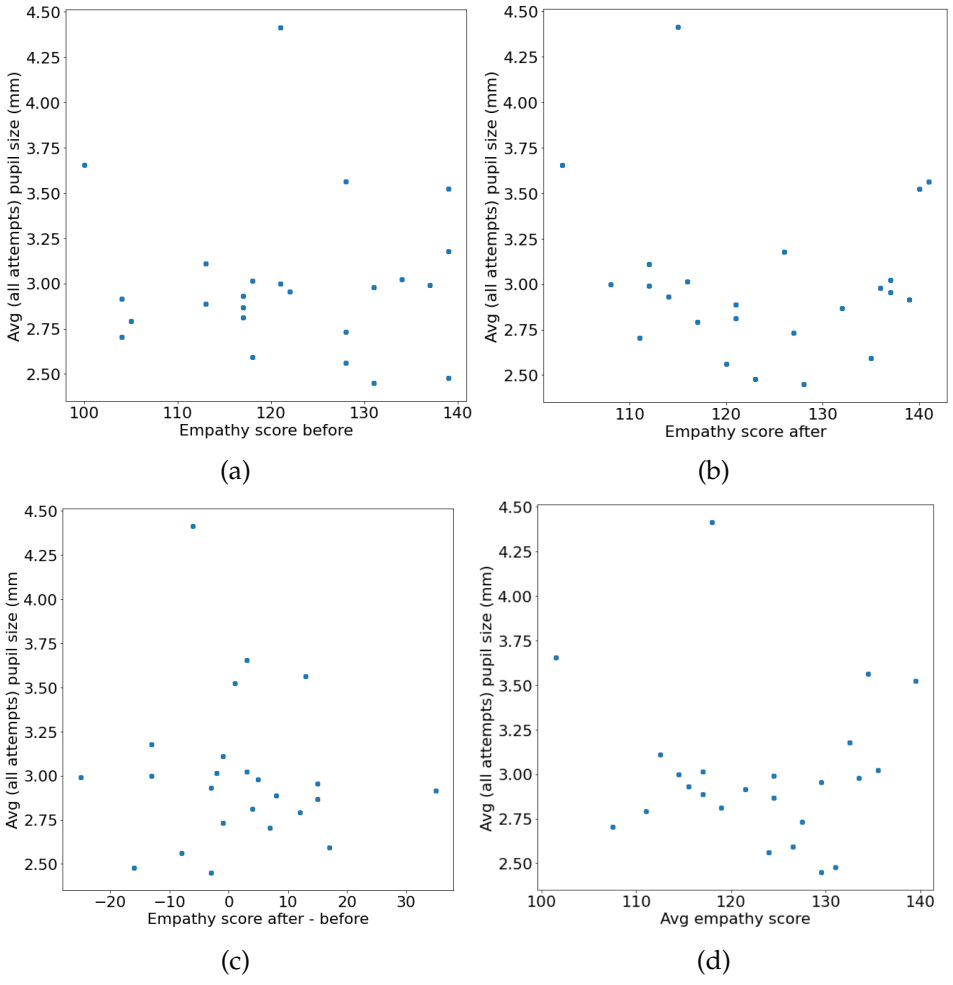
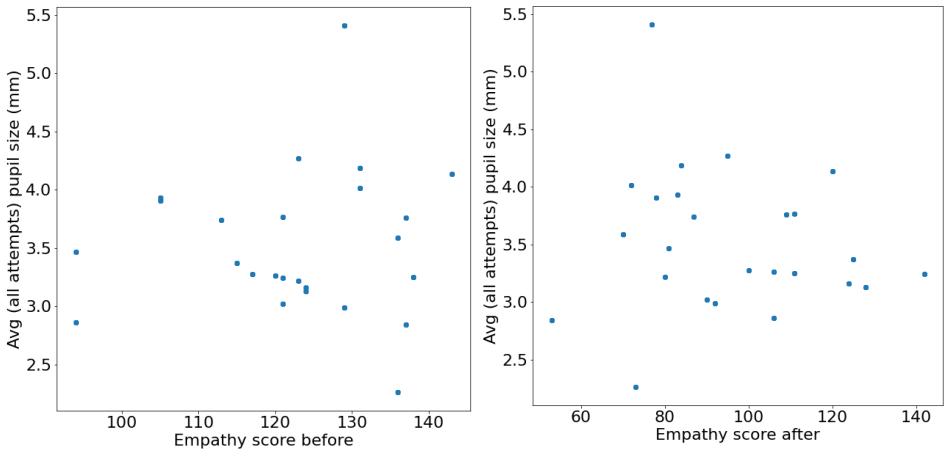


Figure 4.2: Test group: scatter plot of the average pupil size vs empathy variations: a) empathy before, b) empathy after and c) difference empathy after-before before and after the intervention.



(a) (b) (c) (d)

Figure 4.3: Control group: scatter plot of the average pupil size vs empathy variations: a) empathy before, b) empathy after c) difference empathy after-before and d) average empathy group before and after the intervention.

4.2 Power analysis and hypothesis testing

Analysis of the data conducted so far has not confirmed any significant correlation between the empathy scores and the registered mean of the pupil size. Preferably, power analysis is conducted before the experiment to estimate the desired sample size. In our case the experiment sample size was already established by Bhurtel et al, but it is still possible to calculate the statistical power for the given sample and based on the result decide if the null hypothesis can be rejected, with the chosen sample size. In their experiment, Bhurtel et al focused on the empathy score results when calculation the effect size and optimal sample size. In this thesis we will focus on the average pupil diameter and its difference between test and control group. Power analysis depends on four variables:

- Effect size d
- Sample size N
- Level of significance α
- Power 1 - β

The first variable, effect size, can tell us how meaningful the difference between the results from the test group and control group are. There are known multiple measures of effect size. Cohen's d is one of them and is defined by:

$$d = \frac{|\mu_{\mathrm{t}} - \mu_{\mathrm{c}}|}{\tilde{\sigma}}, \tag{4.1}$$

where μ_{t} and μ_{c} are the average values (scores) of the test and control groups respectively and $\tilde{\sigma}$ is the pooled standard deviation given by

$$\tilde{\sigma}^2 = \frac{(N_t - 1)\sigma_{\mathrm{t}}^2 + (N_c - 1)\sigma_{\mathrm{c}}^2}{N_t + N_c - 2}. \tag{4.2}$$

	Avg pupil diameter (μ)	Std deviation (σ)	No of samples (N)	Nc, Nt
Test group	3.005	0.423	24	7
Control group	3.523	0.615	25	7

Table 4.2: Mean, standard deviation and number of samples in test and control group.

By substituting the known values in the formula, we can calculate pooled standard deviation and effect size. Table 4.2 shows the values used in calculation of the effect for each of the groups. The calculated pooled standard deviation $\tilde{\sigma}$ equals 0.53. The calculated effect equals 0.98. The calculated results are presented in Table 4.4. According to the Table 4.3, presenting the descriptors for magnitudes of d = 0.01 to 2.0, as initially

suggested by Cohen, the calculated effect size of 0.98 is *large* towards *very large*. A large effect size could indicate statistical significance and that the null hypothesis can be rejected. Before any conclusions are made, power analysis should be conducted.

Effect size	Very small	Small	Medium	Large	Very large	Huge
d	0.01	0.20	0.50	0.80	1.20	2.0

Table 4.3: Cohen's descriptors for magnitude.

Pooled std deviation (σ)	Cohen's d
0.53	0.98

Table 4.4: Calculated pooled standard deviation and effect.

The sample size equal 24 for test group and 25 for control group. We can assume a low probability of Type-I errors with level of significance α = 0.05 which equals 1.645 z-score from a standard normal probability distribution. For β we will use the standard value 0.2 which gives us power = 0.8 corresponding to 0.8416 z-score from a standard normal probability distribution.

	Avg empathy score before intervention	Avg empathy score after intervention	Avg change in empathy score (after-before)	Avg empathy score in the group
Test group	121.83	123.79	1.96	122,81
Control group	122.68	95.88	-26.8	109,28

Table 4.5: Empathy scores before and after the intervention in test and control group.

Our test hypothesis states that empathy is positively correlated with pupil diameter. To support our hypothesis we would expect the pupil diameter μ_t to be bigger in the group with a higher empathy score (test group). To compare the average pupil size in the groups we can test the following hypothesis:

Null Hypothesis H_0: $\mu_t = \mu_c$ (Pupil diameter in test group same as in control group, no difference)

Test hypothesis H_1: $\mu_t \neq \mu_c$ (Pupil diameter in test group different than in control group)

To investigate if the calculated effect d is statistically significant, we can calculate the number of participants necessary for the difference to be statistically significant. This can be calculated from:

$$N_t \sim N_c \sim= \frac{(Z_\alpha + Z_{1-\beta})^2}{D^2} \tag{4.3}$$

We assume that both groups have approximately the same size ($N_t \sim N_c$). The calculation gives us result 6,44 which implies a minimum sample size of 7 participants which is more than 3 times smaller than the sample size used in the experiment.

4.3 Summary

No correlation between pupil size and empathy score was found in the data. The calculated effect size was measured by Cohen's d and equals 0.98 and according to the Cohen's descriptors of magnitude (Table 4.3) can be interpreted as large towards very large. It means there is a significant difference between the measured average pupil diameter in test and control group. The statistical power of 0.8 can be achieved with sample size 7. The results are discussed in chapter 5.

Chapter 5

Conclusions and future directions

In this thesis we addressed the question if empathy and attention in individuals are related with each other, based on analysis of eye tracker data. We started with a hypothesis that empathy is positively correlated with human eye pupil diameter. If the correlation was possible to prove based on the data collected by Bhurtel et al [1], the next step would be to investigate the relationship between empathy and attention. The analysis was based on the following assumptions:

- Pupil diameter is a measure of attention.

In addition, we attempted to answer the following follow-up questions:

- How significant is this relationship?
- Which statistical tools for data analysis can better uncover the relationship between empathy and pupil diameter?
- What can a data scientist learn from this data?

The collected data was structured in form of a matrix (Table 3.1) which contained the information about the time series (attempts) for each participant, variations of empathy scores and average values of pupil diameter during the attempts. The data from the matrix was used in further data analysis and calculation. We tried to find the relationship between the eye pupil diameter and variations of empathy score in test and control group by calculating correlation coefficients and analyzing scatter plots. The calculated correlation coefficients for the pupil size and empathy variations were very low and negative in almost all case with one exception of average empathy score before intervention vs pupil size in control group (Table 4.2). Intuitively, based on our assumption backed up by the previous research that the eye pupil diameter is positively correlated with attention, we would expect a positive correlation between empathy score and pupil size in test group, where empathy scores slightly increased after the intervention. The calculated correlation values are not only very low, but also negative in almost all the cases presented in the Table 4.2.

There is one exception in control group where the coefficient is positive. Comparing pupil diameter between both groups, power analysis proved that the size effect is large and the sample size of 25 is around 3 times higher than the sample size required for the difference to be statistically significant. We can therefore reject the hypothesis that both groups have the same average pupil diameter. However, the results of power analysis should be interpreted together with the registered empathy scores from the both groups (Table 4.5) and average pupil size registered for both groups. We can see the results are contrary to the expected: pupil diameter is higher in control group where the empathy score is lower and has decreased after the intervention. Intuitively, we would expect the pupil diameter to be bigger in the group with a higher empathy score (test group). All the mentioned findings can point towards a bias in the data. Both the design of the experiment, and the technical aspects of the intervention could affect the results. The flaws in the experiment design were already mentioned by Bhurtel et al as a possible explanation of the decrease of the empathy among some participants. Bhurtel et al mentions that the participants, especially in the test group, reported tiredness after performing a task. Most participants used less time to fill the questionnaire in post-experiment than before. The authors suggest impatience as a reason of decrease in the empathy score. It is possible the participants in the test group were more impatient, less focused and the average pupil size reflects the way they felt. The lack of correlation between empathy and pupil diameter in the analyzed data is contrary to the results from other research [52] and can also indicate flaws in the design of the experiment. However, the questions about the quality of the data undermine the credibility of the results. The results presented in this thesis are not sufficient to answer the research question. Instead of rejecting the null hypothesis based on these results, we could assume that the test hypothesis is correct, but the experiment must be done in a different way.

The design of the experiment has two aspects which can be discussed as possible areas for improvements: (i) technical - how the data is collected, and (ii) content - what is supposed to engage the participants emotionally.

The human eye pupil can react to many factors. The design of the eye tracking studio defining the physical conditions during the experiment can influence both the conscious and unconscious reactions of the participants. The light conditions, noise from the outside, possible interruptions can cause a behaviour of the pupil which can be mistakenly interpreted as increase or decrease of attention or empathy [53]. The physical and psychological condition of the participants is also an important factor. Tiredness or anticipation can influence attention and eye's reaction. Some research [52] suggests the pupil size increases more in reaction to a pleasant emotion than an unpleasant one. Maybe the experiment provided an unpleasant situation to some of the participants and led contraction rather than dilation of the pupil. The light conditions during the experiment must be exactly same for all participants. All the external factors, like noise from the outside or possible interruptions should be prevented. Moreover, the participants should be in a similar psycho-physical state during the

intervention and they should not be under influence of any substances which can affect the pupil diameter or pupillary response. As it is known, some medicine can have this kind of side effects [54].

The experiment consisted of 3 stages: questionnaire before the intervention, intervention, questionnaire after the intervention. In this experiment, the eye pupil was only used during the intervention. To collect more data, the eye tracker could be used also during the stage 1 and 3 when participants filled out the questionnaire to be able to compare pupil diameter before, during and after the intervention.

The pupillary data could be analyzed in different way. Zhang et al analyzed the changes of the pupil diameter over time and focused on the peak values [52]. It could be possible to analyze the relation of the peak values of the pupil diameter with the registered event types and their duration to find a different correlations in the data. The possible relation between fixation durations registered during the interventions, and the empathy scores was not investigated in this thesis due to the time constraints of the project. This should be investigated in the future as the results can give insights into the attention level of the participants and contribute to the explanation of the little increase of the empathy scores in the test group.

Among many other definitions, empathy can be explained as a subject having a similar emotional state to another person as a result of the accurate perception of the other person's situation [55]. One may ask a question if the design of the experiment proposed and conducted by Bhurtel et al [1] gave the participants an opportunity to experience the world in a way a person with a lock-in syndrome does. The lack of increase in the empathy score among the participants can be the result of the lack of experiencing the situation of the person unable to communicate in any other way than with their eyes.

The development of Virtual Reality (VR) technology in the recent years made it possible to create a virtual setting where the user can be a part of the context instead of imagining it or watching it from the outside. There were many attempts of using VR in research experiments, some of them reviewed and documented by Wai Hin Wan et al [56]. In this review, they analyse the results from health-related research where the VR simulation is used in order to enhance the empathy level among the care givers and doctors in the health sector. Many experiments proved enhanced empathy toward patients with various challenges like Alzheimer's disease [57], dyslexia [58] and dementia [59]. The results confirm the increase of the empathy among the participants. Creating a VR simulation of a situation of person with a lock-in syndrome could give a better opportunity for the participants to feel the same as the patients.

Another fact worth attention is the test group chosen in the research described by Wai Hin Wan et al. Most of the researchers chose the test group among the caregivers, doctors and people who can relate to the problem in a more personal way. Bhurtel et al also mentions in his research that the people who had a personal relation to a person with lock-in syndrome scored higher in the empathy assessment than other

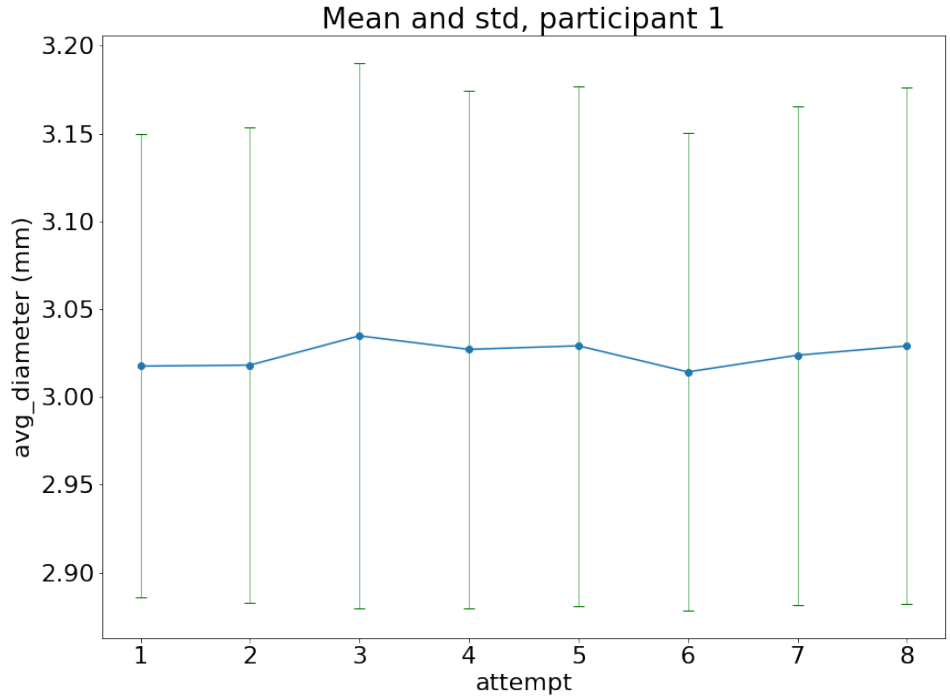
participants.

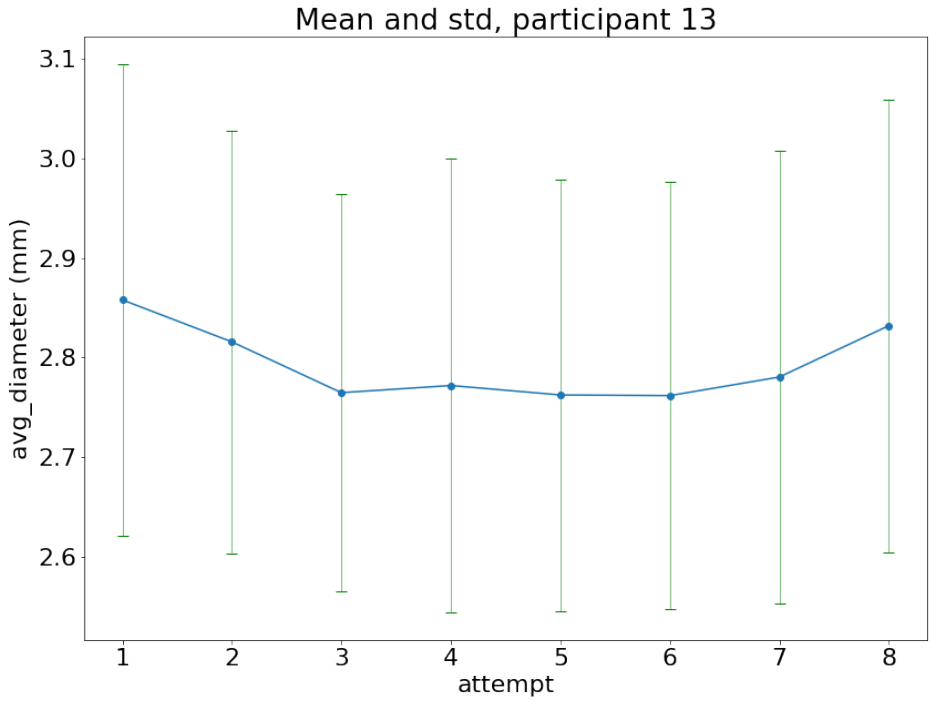
The time frame of this thesis (January 2022 - May 2022) did not allow to conduct a new experiment with any of the proposed changes. In the future, a new experiment with a different design could contribute to collect data to better assess the test hypothesis presented in this thesis.

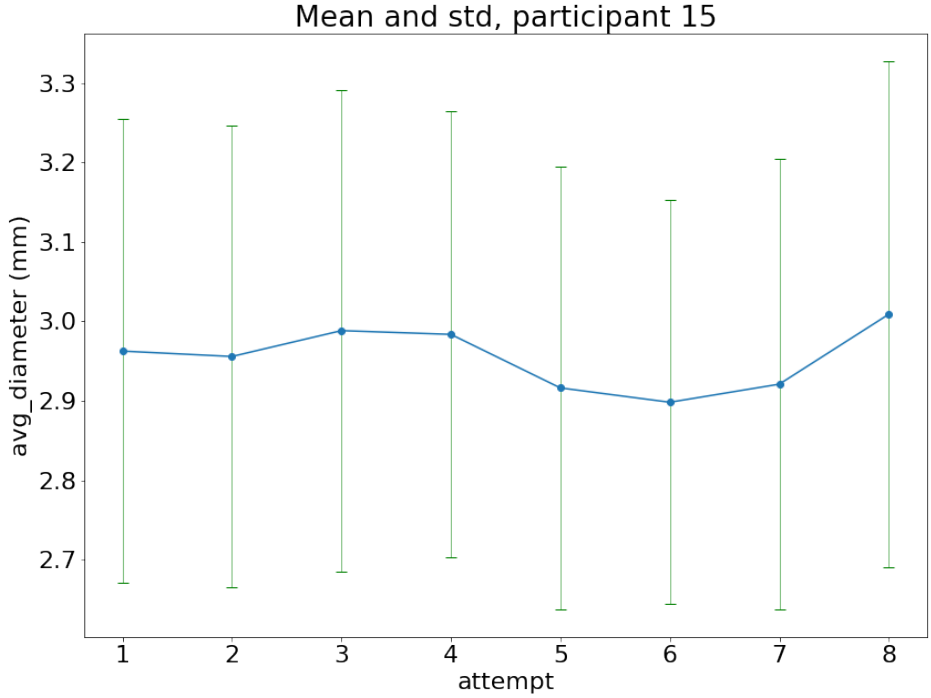
Appendices

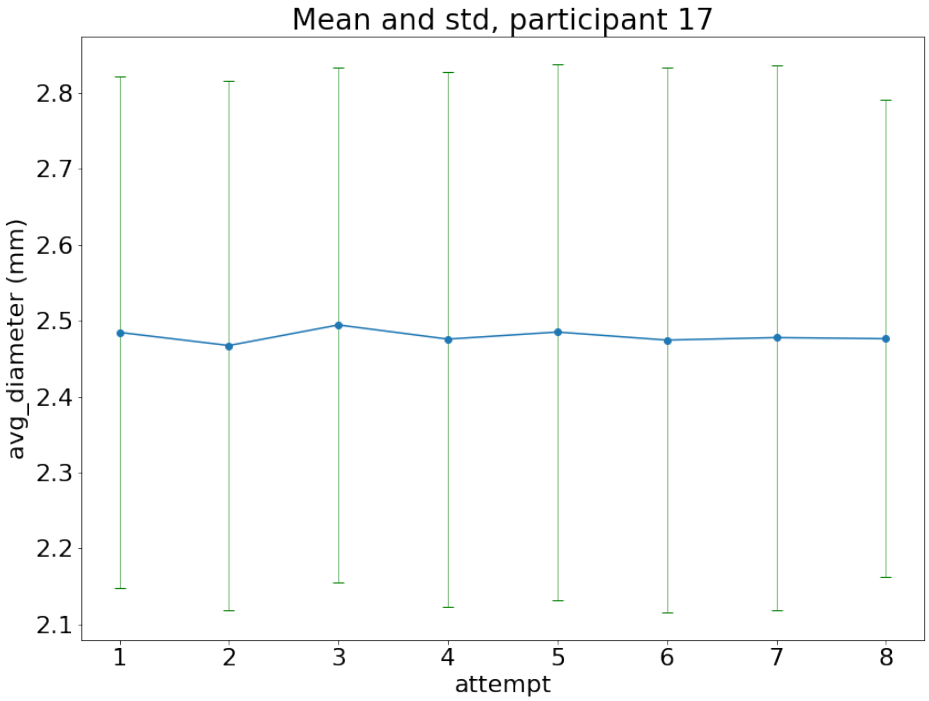
Appendix A

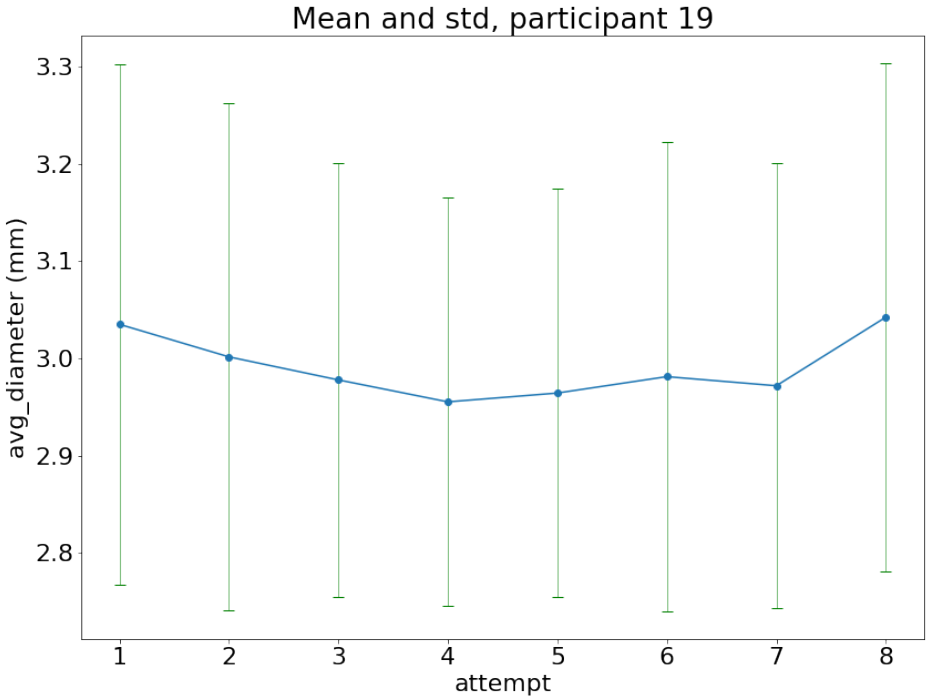
Full results for test group

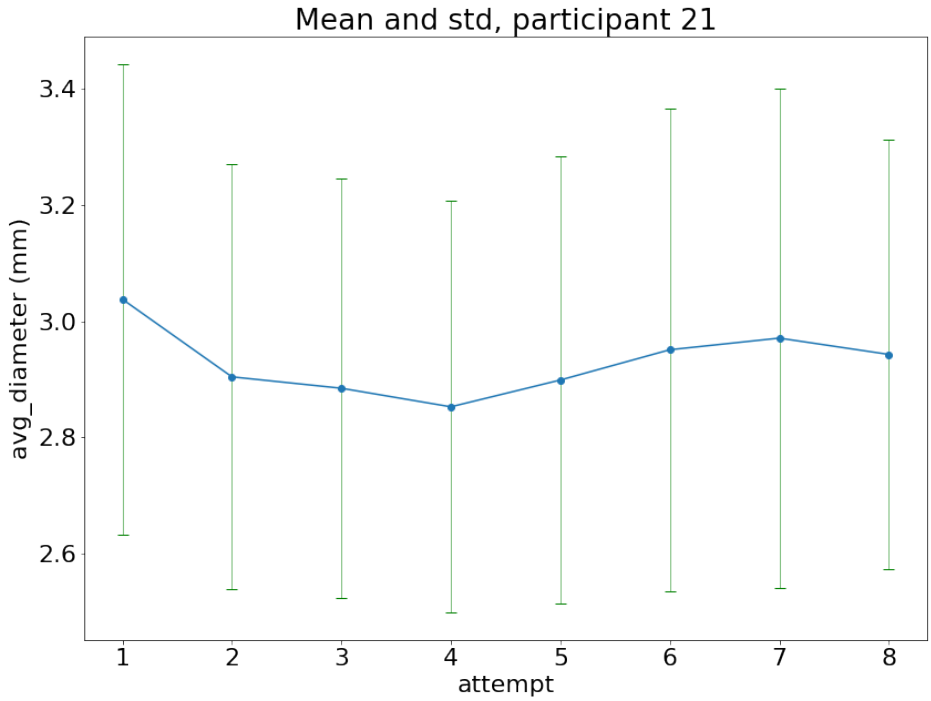


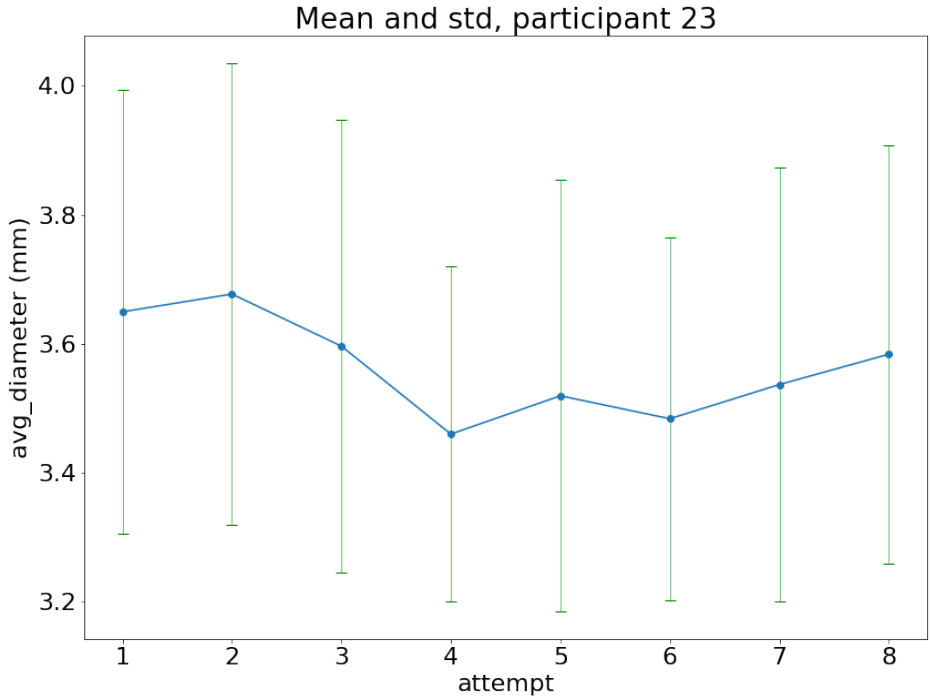

Mean and std, participant 13
avg_diameter (mm)
3.1
3.0
2.9
2.8
2.7
2.6
1
2
3
4
5
6
7
8
attempt

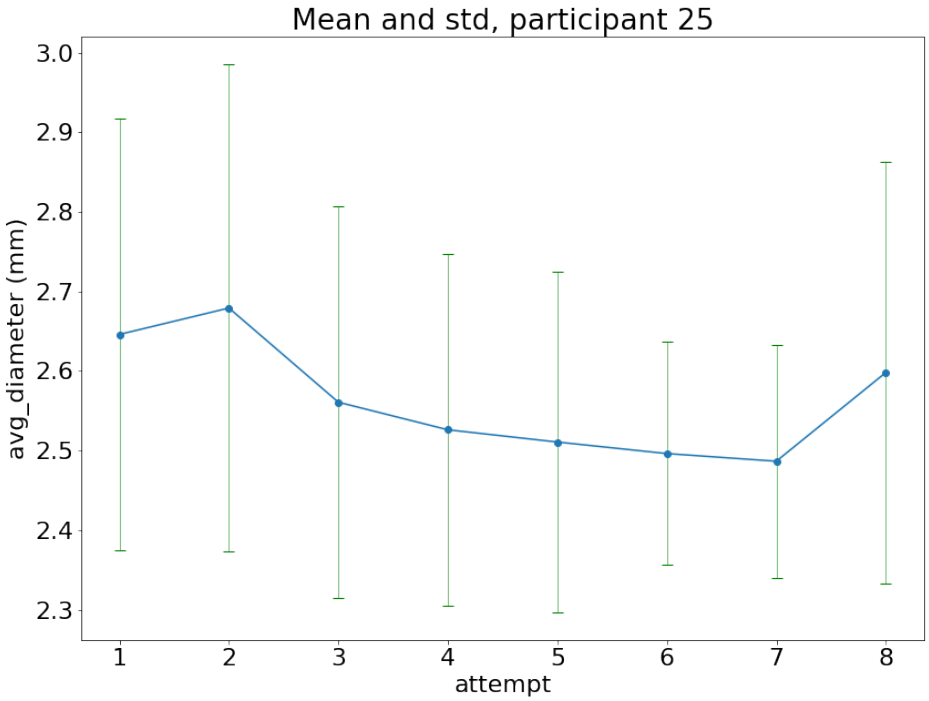

Mean and std, participant 15
avg_diameter (mm)
3.3
3.2
3.1
3.0
2.9
2.8
2.7
1
2
3
4
5
6
7
8
attempt

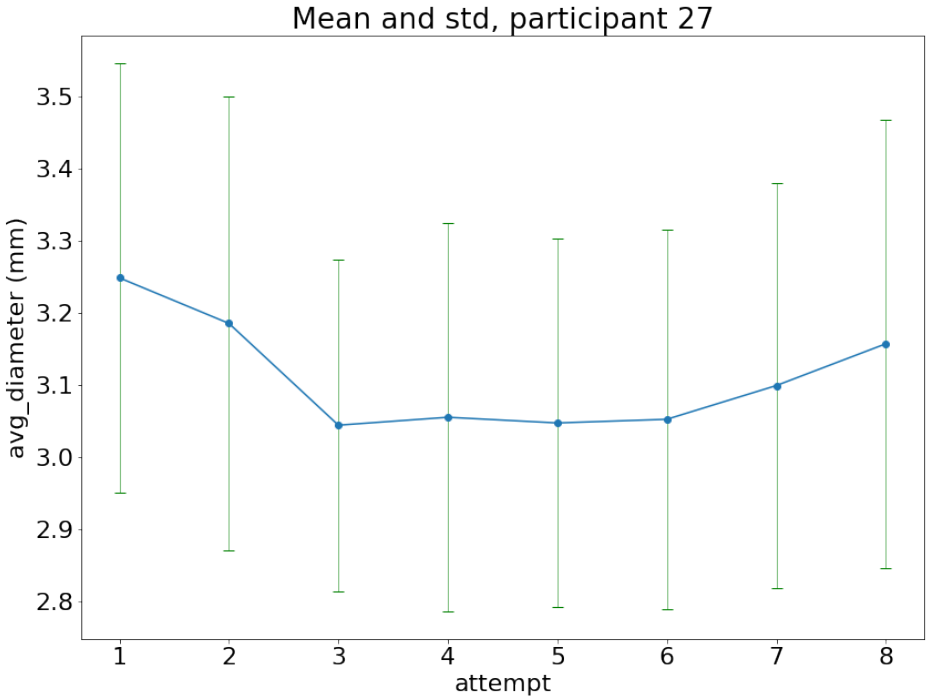

Mean and std, participant 17
avg_diameter (mm)
2.8
2.7
2.6
2.5
2.4
2.3
2.2
2.1
1
2
3
4
5
6
7
8
attempt

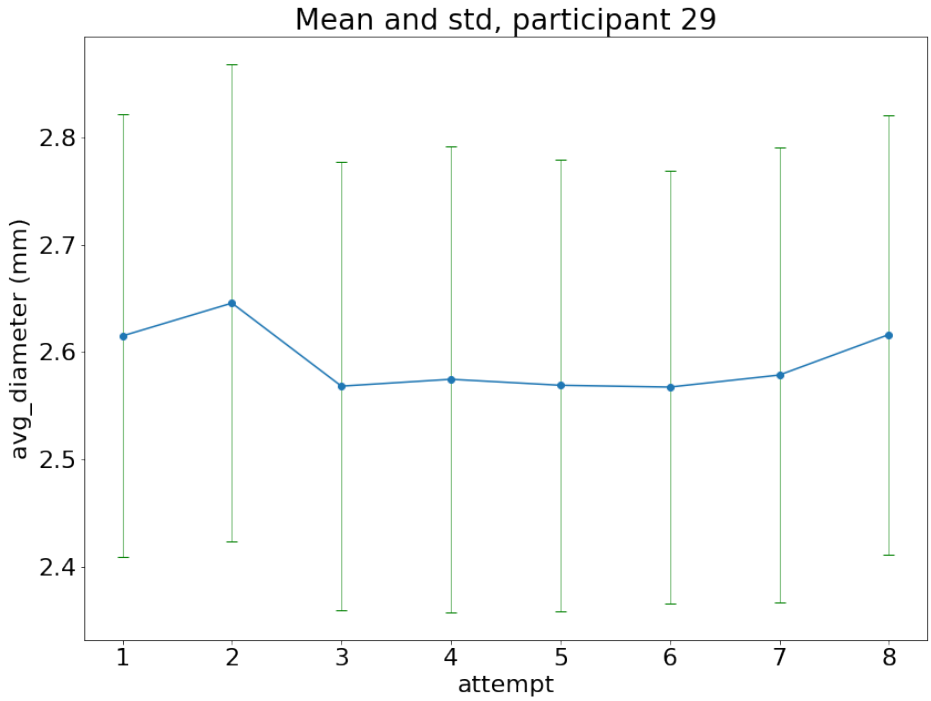

Mean and std, participant 19
avg_diameter (mm)
3.3
3.2
3.1
3.0
2.9
2.8
1
2
3
4
5
6
7
8
attempt

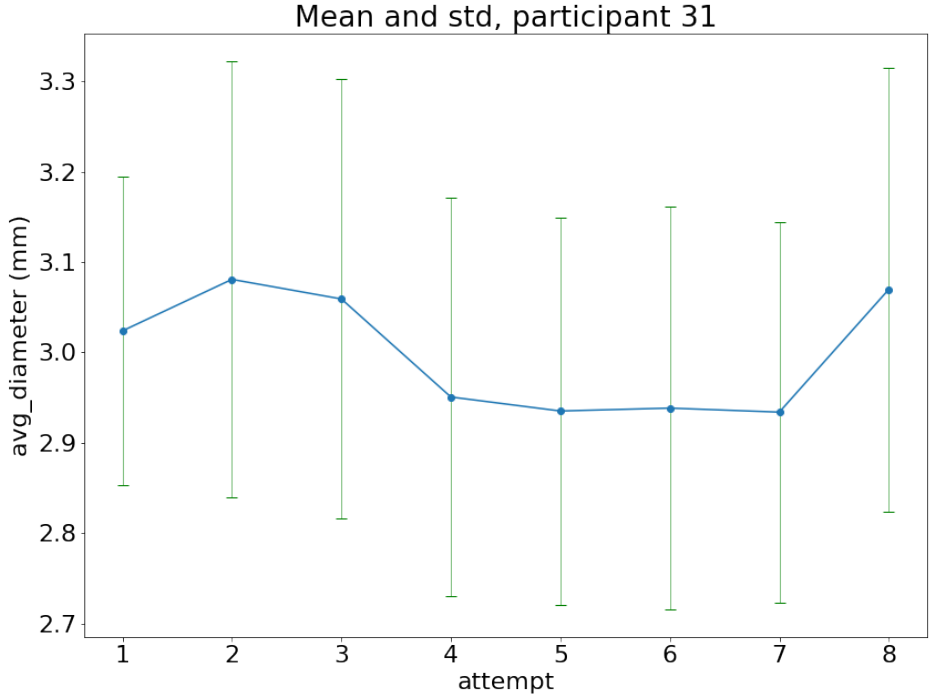

Mean and std, participant 21
avg_diameter (mm)
3.4
3.2
3.0
2.8
2.6
1
2
3
4
5
6
7
8
attempt

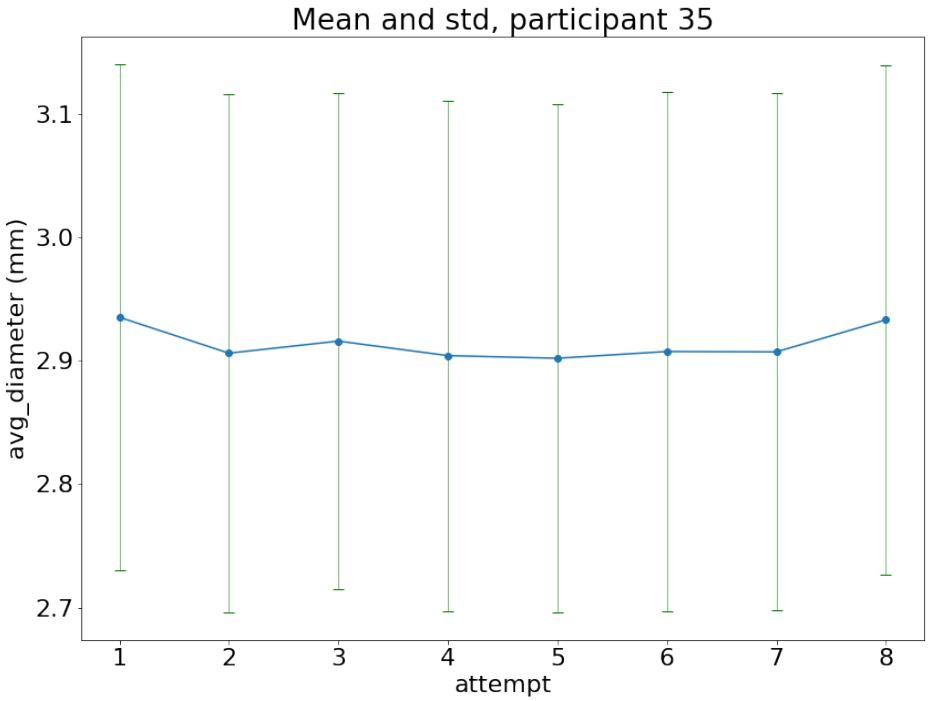

Mean and std, participant 23
avg_diameter (mm)
4.0
3.8
3.6
3.4
3.2
1
2
3
4
5
6
7
8
attempt

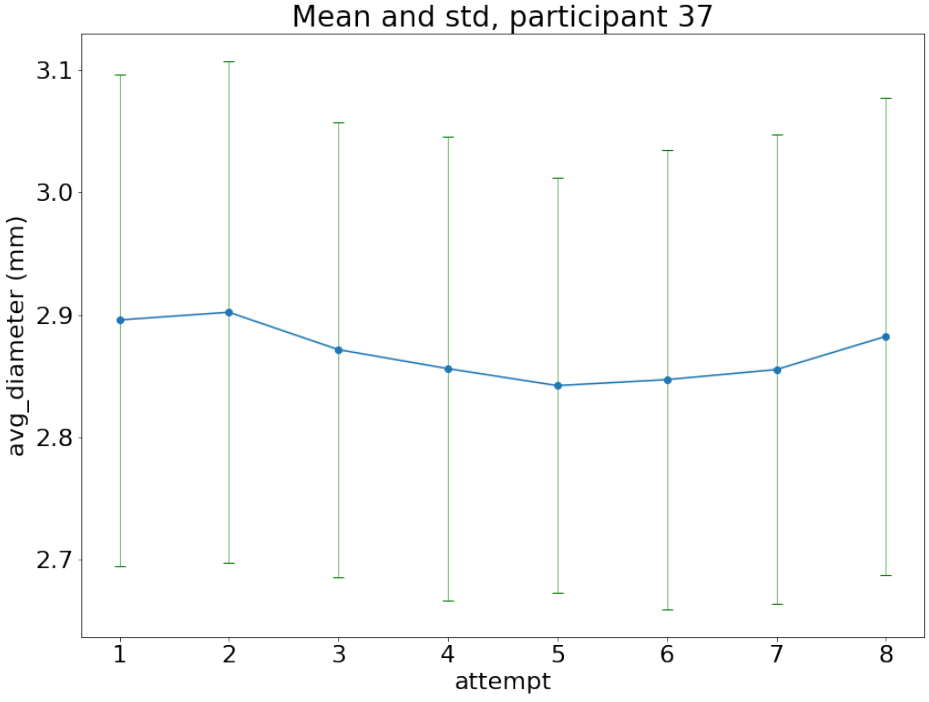

Mean and std, participant 25
avg_diameter (mm)
3.0
2.9
2.8
2.7
2.6
2.5
2.4
2.3
1
2
3
4
5
6
7
8
attempt

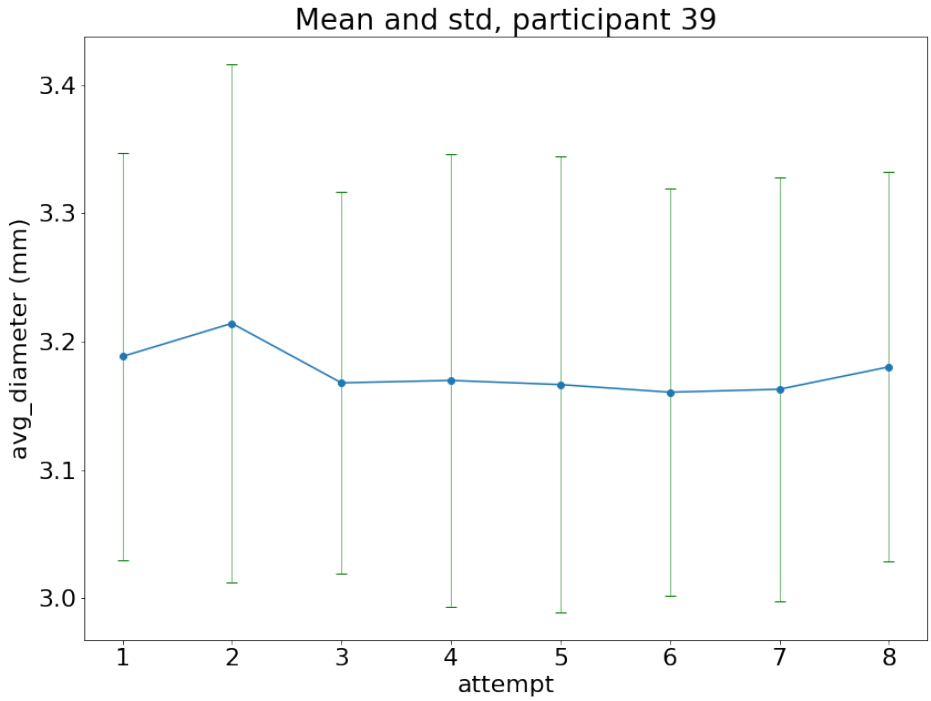

Mean and std, participant 27
avg_diameter (mm)
3.5
3.4
3.3
3.2
3.1
3.0
2.9
2.8
1
2
3
4
5
6
7
8
attempt

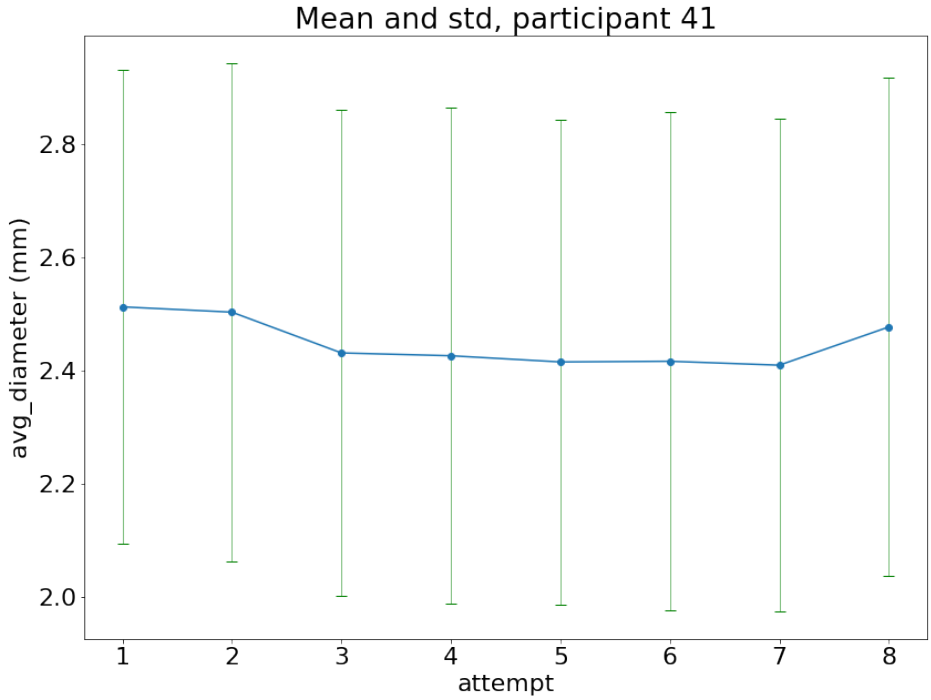

Mean and std, participant 29
avg_diameter (mm)
2.8
2.7
2.6
2.5
2.4
1
2
3
4
5
6
7
8
attempt

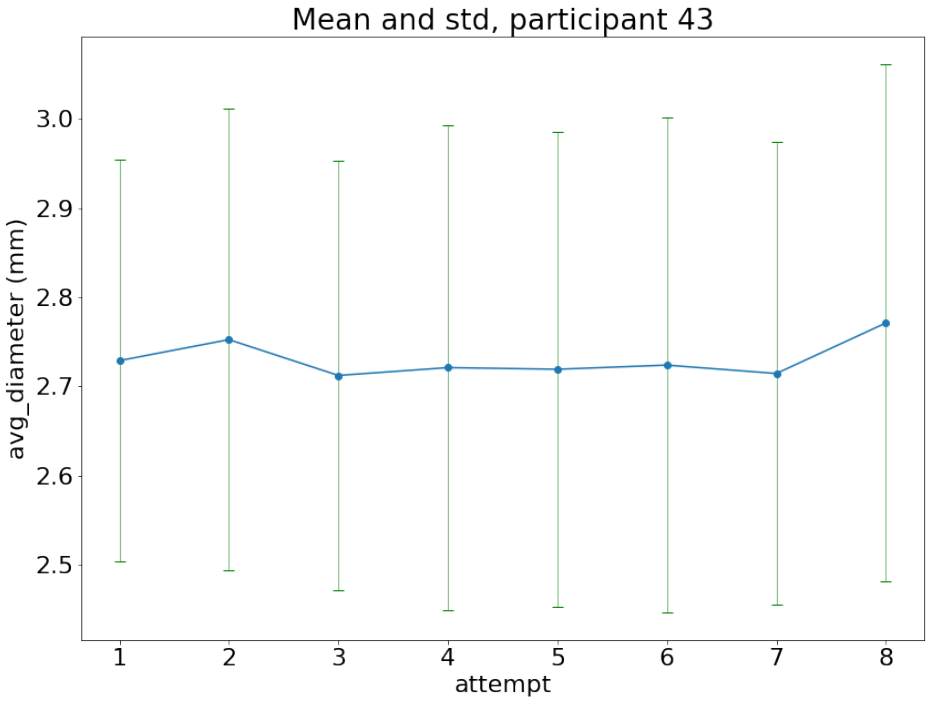

Mean and std, participant 31
avg_diameter (mm)
3.3
3.2
3.1
3.0
2.9
2.8
2.7
1
2
3
4
5
6
7
8
attempt

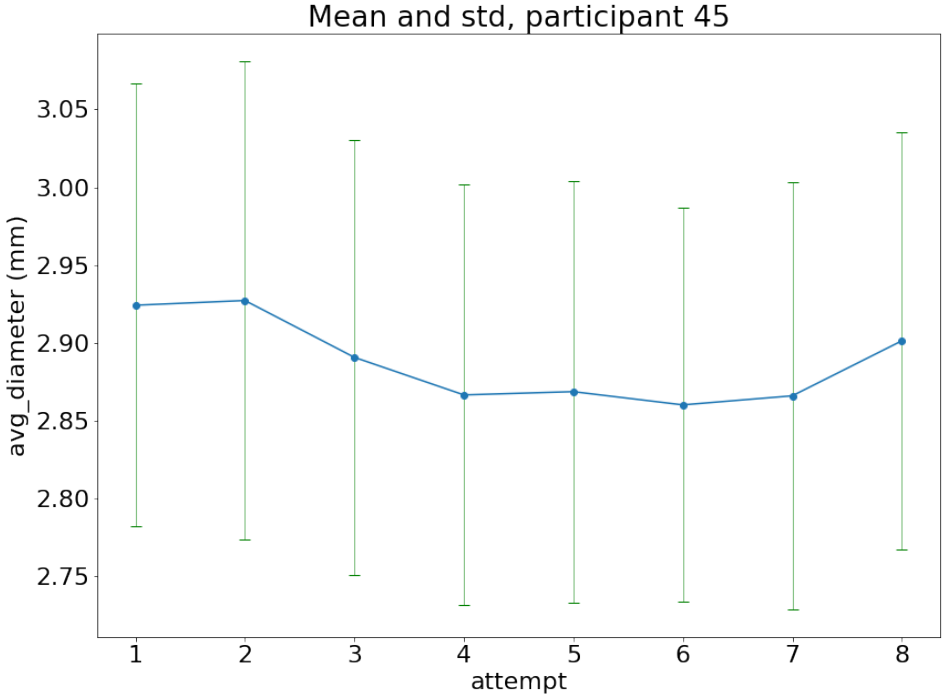

Mean and std, participant 35
avg_diameter (mm)
3.1
3.0
2.9
2.8
2.7
1
2
3
4
5
6
7
8
attempt

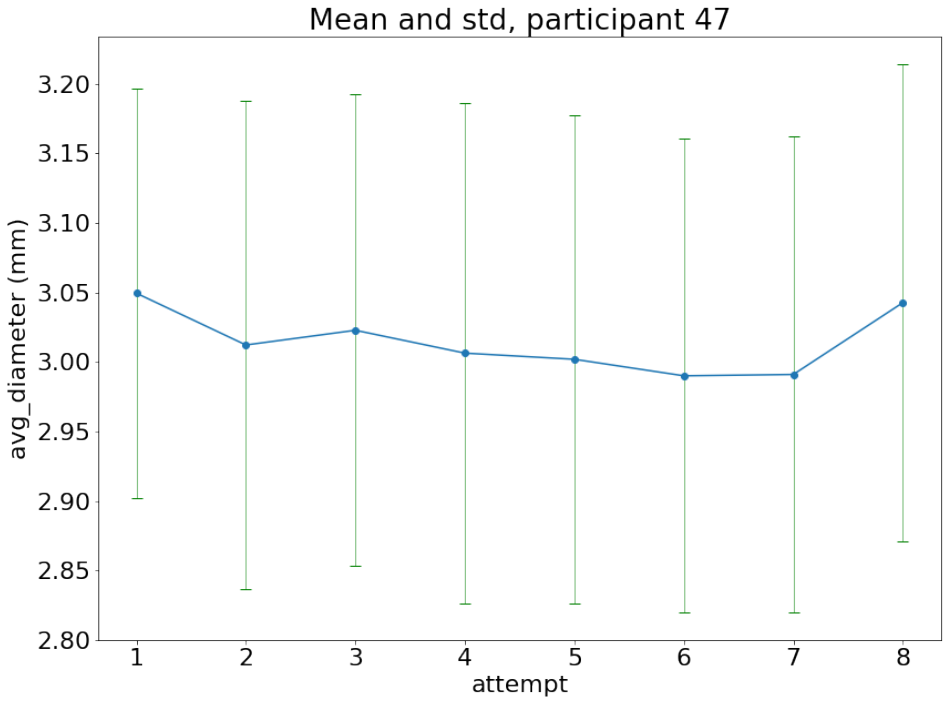

Mean and std, participant 37
avg_diameter (mm)
3.1
3.0
2.9
2.8
2.7
1
2
3
4
5
6
7
8
attempt

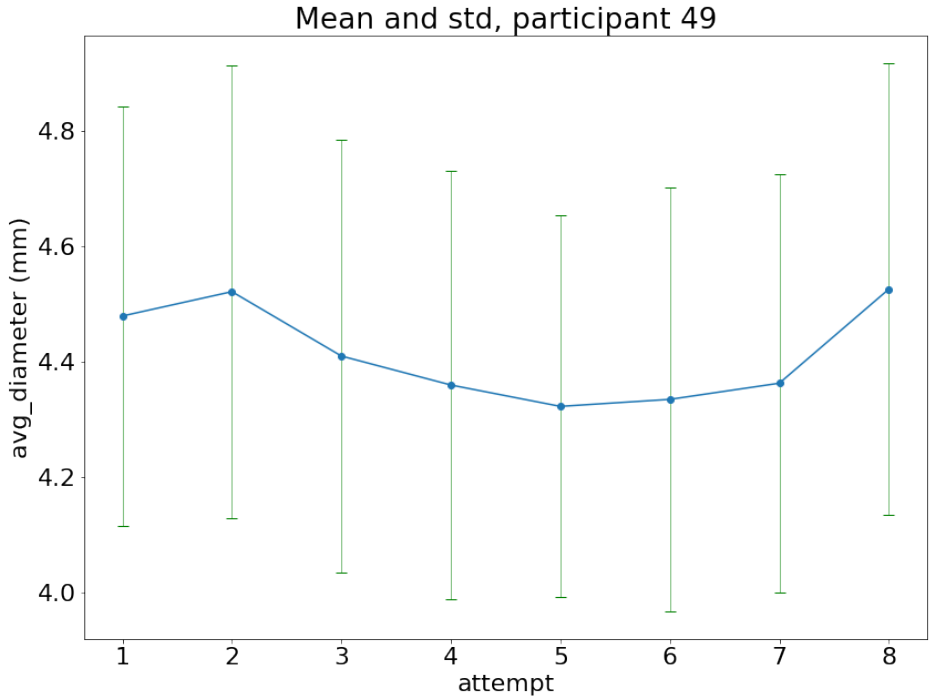

Mean and std, participant 39
avg_diameter (mm)
3.4
3.3
3.2
3.1
3.0
1
2
3
4
5
6
7
8
attempt

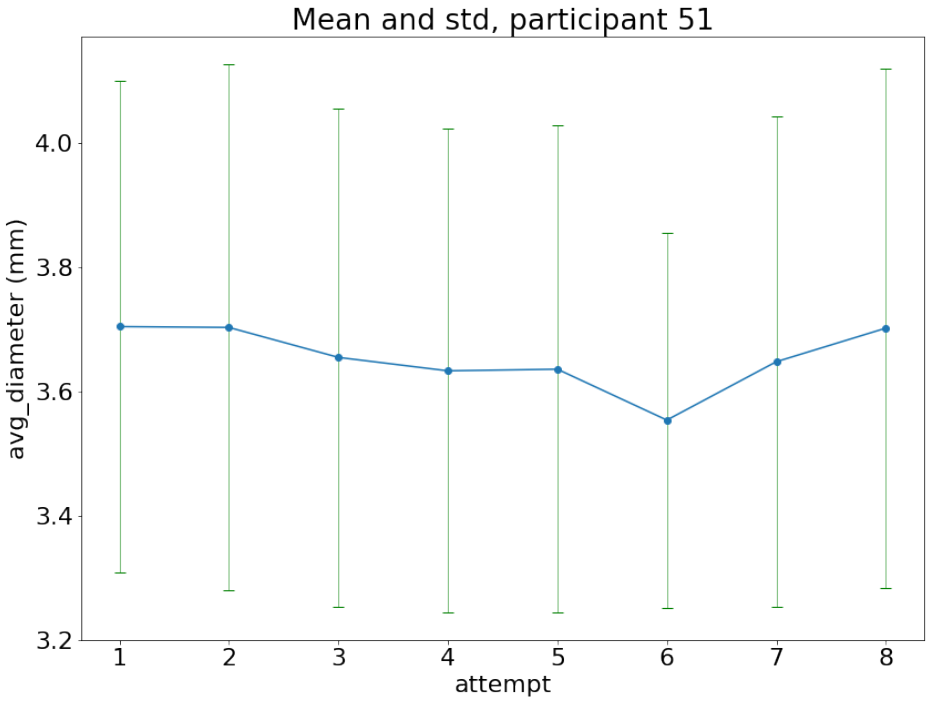

Mean and std, participant 41
avg_diameter (mm)
2.8
2.6
2.4
2.2
2.0
1
2
3
4
5
6
7
8
attempt

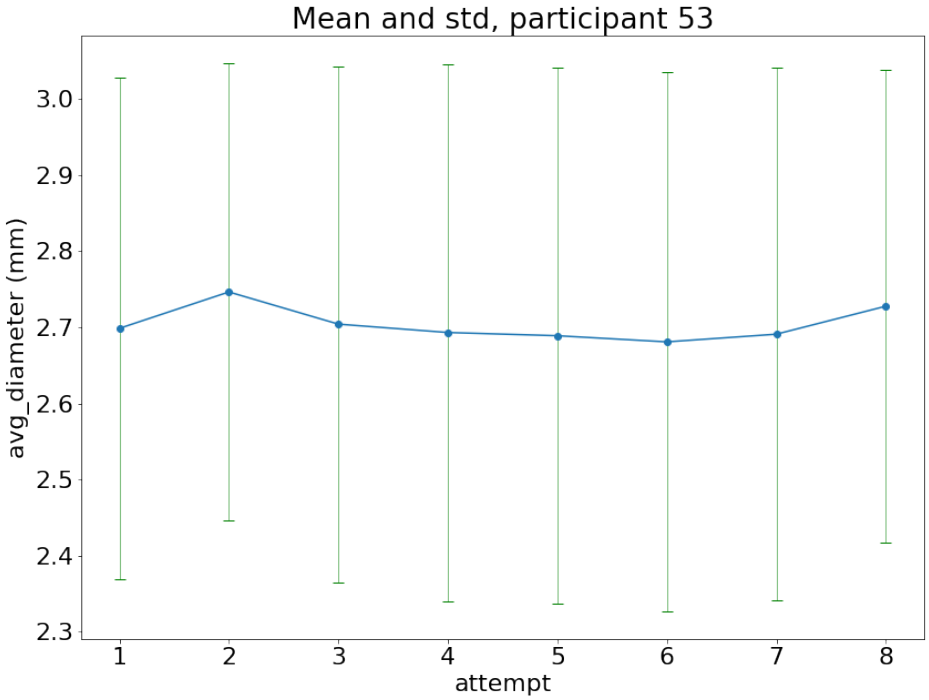

Mean and std, participant 43
avg_diameter (mm)
3.0
2.9
2.8
2.7
2.6
2.5
1
2
3
4
5
6
7
8
attempt

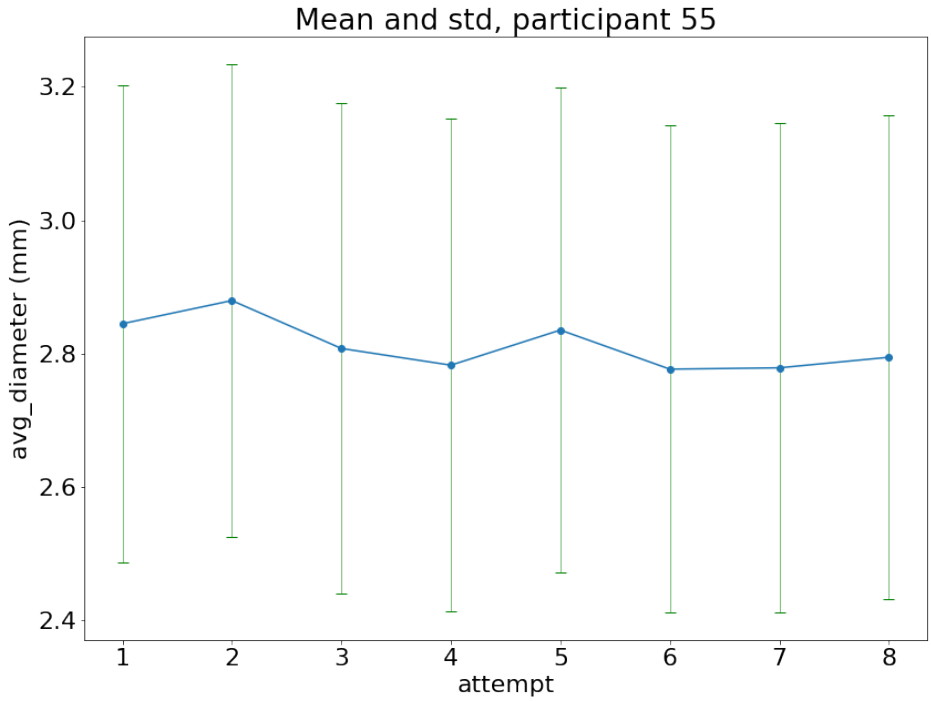

Mean and std, participant 45
avg_diameter (mm)
3.05
3.00
2.95
2.90
2.85
2.80
2.75
1
2
3
4
5
6
7
8
attempt

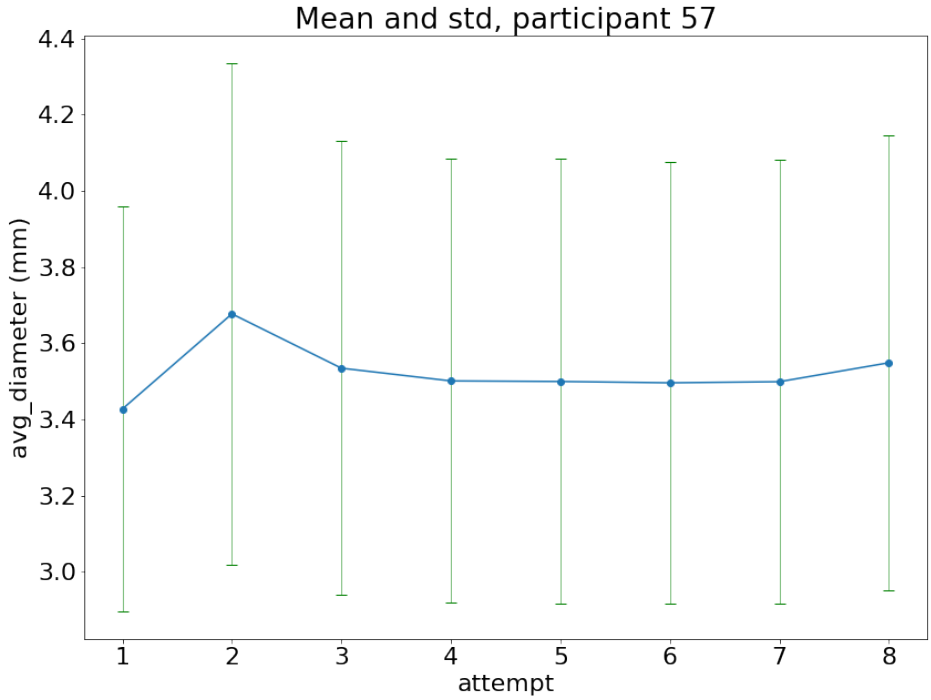

Mean and std, participant 47
avg_diameter (mm)
3.20
3.15
3.10
3.05
3.00
2.95
2.90
2.85
2.80
1
2
3
4
5
6
7
8
attempt

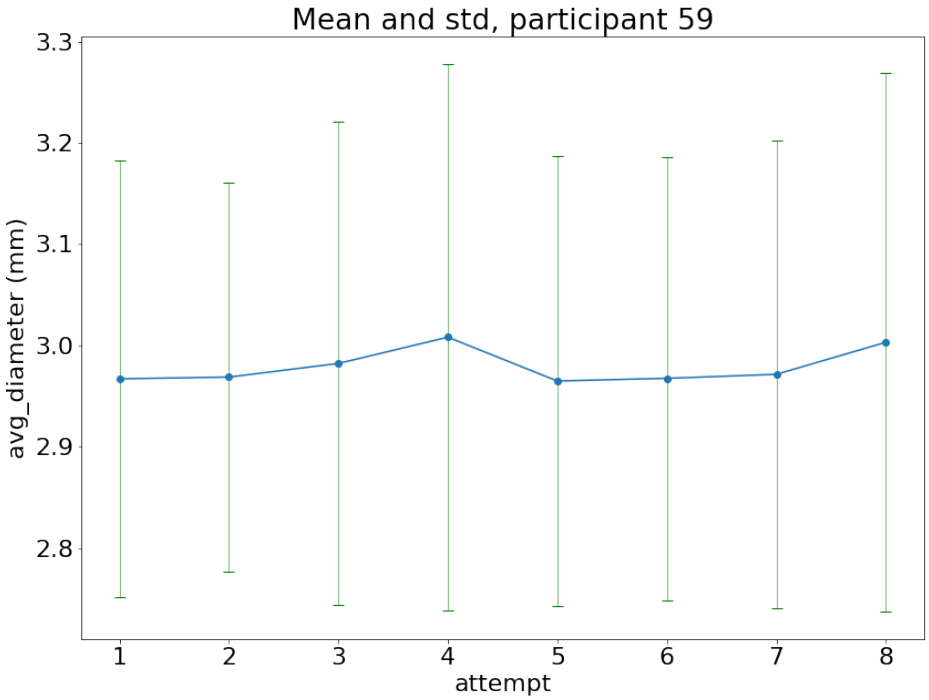

Mean and std, participant 49
avg_diameter (mm)
4.8
4.6
4.4
4.2
4.0
1
2
3
4
5
6
7
8
attempt


Mean and std, participant 51
avg_diameter (mm)
4.0
3.8
3.6
3.4
3.2
1
2
3
4
5
6
7
8
attempt


Mean and std, participant 53
avg_diameter (mm)
3.0
2.9
2.8
2.7
2.6
2.5
2.4
2.3
1
2
3
4
5
6
7
8
attempt

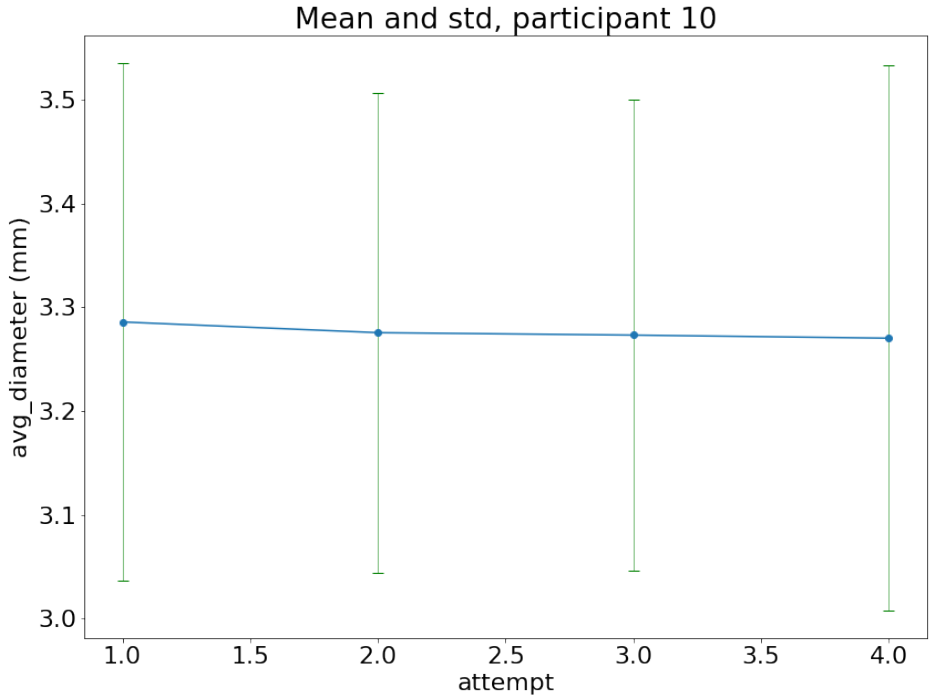

Mean and std, participant 55
avg_diameter (mm)
3.2
3.0
2.8
2.6
2.4
1
2
3
4
5
6
7
8
attempt

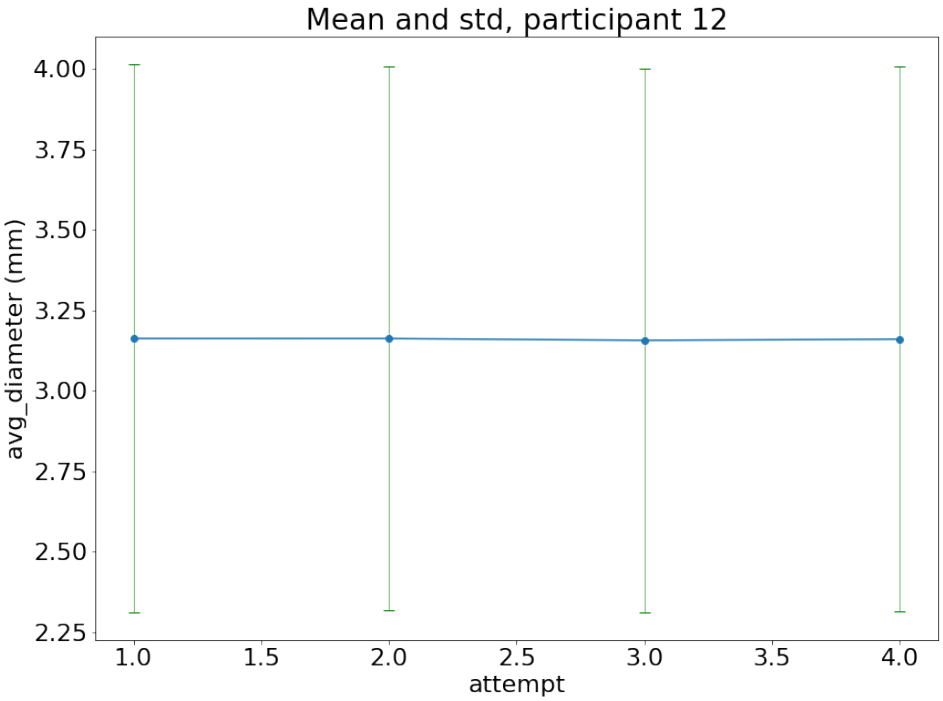

Mean and std, participant 57
avg_diameter (mm)
4.4
4.2
4.0
3.8
3.6
3.4
3.2
3.0
1
2
3
4
5
6
7
8
attempt

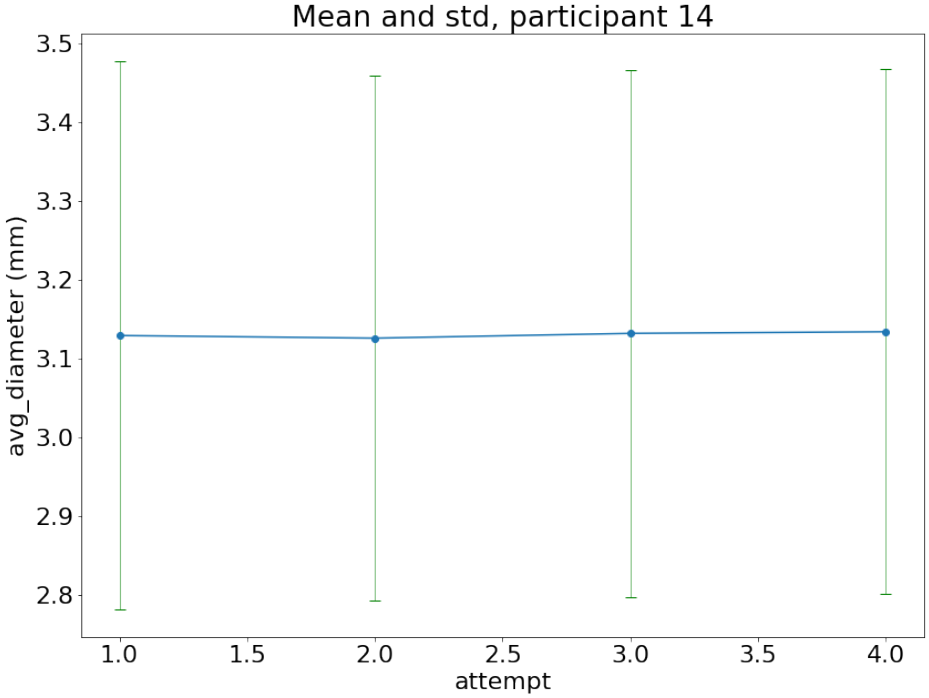

Mean and std, participant 59
avg_diameter (mm)
3.3
3.2
3.1
3.0
2.9
2.8
1
2
3
4
5
6
7
8
attempt

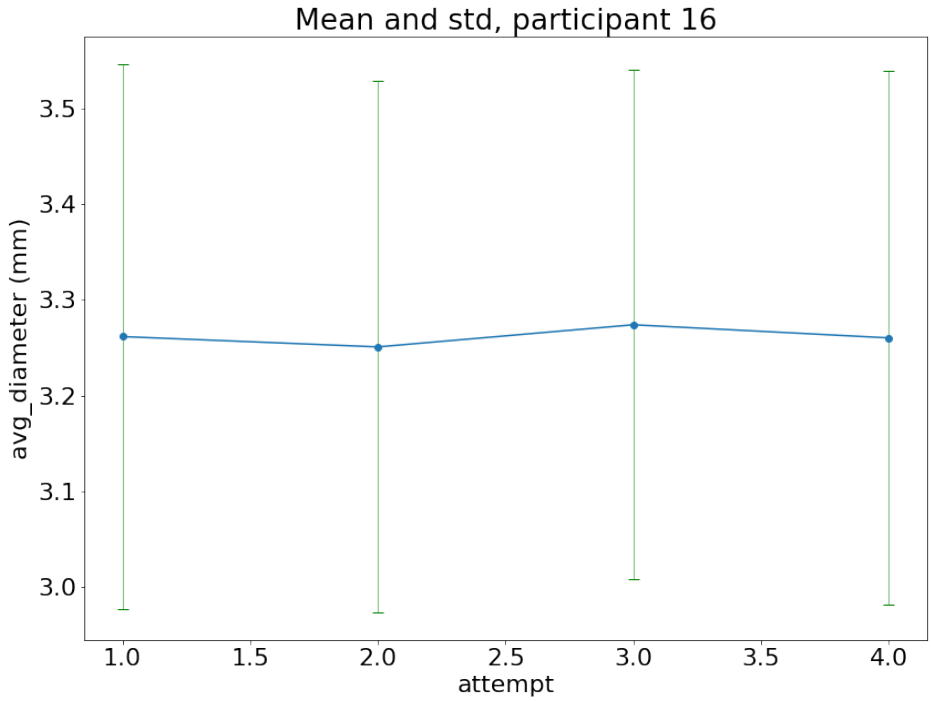
Appendix B

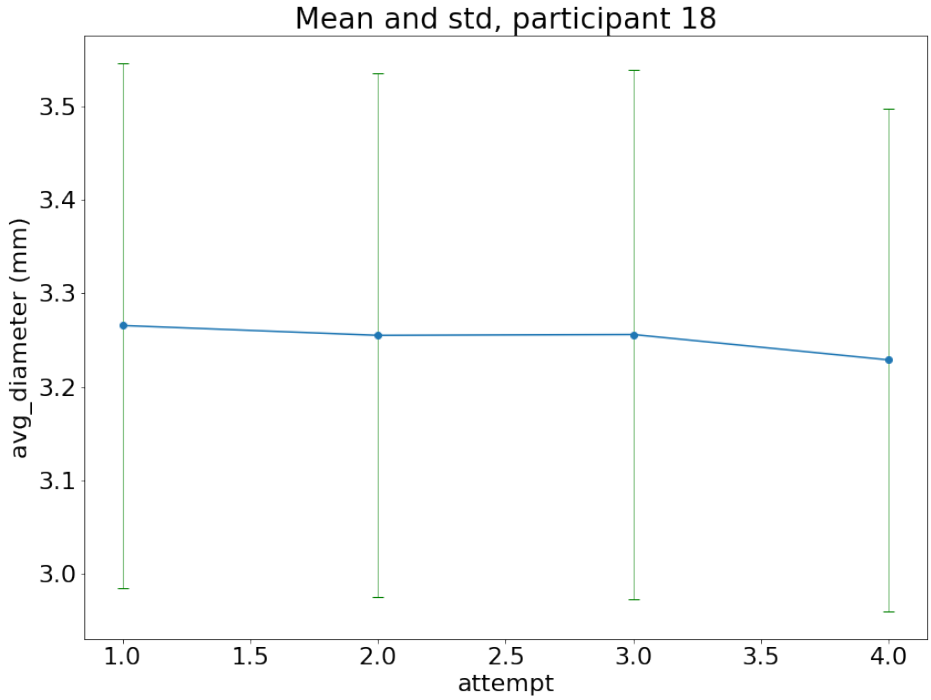
Full results for control group

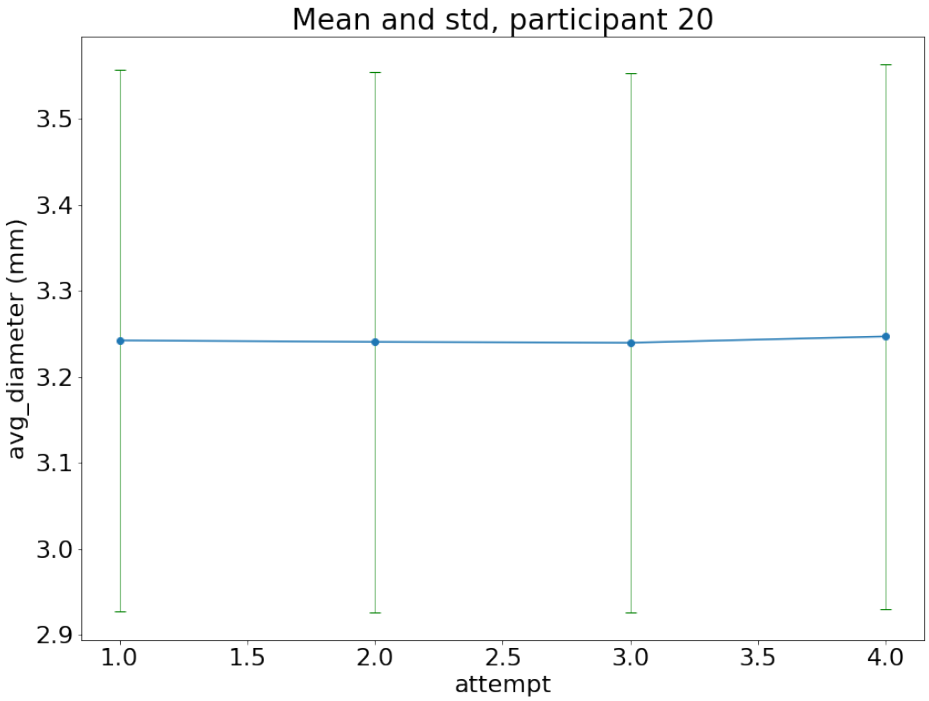


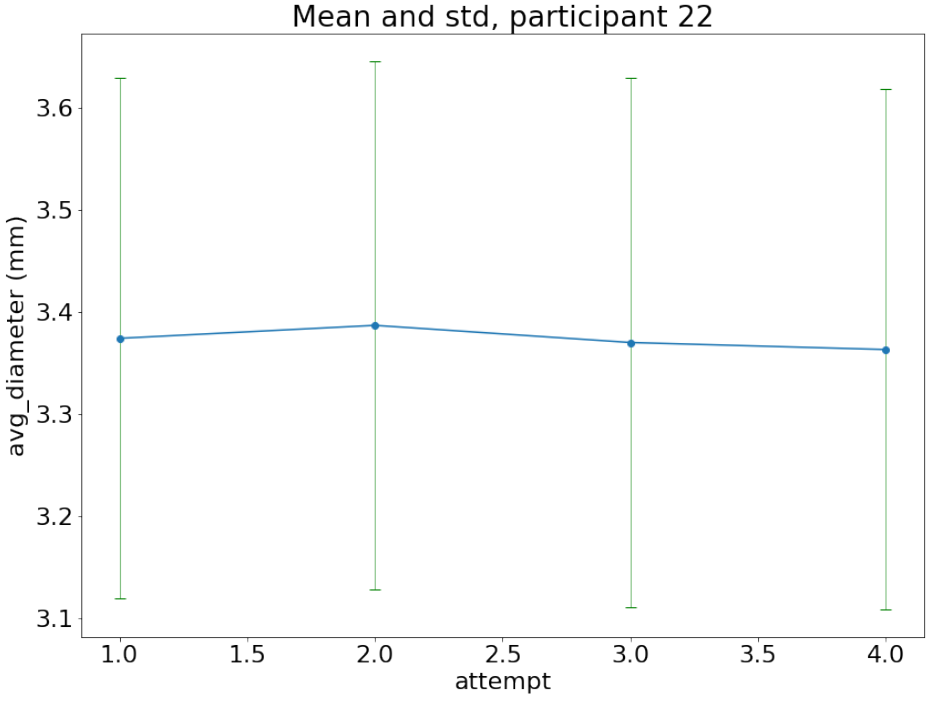

Mean and std, participant 12
avg_diameter (mm)
4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.25
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt

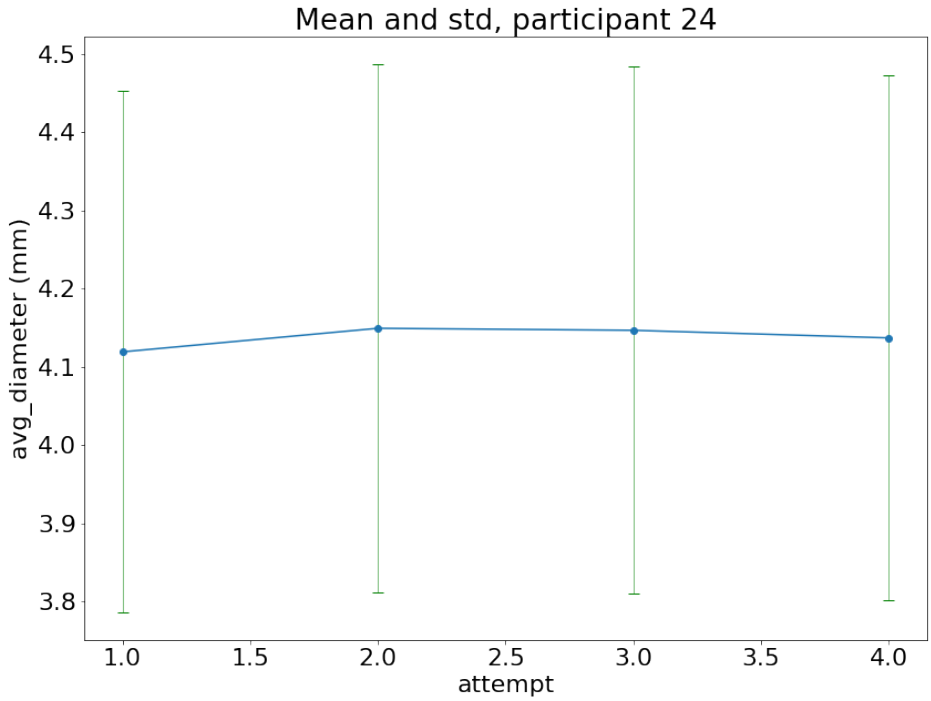

Mean and std, participant 14
avg_diameter (mm)
3.5
3.4
3.3
3.2
3.1
3.0
2.9
2.8
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt

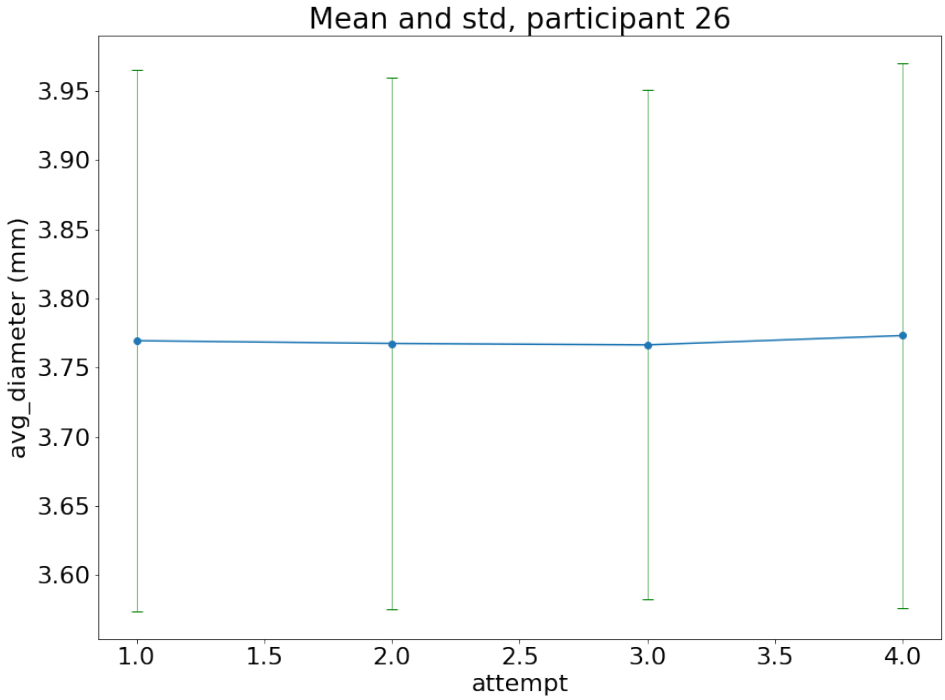

Mean and std, participant 16
avg_diameter (mm)
3.5
3.4
3.3
3.2
3.1
3.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt

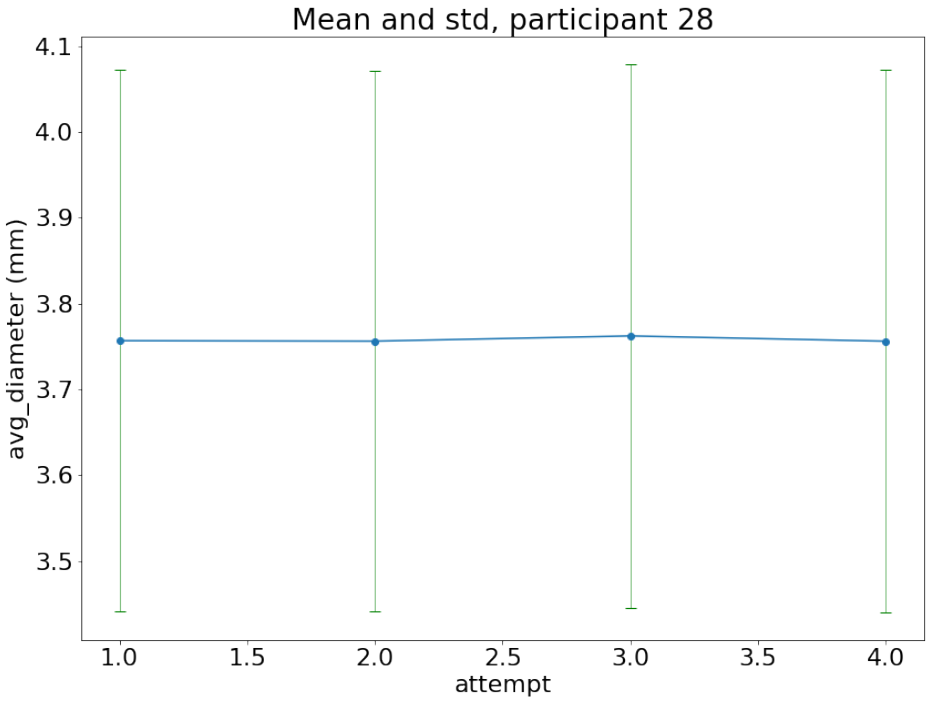

Mean and std, participant 18
avg_diameter (mm)
3.5
3.4
3.3
3.2
3.1
3.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt

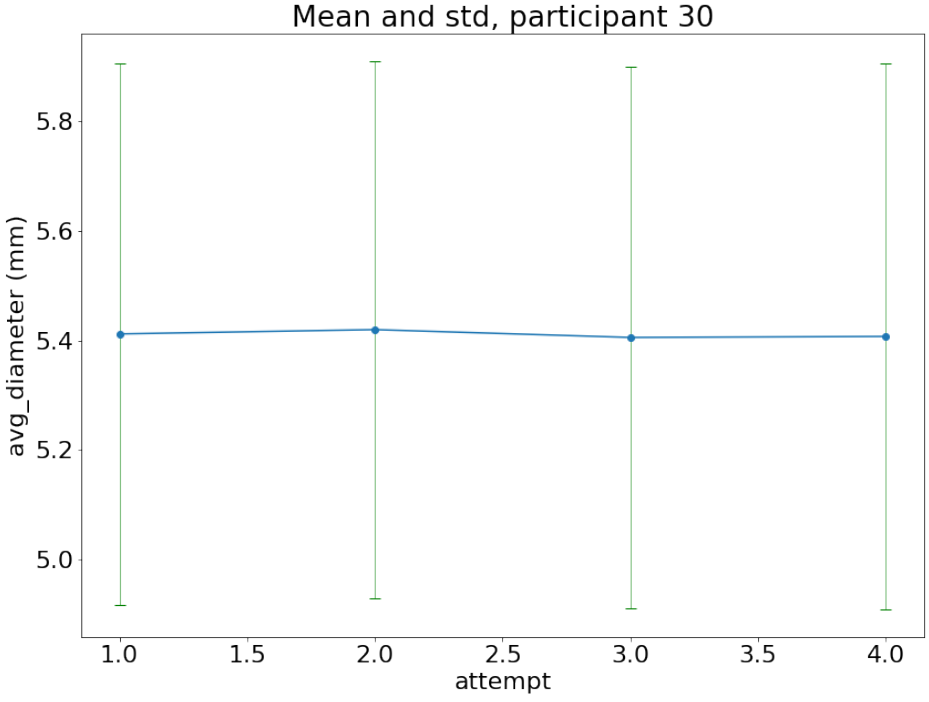

Mean and std, participant 20
avg_diameter (mm)
3.5
3.4
3.3
3.2
3.1
3.0
2.9
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt

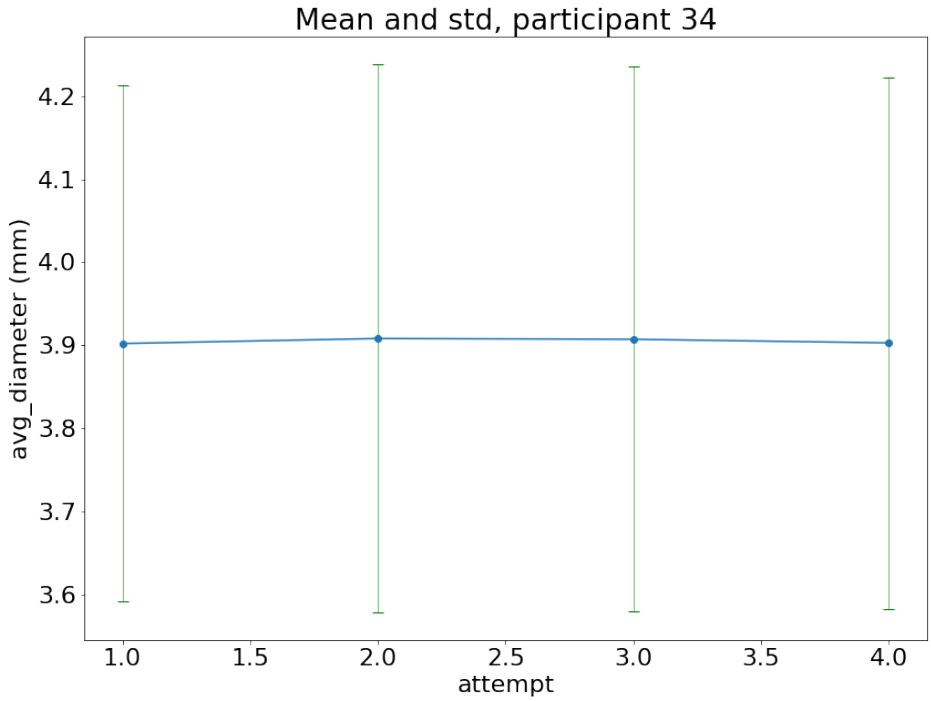

Mean and std, participant 22
avg_diameter (mm)
3.6
3.5
3.4
3.3
3.2
3.1
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt

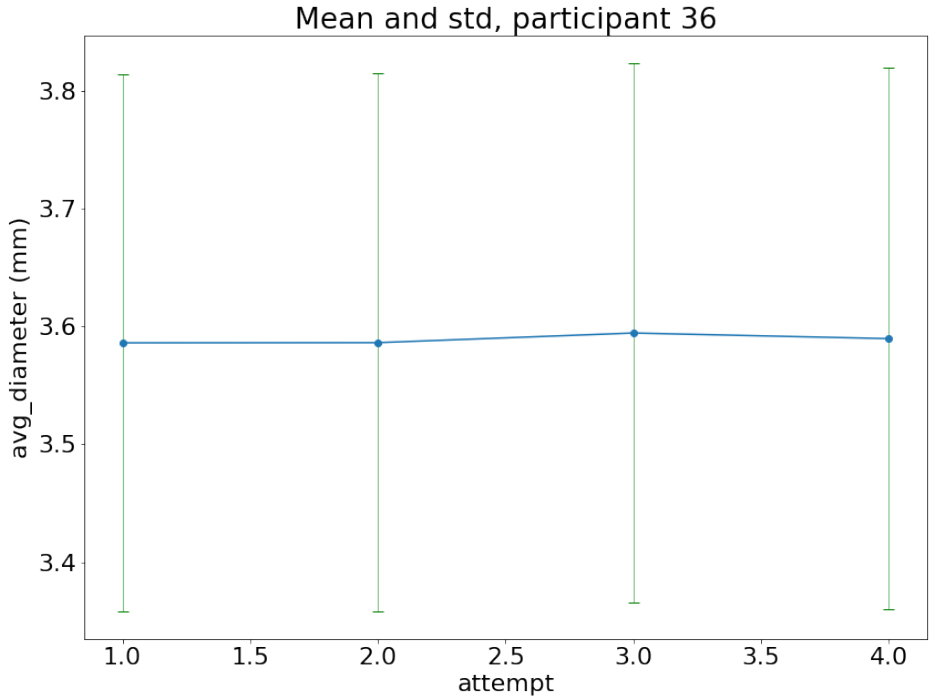

Mean and std, participant 24
avg_diameter (mm)
4.5
4.4
4.3
4.2
4.1
4.0
3.9
3.8
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt

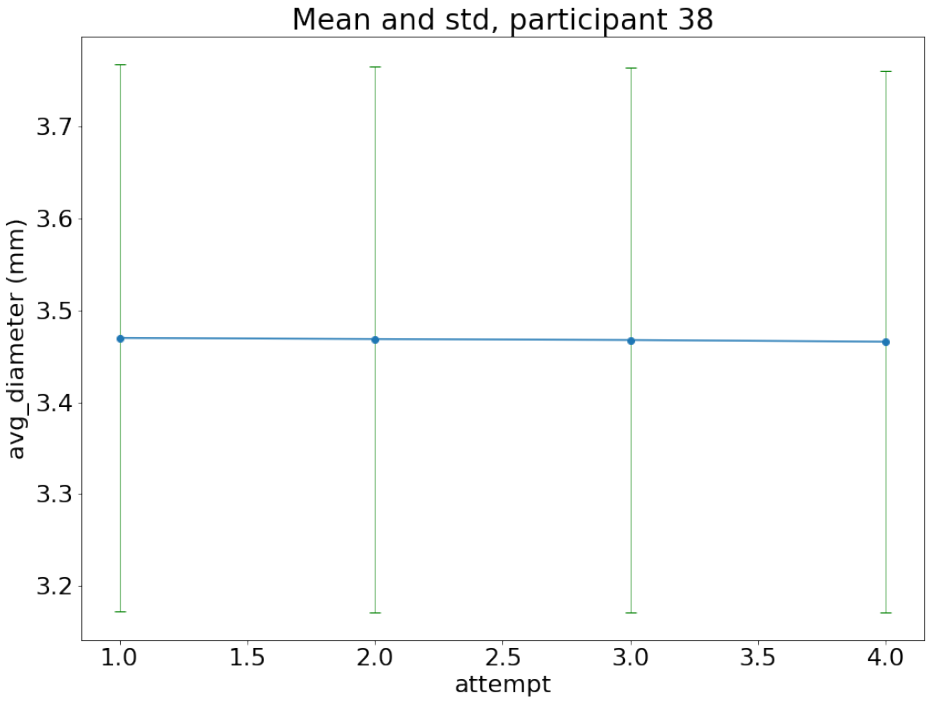

Mean and std, participant 26
avg_diameter (mm)
3.95
3.90
3.85
3.80
3.75
3.70
3.65
3.60
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt

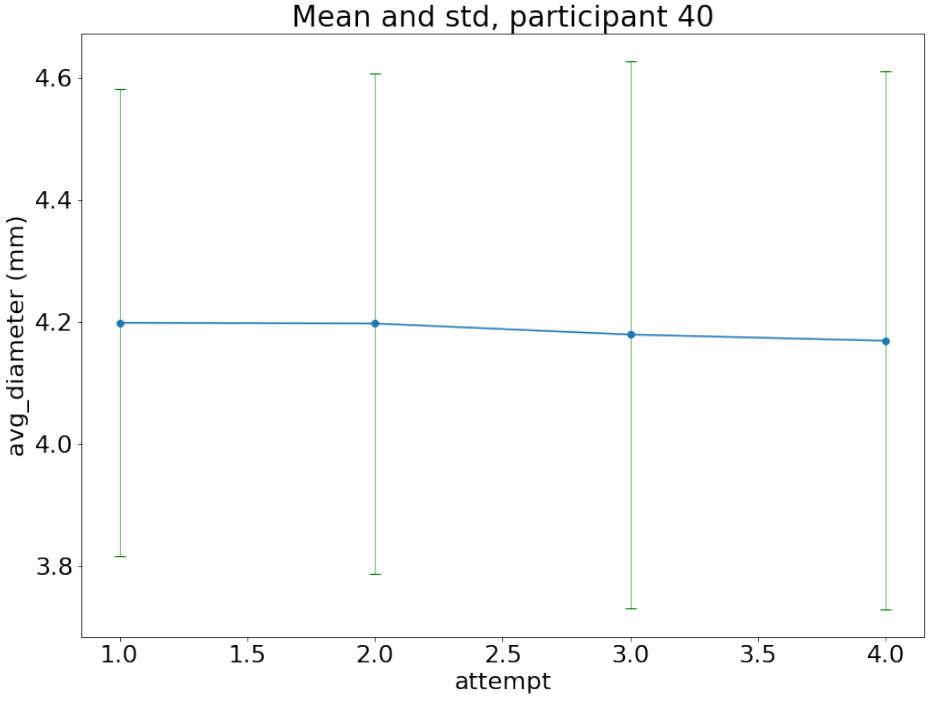

Mean and std, participant 28
avg_diameter (mm)
4.1
4.0
3.9
3.8
3.7
3.6
3.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt

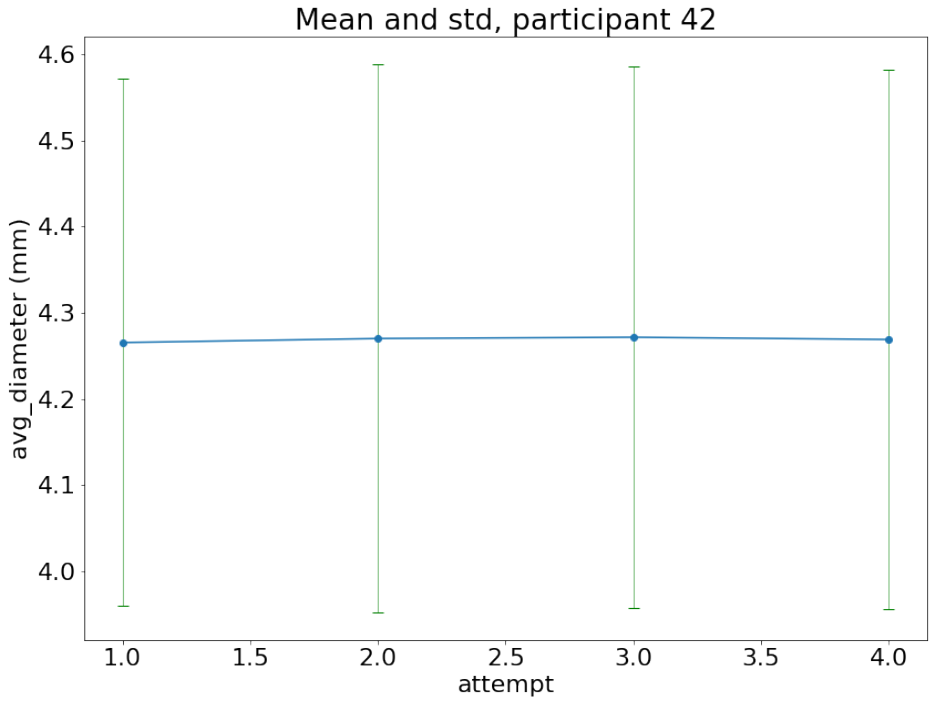

Mean and std, participant 30
avg_diameter (mm)
5.8
5.6
5.4
5.2
5.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt

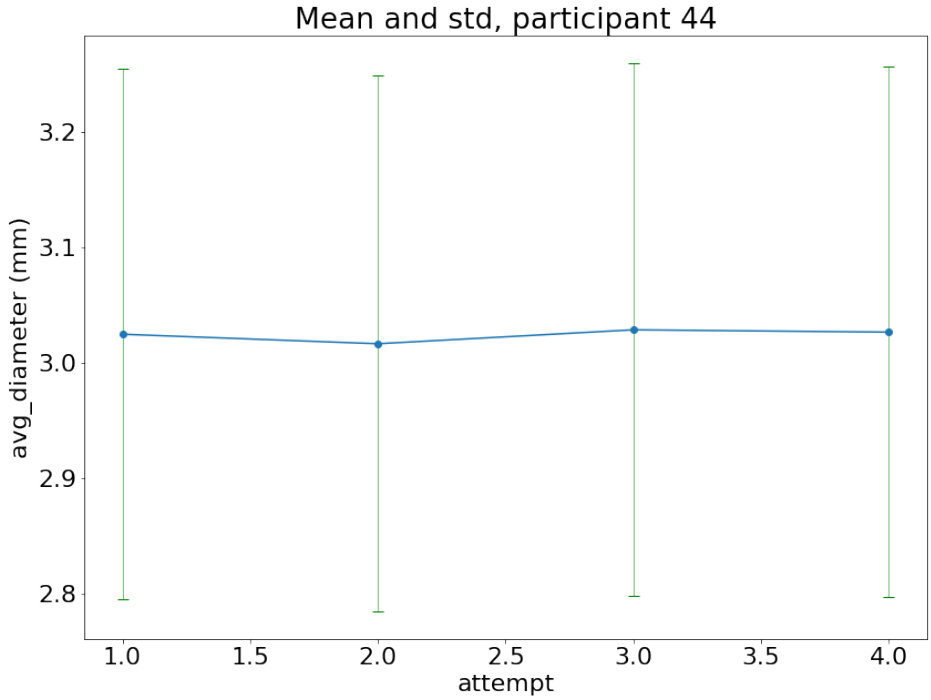

Mean and std, participant 34
avg_diameter (mm)
4.2
4.1
4.0
3.9
3.8
3.7
3.6
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt

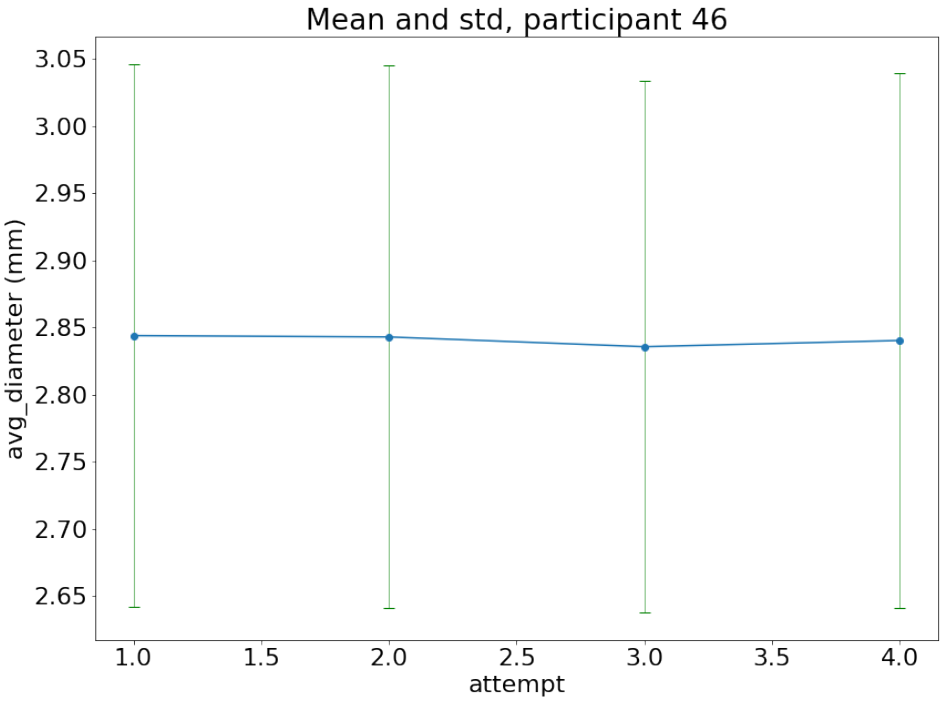

Mean and std, participant 36
avg_diameter (mm)
3.8
3.7
3.6
3.5
3.4
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt

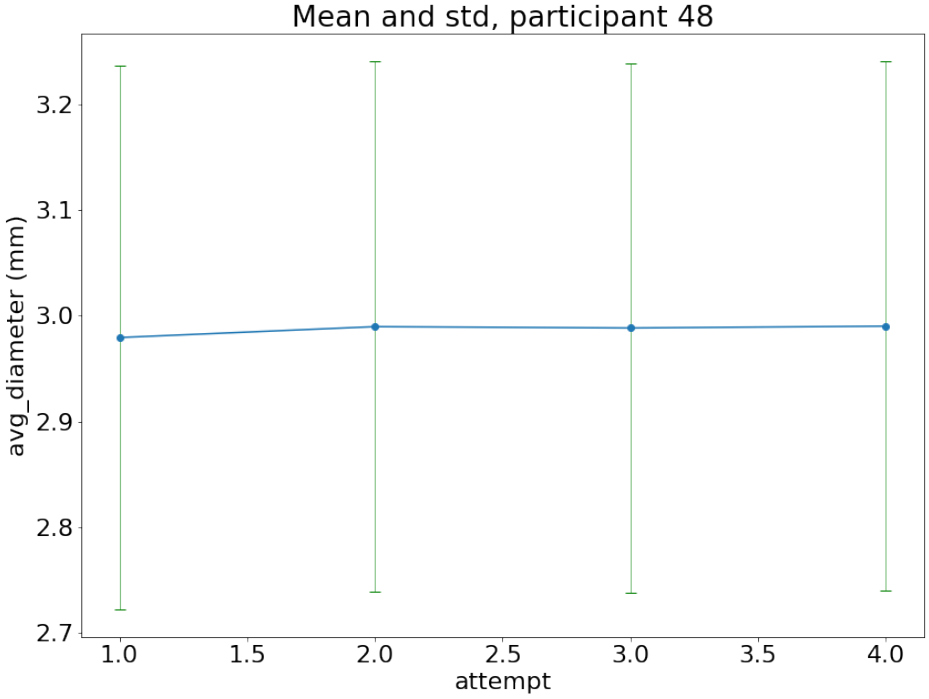

Mean and std, participant 38
avg_diameter (mm)
3.7
3.6
3.5
3.4
3.3
3.2
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt

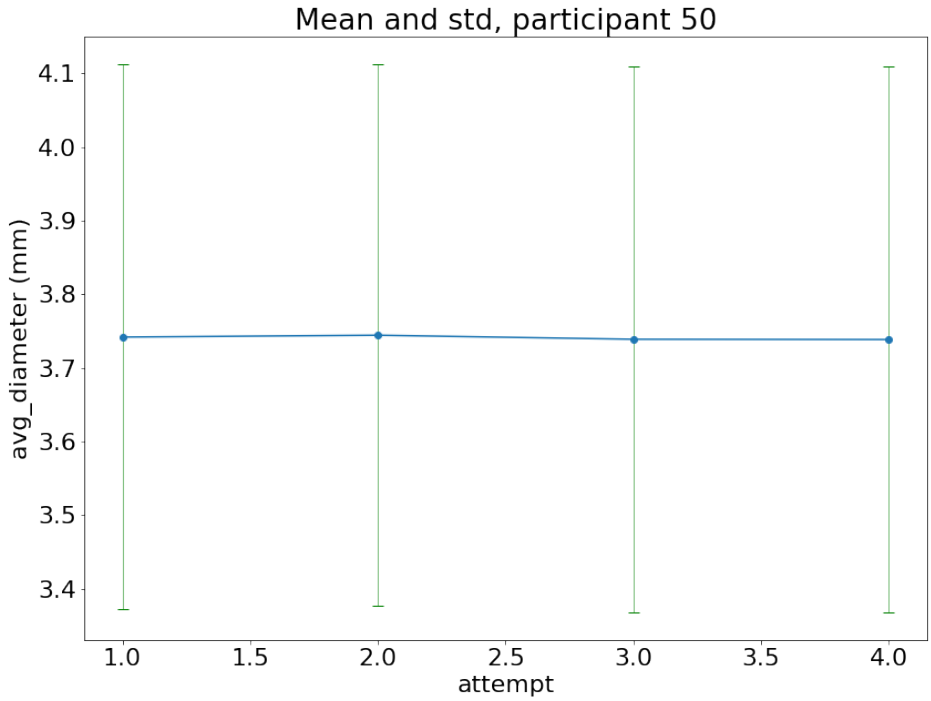

Mean and std, participant 40
avg_diameter (mm)
4.6
4.4
4.2
4.0
3.8
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt

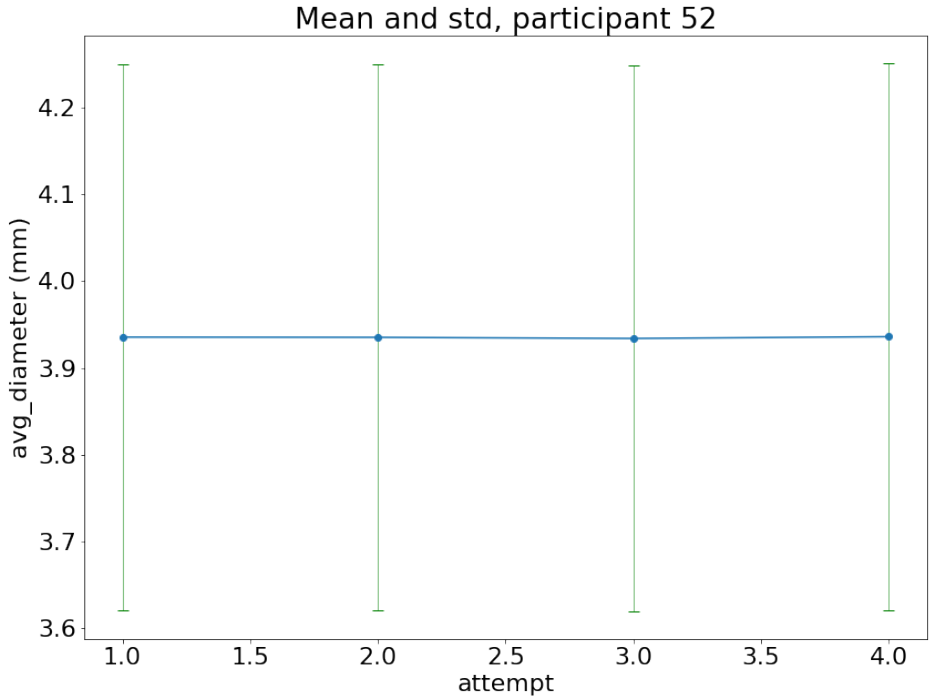

Mean and std, participant 42
avg_diameter (mm)
4.6
4.5
4.4
4.3
4.2
4.1
4.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt

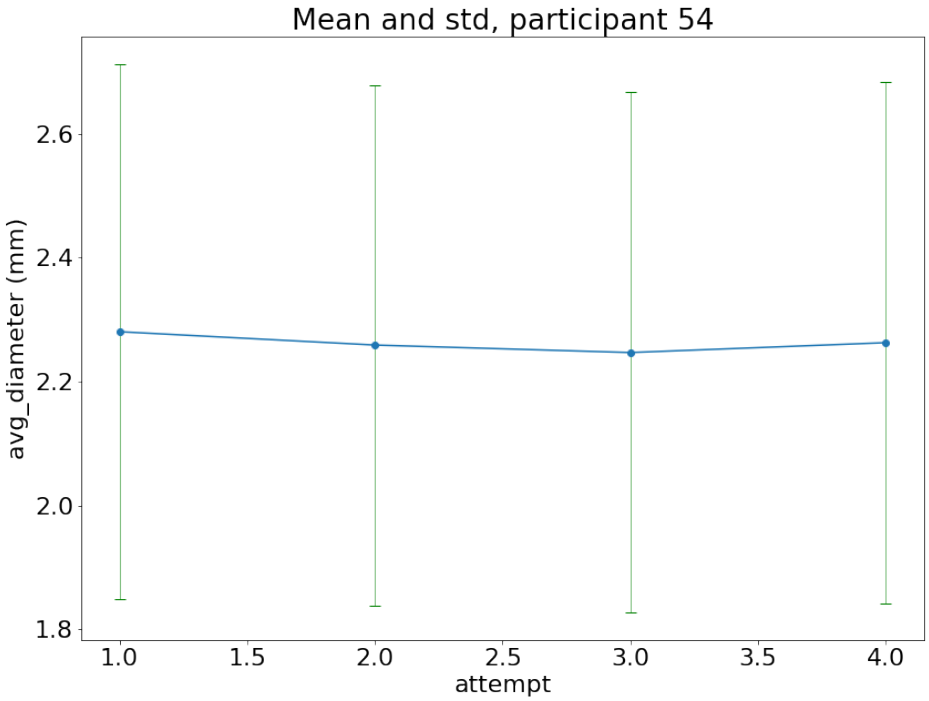

Mean and std, participant 44
avg_diameter (mm)
3.2
3.1
3.0
2.9
2.8
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt

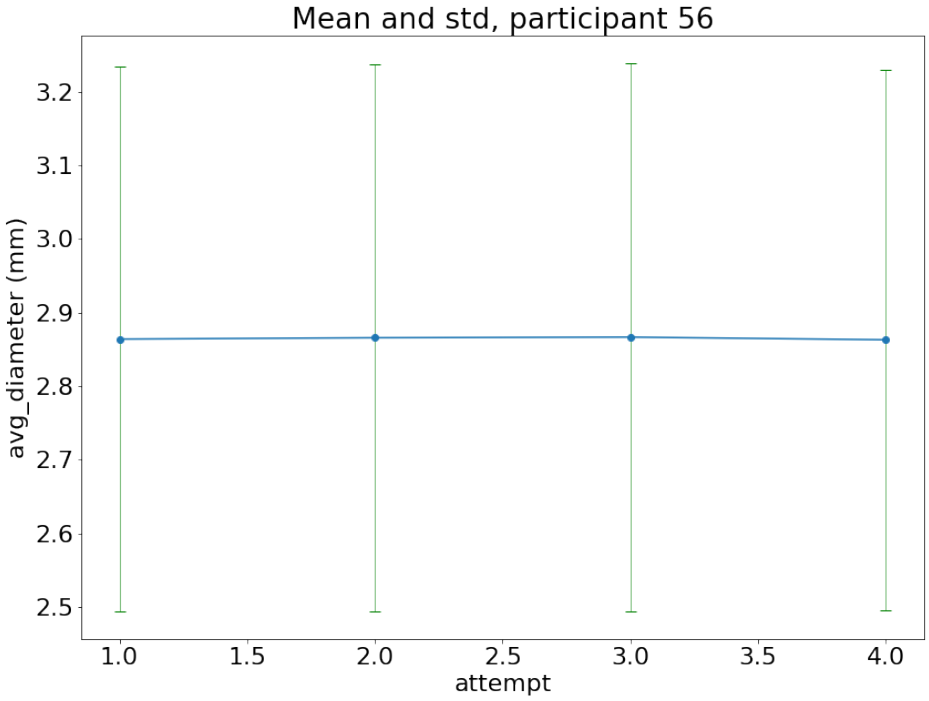

Mean and std, participant 46
avg_diameter (mm)
3.05
3.00
2.95
2.90
2.85
2.80
2.75
2.70
2.65
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt

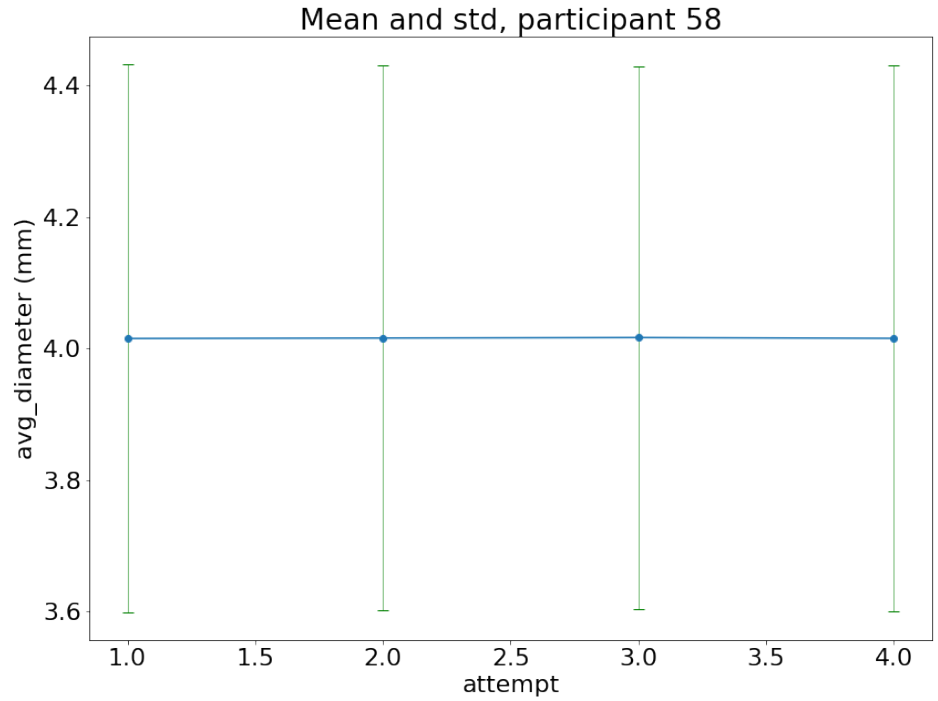

Mean and std, participant 48
avg_diameter (mm)
3.2
3.1
3.0
2.9
2.8
2.7
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt

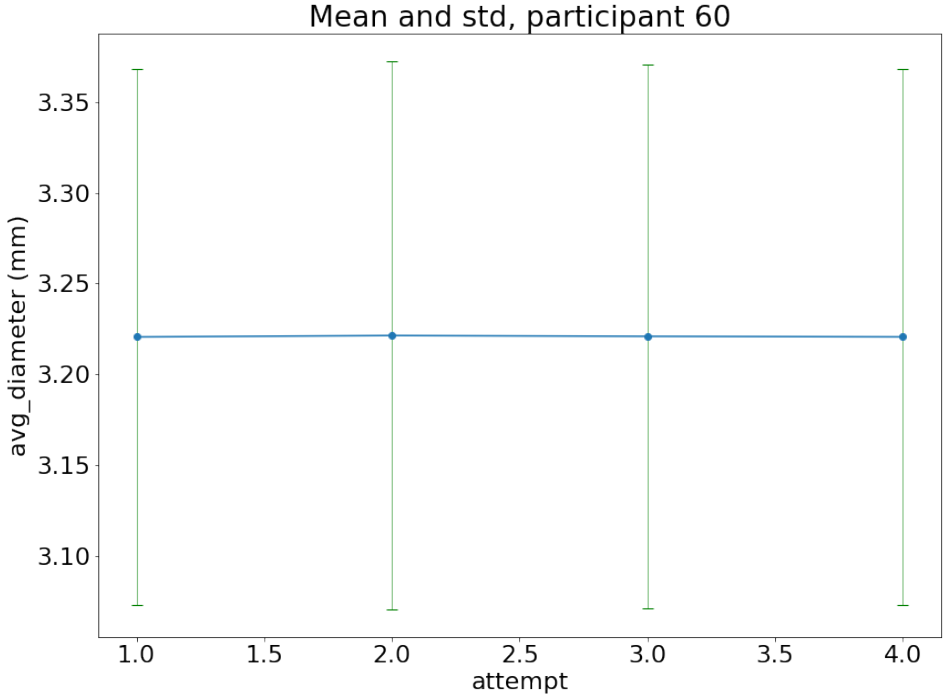

Mean and std, participant 50
avg_diameter (mm)
4.1
4.0
3.9
3.8
3.7
3.6
3.5
3.4
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt


Mean and std, participant 52
avg_diameter (mm)
4.2
4.1
4.0
3.9
3.8
3.7
3.6
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt


Mean and std, participant 54
avg_diameter (mm)
2.6
2.4
2.2
2.0
1.8
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt


Mean and std, participant 56
avg_diameter (mm)
3.2
3.1
3.0
2.9
2.8
2.7
2.6
2.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt


Mean and std, participant 58
avg_diameter (mm)
4.4
4.2
4.0
3.8
3.6
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt


Mean and std, participant 60
avg_diameter (mm)
3.35
3.30
3.25
3.20
3.15
3.10
1.0
1.5
2.0
2.5
3.0
3.5
4.0
attempt

Appendix C

Empathy score questionnaire

Questionnaires:

Furthermore the data is collected through the help of questionnaire, that is QCAE (Questionnaire of cognitive and affective empathy). Questionnaire consists of 31 questions, and we have added 9 more questions which is more related to our research. The questionnaire is mentioned below.

The questions are:

1. I sometimes find it difficult to see things from the ‘other guy’s’ point of view.
2. I am usually objective when I watch a film or play, and I don’t often get completely caught up in it.
3. I try to look at everybody’s side of a disagreement before I make a decision.
4. I sometimes try to understand my friends better by imagining how things look from their perspective.
5. When I am upset at someone, I usually try to ‘put myself in his shoes' for a while.
6. Before criticising somebody, I try to imagine how I would feel if I was in their place.
7. I often get emotionally involved with my friends’ problems.
8. I am inclined to get nervous when others around me seem to be nervous.
9. People I am with have a strong influence on my mood.
10. It affects me very much when one of my friends seems upset.
11. I often get deeply involved with the feelings of a character in a film, play or novel.
12. I get very upset when I see someone cry.
13. I am happy when I am with a cheerful group and sad when the others are glum. It worries me when others are worrying and panicky.
14. I can easily tell if someone else wants to enter a conversation.
15. I can pick up quickly if someone says one thing but means another.
16. It is hard for me to see why some things upset people so much.
17. I find it easy to put myself in somebody else's shoes.
18. I am good at predicting how someone will feel.

19. I am quick to spot when someone in a group is feeling awkward or uncomfortable.
20. Other people tell me I am good at understanding how they are feeling and what they are thinking.
21. I can easily tell if someone else is interested or bored with what I am saying.
22. Friends talk to me about their problems as they say that I am very understanding.
23. I can sense if I am intruding, even if the other person does not tell me.
24. I can easily work out what another person might want to talk about
25. I can tell if someone is masking their true emotion.
26. I am good at predicting what someone will do.
27. I can usually appreciate the other person's viewpoint, even if I do not agree with it.
28. I usually stay emotionally detached when watching a film.
29. I always try to consider the other fellow's feelings before I do
30. something.
31. Before I do something I try to consider how my friends will react to it.
32. I feel frustrated when I cannot communicate my ideas.
33. I complete the sentences of other people when they find it difficult to communicate their ideas to help them.
34. If someone is suffering from stutter, I prefer speaking rather than texting.
35. I try to be patient when people speak with me very slowly.
36. In a situation, when there are handicapped students in a classroom, the pace of the class should be adapted to the handicapped student.
37. In a situation, when there are only one handicapped student in a classroom, the pace of the class should be adapted to him/her independently of how many students are in the class.
38. I feel impatient when people communicate slowly, so I complete other people’s sentences to make the communication more efficient.
39. I believe that the handicapped students should catch the pace of the group in a classroom to not delay the progress of other students.
40. I feel left out (or excluded) when I cannot participate in a conversation.

Appendix D

Link to GitHub code used in analysis of the data sets

Link to GitHub: https://github.com/martadubas/eye-tracker-analysis

Bibliography

[1] Samip Bhurtel, Pedro G. Lind and Gustavo B. Moreno e Mello. 'For a New Protocol to Promote Empathy Towards Users of Communication Technologies'. In: *HCI International 2021 - Late Breaking Posters*. Ed. by Constantine Stephanidis, Margherita Antona and Stavroula Ntoa. Cham: Springer International Publishing, 2021, pp. 3–10.

[2] Samuel Richardson. *A collection of the moral and instructive sentiments, maxims, cautions, and reflexions, contained in the histories of Pamela, Clarissa, and Sir Charles Grandison*. London, England: Printed for S. Richardson; et al., 1755. URL: http://name.umdl.umich.edu/004835423.0001.000.

[3] Kenneth Holmqvist et al. *Eye Tracking : A comprehensive guide to methods and measures*. 1st ed. Oxford: Oxford University Press, 2015.

[4] Donell J. Creel. 'Handbook of Clinical Neurology, Vol. 160'. In: ed. by Patrick Chauvel Kerry H. Levin. 3rd ed. Amsterdam: Elsevier, 2019. Chap. Chapter 33: The electrooculogram.

[5] Yasuo Terao, Hideki Fukuda and Okihide Hikosaka. 'What do eye movements tell us about patients with neurological disorders? - An introduction to saccade recording in the clinical setting'. In: *Proceedings of the Japan Academy. Series B, Physical and biological sciences* 93.10 (2017), pp. 772–801. DOI: https://doi.org/10.2183/pjab.93.049.

[6] Rami N. Khushaba et al. 'Consumer neuroscience: Assessing the brain response to marketing stimuli using electroencephalogram (EEG) and eye tracking'. In: *Expert Systems with Applications* 40.9 (2013), pp. 3803–3812. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2012.12.095. URL: https://www.sciencedirect.com/science/article/pii/S0957417412013371.

[7] Jan Theeuwes, Artem Belopolsky and Christian N.L. Olivers. 'Interactions between working memory, attention and eye movements'. In: *Acta Psychologica* 132.2 (2009). Spatial working memory and imagery: From eye movements to grounded cognition, pp. 106–114. ISSN: 0001-6918. DOI: https://doi.org/10.1016/j.actpsy.2009.01.005. URL: https://www.sciencedirect.com/science/article/pii/S0001691809000158.

[8] Jia Zheng Lim, James Mountstephens and Jason Teo. 'Emotion Recognition Using Eye-Tracking: Taxonomy, Review and Current Challenges'. In: *Sensors* 20.8 (2020). ISSN: 1424-8220. URL: https://www.mdpi.com/1424-8220/20/8/2384.

[9] Helen Riess MD and Liz Neporent. *The Empathy Effect: Seven Neuroscience-Based Keys for Transforming the Way We Live, Love, Work, and Connect Across Differences*. 1st ed. Louisville, Colorado: Sounds True, 2018.

[10] Susan Lanzoni. *Empathy: A History*. 1st ed. London: Yale University Press, 2018.

[11] Rita Charon. 'Narrative Medicine: A Model for Empathy, Reflection, Profession, and Trust'. In: *JAMA* 286.15 (Oct. 2001), pp. 1897–1902. ISSN: 0098-7484. DOI: 10.1001/jama.286.15.1897. eprint: https://jamanetwork.com/journals/jama/articlepdf/194300/jrp10002.pdf. URL: https://doi.org/10.1001/jama.286.15.1897.

[12] Helen Riess. 'Empathy can be taught and learned with evidence-based education'. In: *Emergency Medicine Journal* (2021). ISSN: 1472-0205. DOI: 10.1136/emermed-2021-212078. eprint: https://emj.bmj.com/content/early/2021/12/20/emermed-2021-212078.full.pdf. URL: https://emj.bmj.com/content/early/2021/12/20/emermed-2021-212078.

[13] Hanneke ter Beest, Marlies van Bemmel and Marian Adriaansen. 'Nursing student as patient: experiential learning in a hospital simulation to improve empathy of nursing students'. In: *Scandinavian Journal of Caring Sciences* 32.4 (2018), pp. 1390–1397. DOI: https://doi.org/10.1111/scs.12584. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/scs.12584. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/scs.12584.

[14] Bernhard Hommel et al. 'No one knows what attention is'. In: *Attention, Perception, & Psychophysics* 81 (Sept. 2019). DOI: 10.3758/s13414-019-01846-w.

[15] Nicholas J. Wade and Benjamin W. Tatler. 'Informing contemporary research'. eng. In: The Moving Tablet of the Eye. Oxford: Oxford University Press, 2005. ISBN: 9780198566175. DOI: 10.1093/acprof:oso/9780198566175.003.0001. URL: https://doi.org/10.1093/acprof:oso/9780198566175.003.0001.

[16] Andrew T. Duchowski. *Eye Tracking Methodology*. 3rd ed. New York City: Springer International Publishing, 2017.

[17] Roy Hessels et al. 'Is the eye-movement field confused about fixations and saccades? A survey among 124 researchers'. In: *Royal Society Open Science* 5 (Aug. 2018), p. 180502. DOI: 10.1098/rsos.180502.

[18] Bruno Laeng and Dag Alnaes. 'Pupillometry'. In: *Eye Movement Research: An Introduction to its Scientific Foundations and Applications*. Ed. by Christoph Klein and Ulrich Ettinger. Cham: Springer International Publishing, 2019. ISBN: 978-3-030-20085-5. DOI: 10.1007/978-3-030-20085-5_11. URL: https://doi.org/10.1007/978-3-030-20085-5_11.

[19] Inmaculada Tomeo-Reyes. 'Robust Iris Recognition using Decision Fusion and Degradation Modelling'. PhD thesis. Sept. 2015.

[20] RH Spector. 'The Pupils'. In: *Clinical Methods: The History, Physical, and Laboratory Examinations. 3rd edition*. Ed. by Walker HK, Hall WD and Hurst JW. Boston: Butterworths, 1990. Chap. 58. URL: https://www.ncbi.nlm.nih.gov/books/NBK381/.

[21] S. B. Hutton. 'Eye Tracking Methodology'. In: *Eye Movement Research: An Introduction to its Scientific Foundations and Applications*. Ed. by Christoph Klein and Ulrich Ettinger. Cham: Springer International Publishing, 2019. ISBN: 978-3-030-20085-5. DOI: 10.1007/978-3-030-20085-5_11. URL: https://doi.org/10.1007/978-3-030-20085-5_8.

[22] Cristian Postelnicu and Doru Talaba. 'P300-Based Brain-Neuronal Computer Interaction for Spelling Applications'. In: *IEEE transactions on bio-medical engineering* 60 (Nov. 2012). DOI: 10.1109/TBME.2012.2228645.

[23] Wilfredo Palma. *Time series analysis*. eng. Wiley series in probability and statistics. Hoboken, New Jersey: Wiley, 2016. ISBN: 1-118-63434-9.

[24] Tyler Smith and Sara Johnson. 'Research and Statistics: Distribution, Variability, and Statistical Significance'. In: *Pediatrics in review / American Academy of Pediatrics* 31 (Oct. 2010), pp. 431–2. DOI: 10.1542/pir.31-10-431.

[25] Jim Frost. *Hypothesis testing: An intuitive guide for making data driven decisions*. Statistics By Jim Publishing, 2020. ISBN: 978-1735431154.

[26] Peter Goos and David Meintrup. *Statistics with JMP: Hypothesis Tests, ANOVA and Regression*. Chichester, West Sussex: John Wiley & Sons, Incorporated, 2016. ISBN: 1-119-09704-5.

[27] J. Neyman and E. S. Pearson. 'On the Problem of the Most Efficient Tests of Statistical Hypotheses'. In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 231 (1933). Full publication date: 1933, pp. 289–337. ISSN: 02643952. URL: http://www.jstor.org/stable/91247.

[28] Jacob Kröger, Otto Lutz and Florian Müller. 'What Does Your Gaze Reveal About You? On the Privacy Implications of Eye Tracking'. In: Mar. 2020, pp. 226–241. ISBN: 978-3-030-42504-3. DOI: 10.1007/978-3-030-42504-3_15.

[29] William James. *The principles of psychology*. Henry Holt and Co., 1890. URL: https://doi.org/10.1037/10538-000.

[30] William Ocasio. 'Attention to Attention'. In: *Organization Science* 22.5 (2011), pp. 1286–1296. DOI: 10.1287/orsc.1100.0602. URL: https://doi.org/10.1287/orsc.1100.0602.

[31] Alois Ferscha. 'Attention, Please!' In: *Pervasive Computing, IEEE* 13 (Jan. 2014), pp. 48–54. DOI: 10.1109/MPRV.2014.3.

[32] Antonia Vehlen et al. 'Evaluation of an eye tracking setup for studying visual attention in face-to-face conversations'. In: *Scientific Reports* 11.1 (Jan. 2021), p. 2661. ISSN: 2045-2322. DOI: 10.1038/s41598-021-81987-x. URL: https://doi.org/10.1038/s41598-021-81987-x.

[33] Christian Valuch et al. 'Using eye tracking to test for individual differences in attention to attractive faces'. In: *Frontiers in Psychology* 6 (2015). ISSN: 1664-1078. DOI: 10.3389/fpsyg.2015.00042. URL: https://www.frontiersin.org/article/10.3389/fpsyg.2015.00042.

[34] Ruud L. van den Brink, Peter R. Murphy and Sander Nieuwenhuis. 'Pupil Diameter Tracks Lapses of Attention'. In: *PLOS ONE* 11.10 (Oct. 2016), pp. 1–16. DOI: 10.1371/journal.pone.0165274. URL: https://doi.org/10.1371/journal.pone.0165274.

[35] Jac N. Airdrie et al. 'Facial Emotion Recognition and Eye Gaze in Attention-Deficit/Hyperactivity Disorder With and Without Comorbid Conduct Disorder'. In: *Journal of the American Academy of Child & Adolescent Psychiatry* 57.8 (2018), pp. 561–570. ISSN: 0890-8567. DOI: https://doi.org/10.1016/j.jaac.2018.04.016. URL: https://www.sciencedirect.com/science/article/pii/S089085671830282X.

[36] G. Wainstein et al. 'Pupil Size Tracks Attentional Performance In Attention-Deficit/Hyperactivity Disorder'. In: *Scientific Reports* 7.1 (Aug. 2017), p. 8228. ISSN: 2045-2322. DOI: 10.1038/s41598-017-08246-w. URL: https://doi.org/10.1038/s41598-017-08246-w.

[37] Michelle M. Neumann, Camillia Acosta and David L. Neumann. 'Young Children's Visual Attention to Environmental Print as Measured by Eye Tracker Analysis'. In: *Reading Research Quarterly* 49.2 (2014), pp. 157–167. DOI: https://doi.org/10.1002/rrq.66. eprint: https://ila.onlinelibrary.wiley.com/doi/pdf/10.1002/rrq.66. URL: https://ila.onlinelibrary.wiley.com/doi/abs/10.1002/rrq.66.

[38] Belgüzar Nilay Türkan et al. 'Comparison of change detection performance and visual search patterns among children with/without ADHD: Evidence from eye movements'. In: *Research in Developmental Disabilities* 49-50 (2016), pp. 205–215. ISSN: 0891-4222. DOI: https://doi.org/10.1016/j.ridd.2015.12.002. URL: https://www.sciencedirect.com/science/article/pii/S089142221530024X.

[39] Jonathan Smallwood et al. 'Pupillometric Evidence for the Decoupling of Attention from Perceptual Input during Offline Thought'. In: *PLOS ONE* 6.3 (Mar. 2011), pp. 1–8. DOI: 10.1371/journal.pone.0018298. URL: https://doi.org/10.1371/journal.pone.0018298.

[40] J. Beatty. 'Task-evoked pupillary responses, processing load, and the structure of processing resources'. In: *sychological Bulletin* 91.2 (1982), pp. 276–292. DOI: https://doi.org/10.1037/0033-2909.91.2.276.

[41] J. Beatty and B. Lucero-Wagoner. 'The pupillary system'. In: *The Handbook of Psychophysiology*. Ed. by Berntson G Caccioppo J Tassinary LG. Cambridge University Press, 2000, pp. 142–162. ISBN: 978-0-511-27907-2.

[42] Greg J Siegle et al. 'Use of concurrent pupil dilation assessment to inform interpretation and analysis of fMRI data'. In: *NeuroImage* 20.1 (2003), pp. 114–124. ISSN: 1053-8119. DOI: https://doi.org/10.1016/S1053-8119(03)00298-2. URL: https://www.sciencedirect.com/science/article/pii/S1053811903002982.

[43] Olivia E. Kang, Katherine E. Huffer and Thalia P. Wheatley. 'Pupil Dilation Dynamics Track Attention to High-Level Information'. In: *PLOS ONE* 9.8 (Aug. 2014), pp. 1–6. DOI: 10.1371/journal.pone.0102463. URL: https://doi.org/10.1371/journal.pone.0102463.

[44] Vidas Raudonis et al. 'Evaluation of Human Emotion from Eye Motions'. In: *International Journal of Advanced Computer Science and Applications* 4.8 (2013). DOI: 10.14569/IJACSA.2013.040812. URL: http://dx.doi.org/10.14569/IJACSA.2013.040812.

[45] Claudio Aracena et al. 'Neural Networks for Emotion Recognition Based on Eye Tracking Data'. In: *2015 IEEE International Conference on Systems, Man, and Cybernetics*. 2015, pp. 2632–2637. DOI: 10.1109/SMC.2015.460.

[46] Vicky Tsang. 'Eye-tracking study on facial emotion recognition tasks in individuals with high-functioning autism spectrum disorders'. In: *Autism : the international journal of research and practice* 22.2 (2018). DOI: https://doi.org/10.1177/1362361316667830. URL: https://pubmed.ncbi.nlm.nih.gov/29490486/.

[47] Rajakumari and S. Tamil Selvi. 'HCI and Eye Tracking : Emotion Recognition Using Hidden Markov Model'. In: 2015.

[48] Ping Liu et al. 'Effect of empathy trait on attention to positive emotional stimuli: evidence from eye movements'. In: *Current Psychology* (Apr. 2020). ISSN: 1936-4733. DOI: 10.1007/s12144-020-00723-2. URL: https://doi.org/10.1007/s12144-020-00723-2.

[49] Siri Leknes et al. 'Oxytocin enhances pupil dilation and sensitivity to 'hidden' emotional expressions'. In: *Social Cognitive and Affective Neuroscience* 8.7 (May 2012), pp. 741–749. ISSN: 1749-5016. DOI: 10.1093/scan/nss062. eprint: https://academic.oup.com/scan/article-pdf/8/7/741/27107589/nss062.pdf. URL: https://doi.org/10.1093/scan/nss062.

[50] Renate Reniers et al. 'The QCAE: a Questionnaire of Cognitive and Affective Empathy'. In: *Journal of personality assessment* 93 (Jan. 2011), pp. 84–95. DOI: 10.1080/00223891.2010.528484.

[51] Tobii AB. *Tobii Studio User Manual*. 2016. URL: https://www.tobiipro.com/siteassets/tobii-pro/user-manuals/tobii-pro-studio-user-manual.pdf (visited on 08/04/2022).

[52] Jing Zhang et al. 'Significant Measures of Gaze and Pupil Movement for Evaluating Empathy between Viewers and Digital Content'. In: *Sensors* 22.5 (2022). ISSN: 1424-8220. DOI: 10.3390/s22051700. URL: https://www.mdpi.com/1424-8220/22/5/1700.

[53] Siddhartha Joshi. 'Pupillometry: Arousal State or State of Mind?' In: *Curr Biol* 31.1 (Jan. 2021), R32–R34.

[54] Elena S. Novitskaya et al. 'Effects of some ophthalmic medications on pupil size: a literature review'. In: *Canadian Journal of Ophthalmology* 44.2 (2009), pp. 193–197. ISSN: 0008-4182. DOI: https://doi.org/10.3129/i09-003. URL: https://www.sciencedirect.com/science/article/pii/S0008418209801580.

[55] Stephanie D. Preston and Frans B. M. de Waal. 'Empathy: Its ultimate and proximate bases'. In: *Behavioral and Brain Sciences* 25.1 (2002), pp. 1–20. DOI: 10.1017/S0140525X02000018.

[56] Wai Hin Wan and Angie Lam. 'The Effectiveness of Virtual Reality-Based Simulation in Health Professions Education Relating to Mental Illness: A Literature Review'. In: *Health* 11 (Jan. 2019), pp. 646–660. DOI: 10.4236/health.2019.116054.

[57] Elizabeth Dyer, Barbara Swartzlander and Marilyn Gugliucci. 'Using virtual reality in medical education to teach empathy'. In: *Journal of the Medical Library Association : JMLA* 106 (Oct. 2018), pp. 498–500. DOI: 10.5195/jmla.2018.518.

[58] David Passig. 'The Impact of Virtual Reality on Educator's Awareness of the Cognitive Experiences of Pupils with Dyslexia'. In: *International Federation for Information Processing Digital Library; Joint Open and Working IFIP Conference ICT and Learning for the Net Generation;* 113 (Jan. 2011).

[59] Linda Jutten, Ruth Mark and Margriet Maria Sitskoorn. 'Can the Mixed Virtual Reality Simulator Into D'mentia Enhance Empathy and Understanding and Decrease Burden in Informal Dementia Caregivers?' In: *Dementia and Geriatric Cognitive Disorders Extra* 8 (Dec. 2018), pp. 453–466. DOI: 10.1159/000494660.