**Maria Camila de Avelino Geraldo Dias**

Bachelor of Science in Biomedical Engineering

# Prediction of uncertainty events using human-computer interaction

Dissertation submitted in partial fulfillment of the requirements for the degree of

Master of Science in
**Biomedical Engineering**

Adviser: Prof. Doctor Hugo Gamboa, Auxiliar Professor, NOVA University of Lisbon

Co-adviser: Prof. Doctor Marcus Cheetham, Senior Research Associate, Department of Internal Medicine, University Hospital of Zurich

Examination Committee

Chairperson: Prof.ª Dr.ª Carla Maria Quintão Pereira
Raporteur: Prof.ª Dr.ª Cláudia Regina Pereira Quaresma
Member: Prof. Dr. Hugo Filipe Silveira Gamboa

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

**November, 2018**

**Prediction of uncertainty events using human-computer interaction**



This document was created using the (pdf)LaTeX processor, based in the "novathesis" template[1], developed at the Dep. Informática of FCT-NOVA [2].
[1] https://github.com/joaomlourenco/novathesis [2] http://www.di.fct.unl.pt

# Acknowledgements

I want to thank all the people that stood beside me during these memorable years of study, fun, love and personal and professional growing. I feel very lucky and proud of the choice I made five years ago. Particularly, I want to thank who supported me in the development of this dissertation.

Firstly, I want to thank Professor Hugo Gamboa for giving me the opportunity of developing this really fascinating project, which aroused new interests in me.

I am thankful to Professor Marcus Cheetham as well, who gave several interesting and important ideas and whose psychological knowledge was crucial for the understanding of some concepts. I would also like to thank the Department of Internal Medicine, University Hospital of Zurich, for providing the data.

I am extremely grateful to Cátia Cepeda, who was always available to help me, and to discuss new ideas and results. Without her, I would not have learned as much as I did during this thesis development. Thanks to everyone from Biomedical Engineering Laboratory of LIBPhys (Laboratory for Instrumentation, Biomedical Engineering and Radiation Physics) for welcoming me there.

To all of my great friends, a big thank you. With you, these five years of work were much easier and full of companionship and amusement! To Miguel, for always being there for me.

A huge thanks to my parents, to my sister Mariana and my brother Afonso, for always taking care of me with a lot of love.

# Abstract


The practice of medicine is characterized by complex situations that evoke uncertainty. Uncertainty has implications for the quality and costs of health care, thus emphasizing the importance of identifying its the main causes.

Uncertainty can be manifested through human behaviour. Accordingly, in this dissertation, a machine learning model that detects events of uncertainty based on mouse cursor movements was created. To do so, 79 participants answered an online survey while the mouse data was being tracked. This data was used to extract meaningful features that allowed model testing and training after a feature selection stage. With the implementation of a Logistic Regression, and applying a k-fold cross-validation method, the model achieved an estimated performance of 81%.

It was found that, during moments of uncertainty, the number of horizontal direction inversions increases and the mouse cursor travels higher distances. Moreover, items that evoke uncertainty are associated to longer interaction times and a higher number of visits.

Subsequently, the model was applied to a medical decision making task performed by 8 physicians, in order to understand whether it might be applied in different contexts or not. The results were consistent with the task design.

To better understand the nature of uncertainty, its relationship with personality was explored. Regarding the clinical task, it was found a slight tendency of uncertainty to increase with Neuroticism.

In the future, the created model may be used to help physicians understand their main difficulties.



**Keywords:** Uncertainty, medical decision-making, human-computer interaction, signal processing, machine learning, Five Factor Model

# Resumo

A área de medicina envolve situações complexas que despertam incerteza. As complicações nos cuidados de saúde associados à incerteza e que se refletem na qualidade e custos dos mesmos despertam a necessidade de identificar as suas principais causas.

A incerteza pode manifestar-se através do comportamento humano. Consequentemente, no decorrer desta tese, foi criado um modelo de aprendizagem automática que deteta eventos de incerteza com base em informação extraída a partir de movimentos do cursor do rato. Para tal, 79 participantes responderam a um questionário *online*, durante o qual foram adquiridos os dados do cursor. Estes foram utilizados com o objetivo de extrair características significativas que, após passarem por um processo de seleção, foram utilizadas para testar e treinar o modelo. Seguidamente, com a implementação de uma regressão logística e aplicando um método de validação cruzada, o modelo alcançou um valor estimado de desempenho de 81%.

Foi possível concluir que, durante momentos de incerteza, o número de inversões de sentido no eixo horizontal aumenta e o cursor do rato percorre distâncias superiores à média. Ademais, existe uma tendência para visitar repetidamente, bem como permanecer maiores intervalos de tempo, em elementos do texto que incitam incerteza.

Posteriormente, o modelo foi aplicado a uma tarefa de tomada de decisão médica executada por 8 médicos, com o objetivo de analisar se o modelo é adequado a diferentes contextos. Os resultados obtidos foram satisfatórios.

De forma a compreender a incerteza no seu todo, foi explorada a sua relação com a personalidade. Relativamente à tarefa de tomada de decisão médica, foi verificado um ligeiro aumento da incerteza com o aumento do Neuroticismo.

No futuro, o modelo criado poderá vir a ser utilizado por médicos interessados em depreender as suas maiores dificuldades.

**Palavras-chave:** Incerteza, tomada de decisão médica, interação humano-computador, processamento de sinal, aprendizagem automática, modelo dos cinco fatores

# Contents

# List of Figures

# List of Tables

# Acronyms

| | |
|---|---|
| AJAX | Asynchronous JavaScript and XML. |
| DDI | Disease-disease, Drug-disease and Drug-drug Interactions. |
| FFM | Five Factor Model. |
| FN | False Negatives. |
| FP | False Positives. |
| HCI | Human-Computer Interaction. |
| IGT | Iowa Gambling Task. |
| IP | Internet Protocol. |
| LED | Light Emitting Diode. |
| MISI | Multimorbidity Interaction Severity Index. |
| NEO-FFI | NEO Five-Factor Inventory. |
| PHP | PHP: Hypertext Preprocessor. |
| TN | True Negatives. |
| TP | True Positives. |

# Chapter 1

# Introduction

## 1.1 Motivation

Occasionally, everyone faces hard decisions, like choosing a career option or a place to live. However, some people experience more problems making choices than others. Commonly, indecisive people are more anxious, have lower self-esteem, procrastinate [1] and regret more about decisions made [2]. They tend to be less efficient while making decisions in the workplace as well [1]. Moreover, indecisiveness is a symptom of Obsessive Compulsive Disorder [1, 3]. For all these reasons, it is important to measure indecisiveness and provide counseling if necessary.

Ongoing mental processes, like uncertainty, can be perceived in simple and common tasks, such as during the interaction with a computer. This can be tracked using eye gaze or a mouse cursor. In this dissertation, it was created a machine learning model that identifies moments of uncertainty based on mouse cursor movements. Assessing the percentage of uncertainty instances, indecisiveness can be estimated, providing additional information to the standard questionnaires.

Moreover, the relationship between uncertainty and personality was analysed, in order to examine which factors and personal characteristics influence the former.

When a certain question is associated to uncertainty across several different people, it may not be related to personal attributes but rather to the question's structure or content. Therefore, the developed model may also be used to identify confusing items in a survey, or help teachers understand their students' difficulties while answering to a work sheet, for example. Probably, in the future, it could be provided real-time help in difficult questions.

The created model was used to analyse an occupation highly linked to decision making - medicine. In this study, several physicians evaluated complex medical cases and uncertainty was assessed. The physician's knowledge about the cases may be imperfect or incomplete - which in real life can lead to errors -, and this analysis could be helpful for the understanding of their main difficulties.

This dissertation was developed in Faculdade de Ciências e Tecnologia - Universidade Nova de Lisboa in collaboration with the Department of Internal Medicine, University Hospital of Zurich, where the data was acquired.

## 1.2 Objectives

The main objective of this dissertation was to create a model that identifies moments of uncertainty during the fullfiment of a questionnaire. Eye gaze is the indicator of human visual attention [4] and, accordingly, it would be expected to track eye movements in order to access those events. Nevertheless, mouse tracking has numerous advantages over eye tracking and there is evidence suggesting that eye and mouse trajectories are similar [4, 5]. On that account, eye and mouse movements were compared to understand whether to use mouse cursor data to construct the model or not. To compare them, the Iowa Gambling Task (IGT), which is a card game that simulates real-life decision making, was used. Since the results manifested a fair correlation between mouse and eye movements, the mouse tracking data was used.

The uncertainty model was constructed by extracting features from mouse cursor data of an online survey. Each question of the survey could be classified as an uncertainty instance or not. Afterwards, the relation between the rate of uncertainty occurrences and personality was explored. Personality, on its turn, was measured through the results of a questionnaire that was also used to track the mouse cursor data.

Lastly, the model was applied to a clinical context. Three hypothetical cases of multimorbidity - which concerns the existence of multiple concurrent acute or chronic diseases within an individual - with different degrees of complexity were analysed by physicians. Therapeutic decision making for multimorbid patients is demanding due to the combined effects of pottentially harmful Disease-disease, Drug-disease and Drug-drug Interactions (DDI) - for example, the treatment for one condition may be contraindicated by the presence of other conditions or treatments [6]. The model predicted which DDI evoked uncertainty through mouse movements. Accordingly, it was explored if the amount of difficult DDI in each case was consistent with its complexity (i.e., if the less complex case had a lower amount of difficult DDI and vice-versa), in order to test its validity across different contexts. The association between uncertainty and personality was also assessed with this data.

Summing up, the objectives of this dissertation were:

- To compare eye and mouse cursor movements;
- To construct the mouse movement model for automatic detection of uncertainty events;

- To test the model validity in a clinical context;
- To find a relation between uncertainty and personality.

## 1.3 Thesis Overview

This dissertation is constituted by seven chapters and one appendix. In this chapter, the development of the thesis is justified through the motivation and its main objectives are presented. Chapter 2 introduces the theoretical concepts needed for the comprehension of the study and chapter 3 reports the related work. Chapter 4 describes the data acquisition, including the experiment conditions and the participants of the study and chapter 5 presents the tools and methodologies applied. The results are presented and discussed in chapter 6. The last chapter summarizes the outcomes of the work, describes its limitations and suggests some improvements to apply in the future. Finally, the appendix presents two articles written during this research work.

# Chapter 2

# Theoretical Concepts

In this chapter, the theoretical concepts needed for the understanding of the dissertation are explored. It introduces two important tools used to measure Human-Computer Interaction (HCI), eye and mouse tracking, that were compared in this dissertation. Subsequently, uncertainty and indecisiveness - concepts inherent to decision making - are described, since the main objective of this study was to construct a model that detects uncertainty moments. Lastly, it is given a definition of personality, and the most distinguished model of personality psychology - the Five Factor Model (FFM) - is presented, owing to the fact that a relation between uncertainty and FFM was investigated.

## 2.1 Human-Computer Interaction

In our daily lives, we are always surrounded by computer technology - computers, tablets, mobile phones, televisions, car navigation systems and even household appliances [7]. HCI studies how people interact with this technology [8]. One of its areas of work concerns the design, implementation and evaluation of interactive computer systems [8, 9, 10]. To attract users, the applications must be easy and pleasant to use, trustworthy and persuasive. Taking this into account, systems must be able to sense the users' needs and preferences and respond accordingly [7, 9].

Besides computer system design, the characteristics of the user also influence HCI. For example, many researchers acknowledge the impact of personality in HCI [11]. Taking this into account, HCI has been used in many psychology experiments.

Eye and mouse tracking are powerful and common tools to assess HCI [4]. In this study, they were compared in order to understand if eye gaze data may be replaced by mouse cursor data, since eye tracking data supplies the user's focus at any time [12] but mouse tracking has numerous advantageous over eye tracking, as it will be explained in the following sections.

### 2.1.1 Eye Tracking

#### 2.1.1.1 The Eye

The mammalian eye absorbs and converts the light into electrochemical impulses that are processed by the brain. Firstly, the light is refracted by the cornea, which is the membrane that covers the front of the eye, protecting it. Then, it contacts the iris (the colored part of the eye) and the pupil (the centre of the iris), which, together, regulate the amount of light that enters the eye. Afterwards, the light encounters the lens, that brings objects at various distances into focus with the help of auxiliary muscles. Subsequently, the light is projected onto the retina where the light is converted into electrical impulses. This impulses travel through the neurons of the retina into the optic nerve and, finally, reach the brain [13]. Figure 2.1 shows the basic structures of the vertebrate eye.

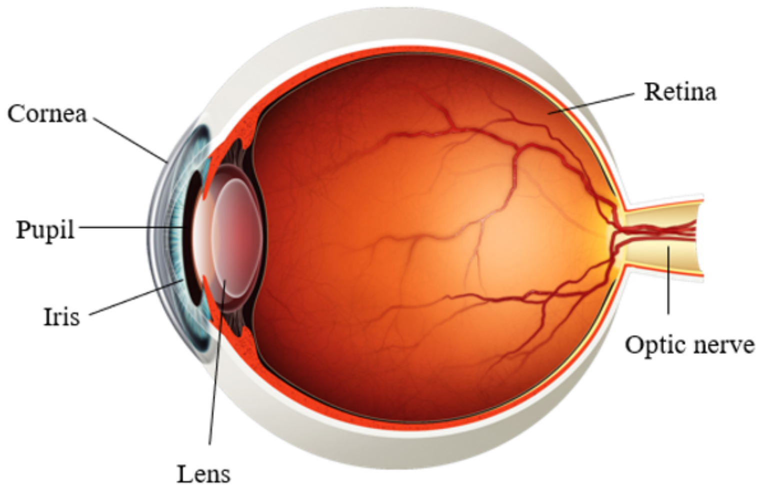



Figure 2.1: Schematic of a vertebrate eye. Adaptad from [14].

#### 2.1.1.2 Eye Tracking

Eye tracking is a sensor technology that records the positions where the user looks at and the eye sequential movements at any time [12, 15].

Nowadays, most eye tracking systems use the *pupil centre corneal reflection* technique. Infrared light from a light source is directed towards the eye, causing reflections that highlight the pupil and the location of the corneal reflection, which are captured by a camera. Afterwards, the image processing software identifies them and, with trigonometric calculations combined with geometrical features, the point where the user is looking at (gaze point [16]) is obtained. Infrared light is used to avoid distractions while the eyes are being tracked [15, 17]. Figure 2.2 shows how the corneal reflection changes according to the gaze point.

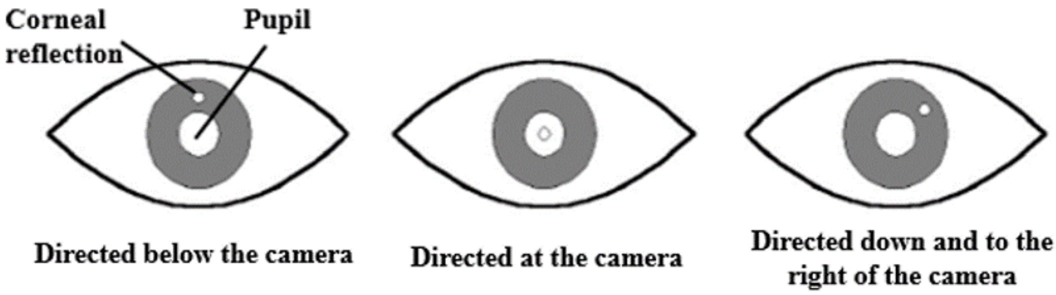



Figure 2.2: Corneal reflection position changes according to the gaze point. Adapted from [15].

The analysis of eye movements has been an area of interest for many years [18] with applications in many areas, such as medical diagnoses and psychological studies [19]. The eye gaze is the indicator of human visual attention and it gives insights into individuals' cognitive states [4, 19]. Nevertheless, an eye tracker is expensive, it requires calibration and, frequently, the results are not satisfactory due to equipment losses of calibration. Additionally, it is only used in studies where the individuals are physically present [4, 5].

### 2.1.2 Mouse Tracking

#### 2.1.2.1 The Mouse

To interact with a computer, a pointing device is needed to move the cursor on the computer monitor. The most commonly used is the mouse.

The mechanical mouse comprises a ball, that rolls according to the movements imposed by the mouse, and two rollers in contact with the ball. The rollers detect movement, one in the horizontal direction, and the other in the vertical direction [20].

Nowadays, the mechanical mouse is being replaced by optical and laser mouses due to problems of deterioration and dirt accumulation over time [20]. The optical mouse, which is illustrated in figure 2.3, is constituted by a Light Emitting Diode (LED) that illuminates the surface and a lens that images the surface of the mouse onto a camera. The mouse works by constantly comparing the images of the surface. The laser mouse is an optical mouse which has a laser diode as the light source which illumination is less diffuse and uniform, exposing surface details that would not be revealed with a LED [21].

A laptop is constituted by a touchpad, and moving a finger across a touchpad moves the cursor in the same direction. There are several other pointing devices, such as a trackpoint, a small point in the middle of the keyboard, and head-tracking devices, for example [22].

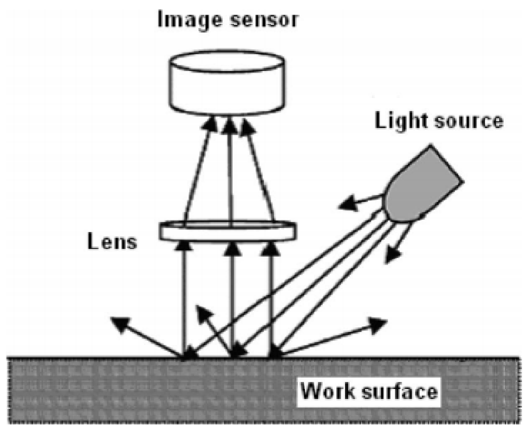



Figure 2.3: Optical mouse operation. From [21].

#### 2.1.2.2 Mouse Tracking

The cursor positions may be collected using a software, a method designated by computer mouse tracking. It is relatively recent and it can provide information about cognitive processes [23], since people tend to move their mouse cursor according to their focus of attention [24].

Several studies suggested that mouse cursor movements are related to eye movements [4, 5, 24] and mouse tracking is associated with numerous advantages over eye tracking. On the one hand, mouse cursor data can be acquired easily, without the presence of users and involving a great number of individuals simultaneously, increasing the amount of data. On the other, this technique does not have the eye tracking calibration problems [4, 5].

There is a wide variety of measures to analyse mouse trajectories, including temporal and spatial measures. The most common temporal features are velocity and acceleration [23, 25, 26]. The spatial variables include distance traveled [27, 28], angles of direction, where great angles correspond to directional shifts [25, 26, 27], curvature [25, 26] and straightness - the ratio between the Euclidean distance from the starting to the ending points and the total path distance -, which can give information about the attraction to an unanswered response, for example [25].

In this dissertation, the mouse tracking data acquired during decision making tasks was used to recognize moments of uncertainty.

## 2.2 Machine Learning

The world is overwhelmed with data. Nonetheless, there is an extensive gap between the generation and the understanding of potentially useful data. On that account, machine learning has become an extremely important field [29].

*Data mining* consists in the process of discovering patterns in data that can be used to predict new data. *Machine learning* provides the techniques to find and describe these patterns. Therefore, it supports problem solving through the analysis of information present in databases. For example, if a store owner wants to keep his costumers, a database of customer profiles and choices may be the solution. Patterns of behaviour can be analysed to predict who will remain loyal and who will not. Afterwards, the latter can be selected for special treatment [29].

In this dissertation, instances of uncertainty behaviour were collected and associated to several features from mouse tracking data. These instances were used to construct a mouse movement model for automatic detection of uncertainty.

## 2.3 Decision Making

Decision making consists in the capacity of choosing between different options based on their possible rewards and risks [30]. The degree of the outcomes' uncertainty varies and, therefore, there are different decision making methods. Nonetheless, all of them include some common stages, such as recognition of the problem, search of information, assessment of alternatives, selection of an alternative, implementation and feedback [31].

Decision-making contexts comprise decisions under *certainty*, where the outcome of each action is known, *risk*, where each action leads to one of a set of possible consequences and their probabilities are known, and, finally, *uncertainty*, where actions lead to a set of possible outcomes but their probabilities are unknown [32]. It is important to highlight that, in this dissertation, the term *uncertainty* is not related to the margin of error of a measurement, as it is applied in physics.

The theories about decisions made under risk propose that humans follow the probabilistic rules. However, several studies have put into evidence that some decisions are made based on an intuitive approach. The dual process theory suggests that human make both rational (analytical) and non-rational (intuitive and emotional) decisions. When the possible outcomes are uncertain, both systems act in collaboration [33].

Some individuals are less tolerant to decisions made under uncertainty, and they are known as indecisive [3].

### 2.3.1 Indecisiveness

Periodically, everyone faces difficult decisions. Nonetheless, some people manifest a tendency to experience problems in making decisions. Indecisiveness is defined as a chronic incapacity to make decisions across a wide variety of domains and situations [1, 2, 34].

Individuals with high indecisiveness avoid, postpone and worry about decisions [1, 2, 3], take a long time to decide, seek more information before deciding [1, 2], use less effective decisional strategies [1] and experience more choice dissatisfaction [2]. Indecisiveness has been related to high anxiety, low problem-solving abilities, low self-esteem, procrastination [1], perfectionism and obsessive-compulsive tendencies [1, 3].

## 2.4 Personality

Although there is no general definition, personality can be defined as "an individual's characteristic patterns of thought, emotion, and behavior, together with the psychological mechanisms - hidden or not - behind those patterns"[35]. These patterns are called *personality traits* [36, 37], and a cluster of related traits constitute a *personality dimension* [38].

There are plenty of personality theories. During decades, there was a considerable disagreement between the number and the names of personality dimensions. Nowadays, the most influential model of personality psychology is the *FFM*, due to its robust evidence across cultures [38, 39, 40].

### 2.4.1 Five Factor Model

The FFM provides an organization of personality traits in five basic dimensions, known as the *Big Five*: Neuroticism, Extraversion, Conscientiousness, Agreeableness, and Openness to Experience [38, 41]. Most of personality traits can be categorized into one of these factors [40].

Each factor can be represented by a scale. The greater the tendency of an individual to behave according to a specific dimension, the higher the score of its scale. Therefore, a person can be characterized by the scores obtained in the five factors [42]. Table 2.1 shows the dimensions' meanings and the traits associated to them, as well as some qualities related to people that score high and low in them.

This model organizes sets of variables that are related to each other and unrelated to other sets [41]. For example, people who are gregarious and assertive tend to be, also, enthusiastic and active, which are qualities related to Extraversion. Nevertheless, people who are gregarious and assertive may or may not be intellectually curious and imaginative. Therefore, those qualities relate to another dimension, Openness to Experience [40].

The most validated tools to measure the FFM's dimensions are questionnaires. The most used is the Revised NEO Personality Inventory, which measures all the traits that define each dimension. The NEO Five-Factor Inventory (NEO-FFI) is a brief version which only assess the five factors. Another questionnaire that is widely used is the Big Five Inventory [40], which is an even shorter version that also measures the five dimensions [43]. In this dissertation, the NEO-FFI was used. An example of a NEO-FFI item is “I am someone who is talkative”, which belongs to the Extraversion scale.

Table 2.1: Dimensions of personality according to FFM, their meanings, traits, and qualities related to people that score high and low in them. Based on [38, 44, 45].

| Dimension | Meaning | Traits | High Scores | Low Scores |
|---|---|---|---|---|
| Neuroticism | A tendency to experience dysphoric affect - sadness, hopelessness, guilt | Anxiety; Hostility; Depression; Self-Consciousness; Impulsiveness; Vulnerability | Irrational; Low self-esteem; Poor control of impulses; Ineffective coping; Pessimistic | Calm; Relaxed; Even-tempered; Unflappable; Stable |
| Extraversion | Preference for companionship and social stimulation | Warmth; Activity; Gregariousness; Assertiveness; Excitement Seeking; Positive Emotions | Cheerful; Dominant; Talkative; Sociable; Warm; Enthusiastic; Energetic; Optimistic | Introverted; Quiet; Reserved; Retiring; Shy; Silent; Withdrawn; Unadventurous |
| Conscientiousness | A strong sense of purpose and high aspiration levels | Competence; Order; Dutifulness; Achievement Striving; Self-discipline; Deliberation | Thorough; Neat; Well-organized; Diligent; Achievement-oriented; Efficient | Disorganized; Lazy; Irresponsible; Careless; Sloppy |
| Agreeableness | Involves aspects of humanity | Trust; Straight-forwardness; Altruism; Modesty; Tender-Mindedness; Compliance | Helpful; Caring; Nurturing; Emotionally Supportive; Cooperative; Trusting; Forgiving | Hostile; Indifferent to others; Self-centred; Jealous; Spiteful; Selfish |
| Openness to Experience | A need for diversity, novelty, and change | Fantasy; Aesthetics; Feelings; Actions; Ideas; Values | Creative; Intellectual; Curious; Innovative; Flexible; Imaginative | Conservative; Shallow; Simple; Closed-minded |

Chapter 3

# State of the Art

This chapter comprises the related work already developed. Initially, the eye and mouse movements comparison studies are introduced. Afterwards, projects concerning decision making issues - indecisiveness, survey response difficulty and medical decision making - are presented.

## 3.1 Eye and Mouse Movements

### 3.1.1 Eye Tracking Data Correction

A problem that arises in every study involving eye tracking data is due to equipment losses of calibration. However, it is possible to correct this data. Hornof and Halverson [46] exposed an approach that depends on required fixation locations to recalibrate the eye tracker. In their experiment, it was required to click in a specific target and, assuming that the participant looks at the target during the click, if the distance between the eye gaze data on the moment of click and the target was higher than a certain threshold, the eye tracker would be automatically recalibrated after the click. An alternative technique consists in two linear regressions (one for horizontal dimension - X axis - and another for vertical dimension - Y axis) between the known data points and the corresponding raw data [47]. The method applied in this dissertation to correct the eye tracking data is similar to the latter, but with some modifications.

### 3.1.2 Eye and Mouse Movements Comparison

Previous studies suggested that mouse movements are related to eye movements during web browsing [4] and during research tasks [5, 48]. The metrics used were the average distance between the eye gaze and the mouse cursor, the analysis of the common regions visited [4, 5] and the correlation between the times spent in each region of the web page by the gaze and the cursor [4]. Guo and Agichtein [48] predicted, with an accuracy of 77%, the regions where the eye gaze and mouse cursor were within 100 pixels of each other using mouse cursor features.

The relationship between the eye and mouse movements was never explored during decision making tasks. Therefore, in this dissertation, these movements were compared in this context and new measures will be presented.

## 3.2 Decision Making

### 3.2.1 Indecisiveness

Germeijs and Verschueren [49] investigated the relation between indecisiveness and personality. It was revealed a strong and positive correlation between indecisiveness and Neuroticism and a negative correlation between indecisiveness, on the one hand, and Conscientiousness and Extraversion, on the other.

Career decisions are common and difficult and, therefore, there are several studies concerning career indecision [1, 50]. Fabio et al. [1] compared career indecision (measured by the Career Decision-making Difficulties Questionnaire) with indecisiveness and concluded that the first is most highly related to emotional intelligence, while indecisiveness is most highly linked to personality traits. The correlation between career indecision and indecisiveness was moderate and, hence, both correlated positively with Neuroticism and negatively with Extraversion (in line with [49]) and emotional intelligence. Lastly, it was also demonstrated that career indecision is inversely related to career decision self-efficacy and, accordingly, it is possible to infer that Extraversion is positively correlated to career decision self-efficacy and that Neuroticism is negatively related to it, which is confirmed by Page et al. [50]. The latter also reported a positive correlation between career decision self-efficacy and Conscientiousness.

Watson [51] analysed the validity of measuring indecisiveness with mouse tracking. Several features were computed and it was only found one significant correlation with indecisiveness - the number of vertical direction changes. Surprisingly, it was not obtained an association between indecisiveness and response times. Consequently, the usage of mouse tracking to measure indecisiveness was not supported.

All the studies above used Indecisiveness Scale questionnaire to measure indecisiveness.

### 3.2.2 Survey Response Difficulty

Mouse cursor movements give insights into mental processes [52]. Accordingly, there are mouse movement patterns associated to uncertainty. Cepeda et al. [53] defined several patterns of mouse movement behaviour extracted from online surveys response. It was distinguished the revisit pattern, which consists in returning to previous answered questions, the hover pattern, verified when the users hover various alternatives before selecting their final answer, the horizontal direction inversions, straight and curvy patterns, among others.

Schneider et al. [52] investigated the effect of ambivalence on mouse cursor trajectories by assessing response times and the maximum deviation from the idealized straight line trajectory toward the unchosen answer. It was concluded that, in case of uncertainty, the maximum deviation is higher. In agreement with [51], it was not found a relation between response times and uncertainty.

Zushi et al. [54] developed a software that tracks mouse movements of students during their learning activities in order to help teachers understand their students' behaviours. It was verified that mouse trajectories become unstable (e.g. excessive number of horizontal direction inversions) when learners are hesitant. It was also perceived that response times and the number of horizontal direction inversions have a negative and strong correlation with the ratio of correct answers. That probably means that horizontal direction inversions and response times are good predictors of uncertainty (which contradict the findings of [51] and [52] related to response times). Accordingly, Conrad et al. [55] used response times and age to detect response difficulty. Conversely, response times do not specify the cause of the delay. Slow responses can be related to multitasking, as answering a call, or the answer could involve mental arithmetic, among several causes [56].

Due to the disadvantages of using only response times as a predictor of response difficulty, Horwitz et al. [56] used mouse cursor trajectories and age to predict it. Hover the question text for more than 2s, horizontal directional inversions, mark a response option text for more than 2s, horizontal and vertical tracking (which occur when the mouse follows the eye while reading) were the variables computed, but only hover, horizontal directional inversions and mark an option turned out to be significant predictors of uncertainty. On the one hand, hover, horizontal and vertical tracking were binary features indicating whether each movement occured or not in each question. On the other, horizontal direction inversions and marker were coded "0"if the movement did not occur, "1"if the movement occured once per question and "2"if it occured more than once per question. To construct the model, the participants classified each question in terms of difficulty. The accuracy of the model was 74.28%, lower than the already existing model that detects moments of uncertainty based only on response times and age [55] with an accuracy of 77.98%. Complementing the information from both models, the accuracy poorly rose to 79.11%. One of the main objectives of this dissertation is similar to this research, however, the approach was different. Firstly, the extracted features were not the same. Additionally, the variables were not considered as binary or in a 3-point scale. For example, the hover time and the number of different answers hovered may provide more information than the perception of its occurrence or non-occurrence.

### 3.2.3 Medical Decision Making

There are not many studies concerning the influence of personality in medical decision making. Nonetheless, Pilárik and Sarmány-Schuller [57] related the decision making process of paramedics with personality. Male decision makers characterized by low scores in Neuroticism and high in Extraversion, as well as fast responses, had the best performance in decision making. It was concluded that the combination of low Neuroticism and high Extraversion leads to adaptive and positive coping strategies. Good performance in females was related to low emotional awareness, ability to deal with stress and fast responses.

The assessment of medical decision through mouse movements, on its turn, was not studied yet. On that account, this dissertation consists in an innovative research.

Chapter 4

# Experimental Procedure

This chapter describes the experimental procedure involved in this dissertation, including the population characterization and the data acquisition conditions. Three different studies were executed - the comparison between eye gaze and mouse cursor trajectories, the construction of the model that detects events of uncertainty through mouse movements obtained with survey responses and, lastly, the application of the model in a clinical context, where mouse data was tracked during a medical decision making task.

## 4.1 Eye and Mouse Movements Comparison

### 4.1.1 Experiment Description

All the three experiments described in this chapter began with the collection of personal data and a written consent for the use of the participants' data according to the guidelines of the Declaration of Helsinki for research involving humans.

Eye and mouse tracking data, used to be compared, were collected during the realization of the IGT. The IGT is a widely explored game that simulates the daily decisions made under uncertainty. It is a card game with four decks that differ in the amount of money that can be won or lost. The game starts by giving the player a fictitious amount of money that should be increased as much as possible. It covers 100 trials, which is unknown to the player, and in each one of them the participant needs to choose one card out of four (by clicking on it with the mouse cursor). After each choice, it is revealed the money won or lost. At a certain moment of the IGT, the player should understand that there are two advantageous decks [58].

The distribution of the decks on the screen was adjusted to acquire the eye-tracking data, with two decks at the top and two decks at the bottom, which is illustrated in figure 4.1.

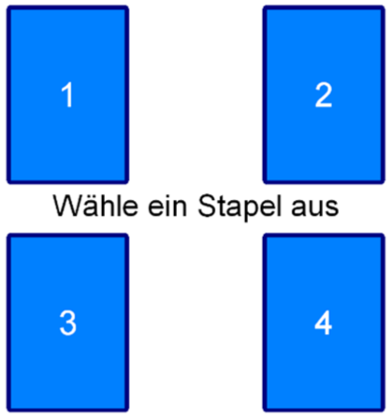



Figure 4.1: Decks' disposition on adapted IGT.

The data acquisition was made in a quiet and dim room without vigilance.

### 4.1.2 Data Acquisition

During the IGT, the mouse tracking data was being sent to a server machine via Asynchronous JavaScript and XML (AJAX), where the data was saved as a file in a database. This file contains the person ID, the trial number, the x and y cursor's position (in pixels) and time.

The eye gaze data file, acquired with the eye tracking system, includes the person ID, the trial number, the x and y position to where the right and left eyes were looking, among other information.

### 4.1.3 Participants

The requirements to participate in this study were:

- Being healthy;
- Have normal, or corrected-to-normal, vision;
- Not having a medical history of neurological or psychiatric illnesses;
- Not being currently medicated;
- Being native or fluent speakers of standard German.

The participants were psychologists or university students from areas like technology, engineering and economics. These students were from the Swiss Federal Institute of Technology and the University of Zurich. They were paid with 20 Swiss Francs, or the equivalent in Euros, or, in the case of psychologists, with hours.

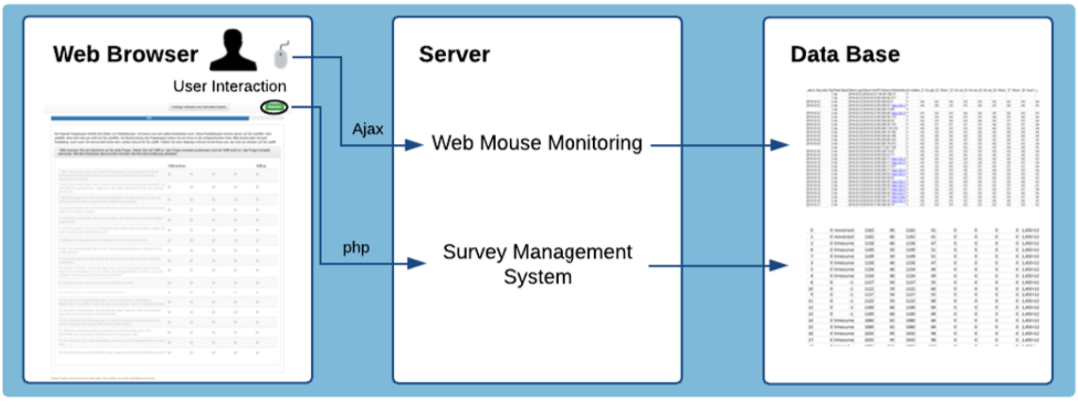



Figure 4.2: Data acquisition architecture. From [53].

## 4.2 Uncertainty Model

### 4.2.1 Experiment Description

This study was based on responses to an online survey constituted by three different personality questionnaires - NEO FFI and other two. The mouse movements obtained from all the questionnaires were used to construct the uncertainty model. To measure personality, only NEO-FFI was used. The personality questionnaires answers were 5-point Likert-type scale.

The data acquisition was made in a quiet room without vigilance.

### 4.2.2 Data Acquisition

While the participants were completing the online survey, the mouse tracking data was being sent to a server machine via AJAX, where the data was saved as a file in a database. The survey results were also saved on this database via the Survey Management System using PHP: Hypertext Preprocessor (PHP). This process is represented in figure 4.2.

The mouse cursor data file contains the frame number, event type (0 during movement, 1 when the mouse button is pressed down in the beginning of a click and 4 when the button is released in the end of a click), question number if hovered, answer number if hovered, x and y cursor's position (in pixels) and time.

The survey file includes the questionnaire start and end times, Internet Protocol (IP) address, name, e-mail, birthday, phone number and the answers to each question. Each personality scale was calculated by accessing the answers related to that scale.

Since files don't contain person identification, the files were associated and the subjects were identified based on times in files.

### 4.2.3 Participants

The requirements to participate in this study were the same as those presented in section 4.1.3. The conditions of the research were also the same. Conversely, the participants were not the same.

## 4.3 Application of the Model to a Clinical Context

### 4.3.1 Experiment Description

This experiment was conducted through an online survey. This survey included the NEO-FFI, to assess the Big Five personality dimensions, and three complex medical cases. The answers of the NEO-FFI were 5-point Likert-type scale.

Physicians evaluated the severity of multimorbid patients' potentially harmful DDI. Three hypothetical cases with different degrees of severity were used, one of them with a high potential for harmful interactions, another with a low potential for harmful interactions, and the last in between. The most complex DDI are covered by the high conflict case, however, the number of complex DDI is higher in the moderate conflict case. The low and the high conflict cases were based on the cases used in [6] and adapted by physicians. The medical history of each patient covers six conditions and six drugs, resulting in 66 interactions. Multimorbidity Interaction Severity Index (MISI) is a web-based tool that provided the cases and the respective patients' histories, as well as all possible DDI randomly ordered, which were adapted to this thesis' needs. Each case included a multimorbidity interaction matrix in order to collect the physicians' judgements about the DDI severity, the probability of occurrence of a problem and the priority to intervene. The physicians also reported how confident they were about the ratings of each DDI. The answers were 4-point Likert-type scale and all the questions were mandatory. The structure of the multimorbidity interaction matrices is illustrated in table 4.1.

Following the DDI evaluation, the participants answered questions related to the experience of using the three rating scales (severity, probability and priority). Lastly, the cases' overall complexity and severity were estimated.

Each medical case took, in average, 40 minutes to conclude. In order to avoid divergent behaviours due to tiredness, the cases were completed in distinct occasions and in a random order (different for each physician).

During this task, the activity of the mouse cursor was tracked. Afterwards, the mouse cursor data was used to detect events of uncertainty.

The data acquisition is described in section 4.2.2.

### 4.3.2 Participants

The requirements to participate in this study were:

- Being a doctor;

Table 4.1: Multimorbidity interaction matrix example with some possible DDI.

| | | **Severity** | | | | **Probability** | | | | **Priority** | | | | **Confidence** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Hypertensive emergency | Metformin | | | | | | | | | | | | | | | | |
| Renal failure | Depression | | | | | | | | | | | | | | | | |
| Diabetes Mellitus II | Escitalopram | | | | | | | | | | | | | | | | |
| Lisinopril | Prednison | | | | | | | | | | | | | | | | |
| ... | ... | | | | | | | | | | | | | | | | |

- Being healthy;
- Have normal, or corrected-to-normal, vision;
- Not having a medical history of neurological or psychiatric illnesses;
- Not being currently medicated;
- Being native or fluent speakers of standard German.

The participants were physicians from the University Hospital of Zurich with a maximum of 4 years of experience in internal medicine.

# Chapter 5

# Methods

In this chapter, the materials and methods used are described and justified. Firstly, it focus on eye and mouse movements comparison study and, subsequently, on the uncertainty model construction. The latter comprises features extraction and selection, the classification process, as well as the relation between uncertainty and personality. Finally, the created model was applied to the medical decision making task.

## 5.1 Eye and Mouse Movements Comparison

### 5.1.1 Technological Materials

For the comparison between eye and mouse movements study, the SMI RED, by Senso-Motoric Instruments [59] was used to acquire the eye gaze data.

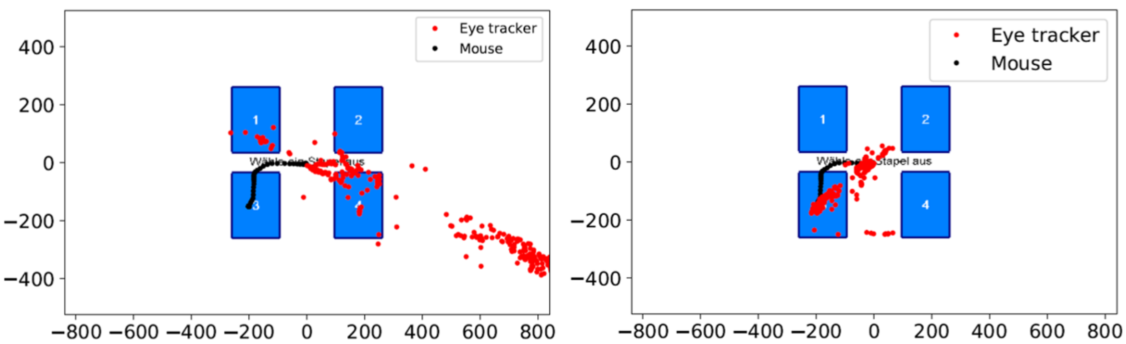
The IGT was programmed in the Presentation software from Neurobehavioral Systems [60].

Data analysis was executed using Python language [61] by the code editor PyCharm 2017.2.4 [62]. The python packages used were NumPy, the essential library for computing [63], Matplotlib, a plotting package [64], SciPy, that provides mathematical functions [65] and Pandas, which supplies data structures and data analysis tools [66].

### 5.1.2 Eye Tracking Data Correction

Part of the data acquired with the eye tracker was lost, since commonly the eye tracker can not recognize the gaze point. Taking this into account, to correctly compare the mouse and eye tracking data, it was necessary to remove the mouse cursor data correspondent to the eye gaze lost data.

Furthermore, some of the eye tracking data was unsatisfactory, since the equipment loses calibration regularly. This problem was essentially solved with two linear regression (one for horizontal dimension - X coordinates - and another for vertical dimension - Y coordinates) between the known data and the corresponding raw data in two parts. The players were instructed to focus on the centre of the screen in the beginning of each trial, and, therefore, the eye tracking data was adjusted accordingly, by accessing the median of the initial value of all trials for each participant and considering it the centre. After this step, the major part of the trials was well calibrated. Nonetheless, some data was still not acceptable and, therefore, the second part of the calibration was applied to these trials. To do so, it was assumed that when people click in a target they tend to look at it. The distance between the mouse and eye coordinates during the click was accessed for each trial. In the trials which distance was higher than 80 pixels, the ratio between the mouse and eye coordinates in the moment of the click was computed and multiplied to the eye tracking data. The advantage of this two-phased linear regression comprises the non-modification of the trials already well-calibrated only with the first step. Figure 5.1 shows an example of non-calibrated and calibrated trials.




(a) (b)

Figure 5.1: Eye gaze (red) and mouse (black) positions a) before and b) after the calibration process.

Even with the application of the two-phased linear regression, some trials still had not suitable eye gaze data. To remove this data, the ratio between the amount of data outside the central square where the decks were and the total data – *outside ratio* - was computed. The central square is represented in figure 5.2 and it is characterized by absolute values of X and Y smaller than 400. This value was chosen following the observation of several trials, as well as the criterion to eliminate non-calibrated trials - the trials with an outside ratio higher than 0,2 were discarded. A difference between this method to correct eye tracking data and the approaches presented in other studies (e.g., [47]) is this elimination of the trials that, after the correction, are still not calibrated.

Figure 5.3 sumarizes all the steps applied to correct the eye tracking data.

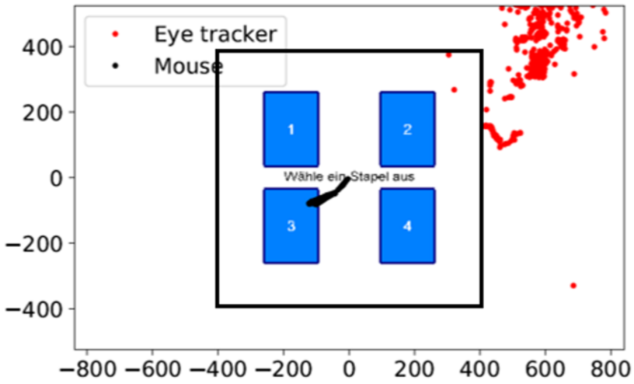



Figure 5.2: Representation of the central square where the decks are on IGT.

### 5.1.3 Comparative Measures

To compare the eye gaze and mouse cursor movements, some measures were computed. To do so, few regions inside the IGT were considered - each deck constituted a region and the last region was constituted by the space outside the decks. Two of the metrics calculated, *common regions ratio* and *times correlation*, were adapted from [4, 5]. Another variable was introduced, *intersection ratio*, to access the amount of time that the eye gaze and the mouse cursor were visiting the same region.

#### 5.1.3.1 Common Regions Ratio

For each trial, it was calculated the ratio between the number of regions that were visited by both eye gaze and mouse cursor and the total number of regions visited. This feature, described in equation 5.6, was called *common regions ratio*. The considered regions were "Deck 1", "Deck 2", "Deck 3" and "Deck 4". For example, if in a certain trial the eye gaze visited "Deck 1" and "Deck 4" and the mouse visited only the "Deck 4", the value of this relation would be 1/2. The mean value of all trials for each participant was determined. This method was based on that implemented by [4, 5]. In this study, it was only considered, for the eye gaze, the regions visited for more than 50 ms, the minimum fixation duration according to [67].

$$\textit{Common regions ratio} = \frac{\textit{Number of decks visited by eye and mouse}}{\textit{Total number of decks visited}} \tag{5.6}$$

#### 5.1.3.2 Times Correlation

The time spent in each region by the eye gaze and the time spent in each region by the mouse cursor were computed for all trials and correlated [4]. The Pearson correlation coefficient, which measures the linear relationship between two datasets, was assessed. It varies between -1 an +1, with the extremes implying the strongest linear relationships (inverse and direct, respectively) and 0 indicating no correlation [68]. It was considered the mean value of the correlations of all regions for each participant. This variable,

**Calibration - step 1: Adjustment of the eye tracking data by correcting a known fixation.**
In this case, the initial value should be in the centre - position (0,0) - and the following algorithm was applied:

$$x = x - x_{0m} \tag{5.1}$$

$$y = y - y_{0m} \tag{5.2}$$

Where $x_{0m}$ and $y_{0m}$ are the median of the initial $x$ and $y$ values of all trials, respectively. These two variables vary from participant to participant.

**Assessment of the distance between eye and mouse coordinates in the moment of the click in each trial.**

Distance < 80 pixels

yes

no

**Calibration - step 2: Calculation of the ratio between eye and mouse coordinates in the moment of the click and subsequent multiplication.**

$$x = x \times x_{ratio} \qquad \text{where} \qquad x_{ratio} = \frac{x_{click,mouse}}{x_{click,et}} \tag{5.3}$$

$$y = y \times y_{ratio} \qquad \text{where} \qquad y_{ratio} = \frac{y_{click,mouse}}{y_{click,et}} \tag{5.4}$$

$x_{click,mouse}$, $y_{click,mouse}$, $x_{click,et}$ and $y_{click,et}$ are the $x$ and $y$ mouse and eye tracker coordinates in the moment of the click, respectively.
Each trial has a different $x_{ratio}$ and $y_{ratio}$.

**Elimination of uncalibrated trials: Computation of the ratio between the amount of data outside the decks area and the total data.**

$$Outside\ ratio = \frac{Number\ of\ samples\ in\ decks'\ area}{Number\ of\ samples} \tag{5.5}$$

$outside_{ratio}$ < 0.2

no

yes

Calibration concluded

Elimination of the trial

Figure 5.3: Eye tracking data correction

defined in equation 5.7, was called *times correlation*. Beyond the regions considered previously, it was also measured the time spent in the region “Outside the decks”. This region was not considered in section 5.1.3.1 since it is a region visited by both eye gaze and mouse cursor in almost every trial, and, therefore, the common regions ratio would increase significantly due to a region which is nearly compulsory to visit (for example, to go from a deck to another it is required to visit the region “Outside the decks”, as well as in the beginning of each trial, where the participants were instructed to focus on the middle of the screen).

$$\textit{Times correlation} = \frac{\sum \rho_{ei,mi}}{\textit{Number of regions}} = \frac{\rho_{e1,m1} + \rho_{e2,m2} + \rho_{e3,m3} + \rho_{e4,m4} + \rho_{eo,mo}}{5} \tag{5.7}$$

In equation 5.7, $\rho_{ei,mi}$ corresponds to the Pearson correlation coefficient between the time spent in region $i$ by the eye gaze and the mouse cursor, with regions “Deck 1” ($\rho_{e1,m1}$), "Deck 2"($\rho_{e2,m2}$), "Deck 3"($\rho_{e3,m3}$), "Deck 4"($\rho_{e4,m4}$) and "Outside the decks"($\rho_{eo,mo}$).

#### 5.1.3.3 Intersection Ratio

The ratio between the time where the eye gaze and the mouse cursor were in the same region and the trial time was also quantified and it was called *intersection ratio*, as it is mentioned in equation 5.8. It was considered all the five regions mentioned above. The mean value of all trials for each participant was calculated.

$$\textit{Intersection ratio} = \frac{\textit{Time of eye and mouse intersection}}{\textit{Trial time}} \tag{5.8}$$

This variable covers the spatial information given by *common regions ratio* and the temporal information given by *times correlation*, but the last two provide insights that intersection ratio might not provide. For example, if a participant looks from a deck to another and, after doing this eye movement, moves the mouse cursor accordingly, there is no intersection ratio, but the eye and mouse movements have some correlation since the path is the same. In this hypothetical case, the regions would be the same, directly affecting the common regions ratio feature. Taking this into account, these three features complement each other.

#### 5.1.3.4 Eye and Mouse Alterations

It was also analysed the number of alterations between decks in both eye and mouse tracking data. The number of alterations is the number of times in a trial that the player changes from a deck to another. For example, if in a certain trial the player changes from “Deck 1” to “Deck 3” and finally chooses “Deck 2”, the number of alterations is 2. This number was accessed for all trials and a correlation between the mouse cursor alterations and the eye gaze alterations was explored. For the eye gaze, it was only considered alterations that lasted more than 50 ms, for the reason presented in section 5.1.3.1.

### 5.1.4 Exploring and Non-exploring Conditions

When the participants know in advance the deck that will be selected, they make a fast decision, therefore eye and mouse move directly from the centre of the screen to the chosen deck. In this condition – condition A -, the both movements should highly correlate. However, when the participants explore the options and waver between different decks – condition B -, those movements may not be correlated. Condition A is verified when the eye gaze and the mouse cursor only visit one region, and condition B in the other trials. The variables *common regions ratio*, *times correlation* and *intersection ratio* were accessed for these two conditions. The *alterations correlation* feature was not calculated to compare these conditions since, in condition A, there are no alterations.

## 5.2 Uncertainty Model

### 5.2.1 Technological Materials

The LimeSurvey web application was used to conduct the personality questionnaires and to implement MISI.

Pycharm 2017.2.4 [62] was used to do the data analysis using python language [61]. Similarly to the first study, the Python packages used were NumPy [63], SciPy [65], Pandas [66] and, additionally, Scikit-learn, used for data mining and data analysis [69], and Seaborn, a statistical data visualization library [70].

Lastly, MISI was the tool that provided medical cases, which were adapted to this thesis' needs [6].

### 5.2.2 Data Pre-processing

Some files containing mouse cursor positions had bugs. The errors found and the corrections applied were:

1. For each new pointer movement, the files had a new line, but, periodically, two different lines were together and they were splited in two distinct lines;

2. Some lines were out of order and, therefore, they were reordered by frame number;

3. Files without the number of frames were identified;

4. If different files had the same IP address and questionnaire, they were concatenated;

5. Same position lines, $(x;y)_i = (x;y)_{i+1}$, were located and the line with $(x;y)_{i+1}$ was removed;

6. Same time lines, $t_i = t_{i+1}$, were identified and the line with $t_{i+1}$ was eliminated;

7. Lines with NotANumber values were removed;

8. Data acquired with touch screen devices was excluded since the mouse cursor movements is lost in these devices.

To identify data acquired with touch screen devices, the ratio between the events where the mouse is moving (events=0) and the click down events (event=1) was computed. If less than 2, it was considered a touch screen device, as expressed by equation 5.9.

$$\frac{\textit{Number of events} = \textit{0}}{\textit{Number of events} = \textit{1}} < 2 \tag{5.9}$$

### 5.2.3 Features Extraction

Several features related to uncertainty behaviour were computed for each question of the survey. With these variables, it was created a model that detects the difficult questions for an individual. In this section, the temporal, spatial and contextual features are presented.

#### 5.2.3.1 Temporal Features

Firstly, to access the temporal information, it was necessary to remove the time associated to abandon events. Sometimes, due to external factors (e.g. receiving an e-mail or answering a call), an individual may abandon the survey. Without correction, the questions where the abandons occur could be associated to uncertainty as a result of the time spent there. Therefore, the abandon events were identified - when the mouse cursor is not moving for more than 10 times the mean question time - and removed.

Short times in questions were also ignored. They can be caused by quick visits to the question above or bellow since the question height is small, or by scroll. These events occur when the time spent in a question is lower than 100 ms [71].

The temporal features are *accumulated time*, *time before click*, *pause before click*, *correction time*, *hover selected answer* and *velocity*.

The *accumulated time* is the total time in an item, i.e., the sum of all time intervals in a question, as expressed in equation 5.10, where $\Delta t_{qi}$ represents, hence, a time interval spent in question *i*.

$$\textit{Accumulated time} = \sum \Delta t_{qi} \tag{5.10}$$

*Time before click* is the sum of all time intervals in a question until the first click, as shown in equation 5.11. For example, if a participant enters in a question for the first time at t = 20s, stays in the item for 10s without clicking, abandon the question, comes back at t = 45s and clicks for the first time at t = 50s, the time before click is 15s.

$$\textit{Time before click} = \sum_{enter}^{1^{st} click} \Delta t_{qi} \tag{5.11}$$

The *pause before click*, i.e., the time interval that an individual remains stopped before clicking an answer, was also computed, based on [72]. If the participant clicks more than once in a single question (to correct a previous answer), this value is averaged.

*Correction time* is the sum of all time intervals in a question from the first click to the last click (last correction), as it is indicated in equation 5.12. If there is not any correction, the result is zero. The computation of this feature is similar to that of *time before click*, but instead of being the sum of the time intervals until clicking for the first time, it is from the first choice to the last one.

$$Correction\ time = \sum_{1^{st}click}^{lastclick} \Delta t_{qi} \tag{5.12}$$

*Hover selected answer* is the ratio between the sum of the time intervals spent hovering the selected answer of a certain question and the total hover time in that question. It was based on a feature extracted by [53, 56]. In this study, when an individual is in the response area, i.e., close to one of the possible answers, it is considered that the participant is hovering that answer. This feature is described in equation 5.13, where $\Delta t_{hover\ selected\ answer,\ qi}$ represents a time interval spent hovering the selected answer of question $i$ and $\Delta t_{hover,qi}$ is a time interval spent hovering the answers of question $i$.

$$Hover\ selected\ answer = \frac{\sum \Delta t_{hover\ selected\ answer,\ qi}}{\sum \Delta t_{hover,qi}} \tag{5.13}$$

The *velocity* was also calculated and it is expressed in equation 5.14, where $n$ is the number of samples. $\Delta x_i$, $\Delta y_i$ and $\Delta t_i$ are defined in equation 5.15. To create the model, it was used the *mean velocity*. Correspondingly, it was necessary to compute, on the one hand, the mean value across all samples (equation 5.14 symbolizes the velocity between two samples) and, on the other, the mean value across all visits to a certain question. To compute this variable, in order to have equal temporal intervals proportional to the mean time variance, it was applied a cubic spline interpolation. Using this method, a series of unique cubic polynomials are adjusted between the data points, resulting in a smooth continuous curve [73].

$$Velocity = \frac{\sqrt{\Delta {x_i}^2 + \Delta {y_i}^2}}{\Delta t_i}, \qquad i = 1, \ldots, n-1 \tag{5.14}$$

Where

$$\Delta x_i = x_{i+1} - x_i \qquad \Delta y_i = y_{i+1} - y_i \qquad \Delta t_i = t_{i+1} - t_i \tag{5.15}$$

#### 5.2.3.2 Spatial Features

Firstly, it was applied a cubic spline interpolation to smooth the spatial signal, producing intervals equal to the mean distance variance. Subsequently, the spatial features *distance*, *distance from answer* and *straightness* were computed.

The *distance* travelled in a question was measured and it is defined in equation 5.16, where $s_{qi}$, displayed in equation 5.17, represents the travelled distance in a visit to question *i*. $\Delta x_i$ and $\Delta y_i$, on their turn, are illustrated in equation 5.15.

$$Distance = \sum s_{qi} \tag{5.16}$$

Where

$$s_{qi} = \sum_{i=1}^{n-1} \sqrt{\Delta {x_i}^2 + \Delta {y_i}^2} \tag{5.17}$$

The *distance from answer*, i.e., distance from the path inside a question to the selected answer, was also computed. This variable is illustrated in equation 5.18, where $x_{answer}$ and $y_{answer}$ are the $x$ and $y$ coordinates of the question's last click. For the construction of the model, it was calculated the mean *distance from answer*, which covers the mean value across all samples and, then, the mean value across all visits to a specific question.

$$Distance\ from\ answer = \sqrt{(x_i - x_{answer})^2 + (y_i - y_{answer})^2}, \qquad i = 1, ..., n-1 \tag{5.18}$$

*Straightness* is the ratio between the Euclidean distance from the moment of entering in a question until leaving it and the total distance travelled in that question [25]. It is defined in equation 5.19 and $s_{qi}$ is described in equation 5.17. The mean *straightness* over all the visits to a specific question was used.

$$Straightness = \frac{\sqrt{(x_1 - x_n)^2 + (y_1 - y_n)^2}}{s_{qi}} \tag{5.19}$$

#### 5.2.3.3 Contextual Features

The contextual features comprise the number of *interactions* with each question (i.e., the number of times in each question) as well as the number of *revisits*, which is the event of going back to a previous question without changing its answer.

The number of corrections was also calculated. There are two types of corrections - *corrections within item* and *corrections between item*. The first occurs when an individual selects an answer, remains in the same question and changes the option, while the latter happens when a person selects an answer, moves forward to next questions and, after answering at least one more question, goes back and changes the previous answer.

The number of <*-turns*, i.e., horizontal direction changes [53, 54, 56], was extracted by computing horizontal trajectory derivative changes from positive to negative or vice-versa.

Lastly, the relative number of *hovered answers* was computed and it is illustrated in equation 5.20.

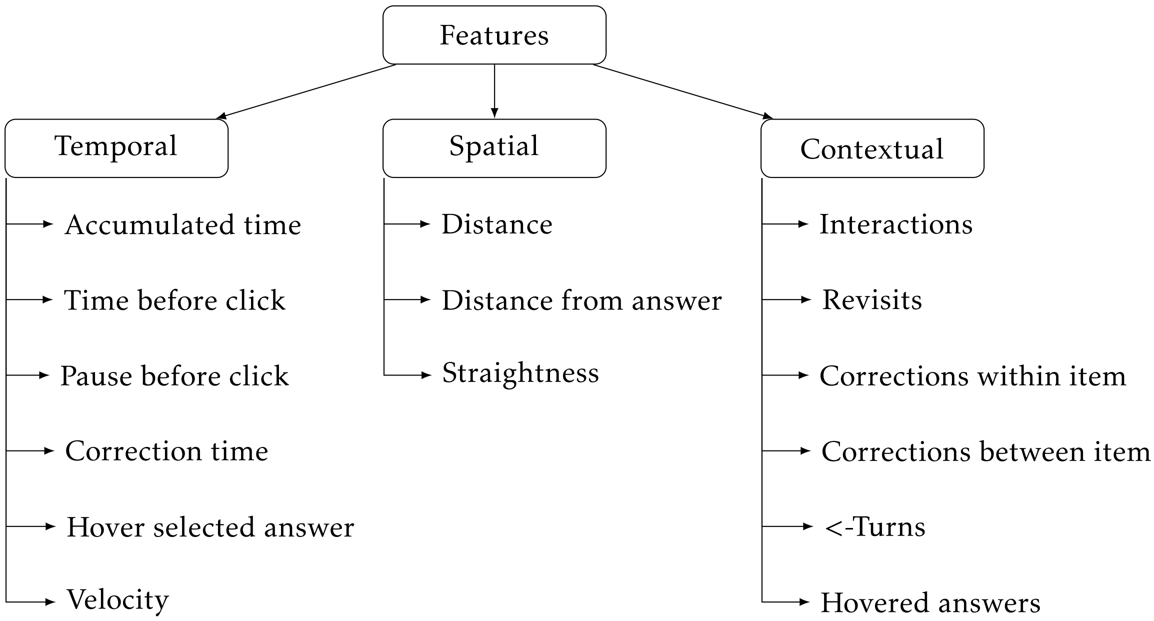



Figure 5.4: Features extracted.

$$\textit{Hovered answers} = \frac{\textit{Number of hovered answers}}{\textit{Total number of answers}} \tag{5.20}$$

Figure 5.4 summarizes all the extracted features.

#### 5.2.3.4 Features Normalization

Distinct people express uncertainty differently. For example, maybe the time spent in a difficult question by a fast person is equal to the time spent in an easy question by a slower individual. Accordingly, the features were normalized for each person separately using the formula presented in equation 5.21, where $z_i$ represents the sample $x_i$ after normalization, $\overline{x}$ and $\sigma$ are the mean and standard deviation of the samples, respectively. This normalization is known as z-score [74]. Applying this transformation, the samples are reshaped so that its mean and standard deviation become 0 and 1, respectively [75].

$$z_i = \frac{x_i - \overline{x}}{\sigma} \tag{5.21}$$

Nonetheless, with all the features normalized, it is only possible to identify the most difficult questions for each individual. In the hypothetical case of uncertainty in all questions (or a great part of them), this would be a problem. Therefore, the original values of each feature were also used to construct the model. Taking this into account, 30 features were used - 15 normalized and 15 not normalized.

Subsequently, all the features from all the participants were concatenated and normalized in order to standardize the range of the variables.

### 5.2.4 Features Selection

There is a negative effect of using irrelevant features in machine learning systems. Some classifiers are not sensible enough to detect the influence of relevant features in the presence of many variables [76]. Taking this into account, it is advantageous to precede learning with a feature selection stage [29].

Accordingly, the highly correlated features were eliminated [29], since the information they provide is almost the same. The Pearson correlation coefficient was accessed and, if two features had an absolute coefficient higher than 0.9, one of them was left out.

### 5.2.5 Training and Testing

In order to train the uncertainty model and, afterwards, test it, several examples of uncertainty and certainty while answering survey questions were needed. With the combination of features and the labels associated to them (in this case, certainty or uncertainty), the model is trained. To test it, the results obtained with the implementation of the model are compared to the baseline (or "true") outcomes, i.e., the certainty and uncertainty labels mentioned above. Since the mouse cursor data has both spatial (vertical and horizontal directions) and temporal information, numerous mouse movements videos were observed and those needed examples were selected. Clearly, detecting uncertainty involves a subjective evaluation, which could be a barrier to construct an accurate model. To escape this problem, 3 individuals made the analysis separately and 1 of them had few information about the study. A question was selected (for certainty or uncertainty) only if at least 2 people had chosen it.

It is recurrent to use, for example, two-thirds of the data for training and the remaining one-third to test the model. Conversely, the training or testing sets might not be representative. This problem may be soluted by repeating the process of training and testing various times with different samples. Taking this into consideration, it was used the *10-fold cross validation*. In this procedure, the data is divided in ten approximately equal partitions, where one of them is used for testing, while the remaining nine are used for training, and the process is repeated ten times. In each iteration, the datasets change and, accordingly, every instance is used for both training and testing, and exactly once for testing. Finally, the ten estimated accuracies are averaged to obtain the overall accuracy. The number of folds might have been different, but a considerable amount of tests has led to the conclusion that ten is the number that reaches the best estimate of error. Even though these results are questionable, the 10-fold cross-validation has become the conventional practice [29].

### 5.2.6 Classification

The applied classification method was *Logistic Regression*, due to its effectiveness when the outcome variable is dichotomous (in this case, the outcome might be certainty or

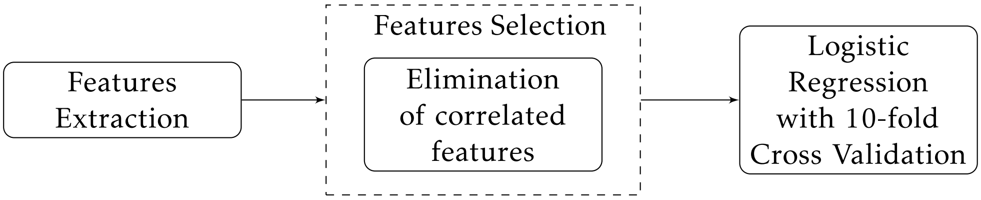



Figure 5.5: Uncertainty model construction.

uncertainty). In this technique, the probability of occurrence of an event is estimated by fitting the data to a logistic curve. Accordingly, non-linear relationships between the input features and the outcome variable can be handled [77].

The fundamental mathematical concept underlying *Logistic Regression* is the logit. The logit is the natural logarithm of odds ratio, which is the ratio between the probability of occurrence of an event (in this case, uncertainty) and the probability of non-occurrence of the same event. The logistic model has the form presented in equations 5.22 and 5.23, where $p$ represents the probability of an event, $\beta_i$ illustrates the regression coefficients and $x_i$ are the input features [76].

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n \tag{5.22}$$

Solving for $p$,

$$p = \frac{1}{1 + e^{-(\beta_0 + ... + \beta_n x_n)}} \tag{5.23}$$

When $p \geqslant 0.5$ it is predicted $Y = 1$ (uncertainty), otherwise, $Y = 0$, where $Y$ is the outcome variable [78]. From equation 5.23, it is possible to verify that a positive $\beta_i$ increases (and a negative $\beta_i$ decreases) the probability of $Y = 1$.

Figure 5.5 summarizes the uncertainty model construction.

### 5.2.7 Model Performance Evaluation

In binary classification, data is constituted by two opposite classes, positives and negatives. Accordingly, the possible outcomes comprise True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN), as illustrated in table 5.1 [79]. In this study, the positives are the questions linked to uncertainty.

Table 5.1: Confusion matrix.

| | **Actually positive** | **Actually negative** |
|---|---|---|
| **Predicted as positive** | TP | FP |
| **Predicted as negative** | FN | TN |

The true positive rate, or *sensitivity*, and the true negative rate, or *specificity*, were computed [29]. In this case, the *sensitivity* represents the probability of a question that evokes uncertainty being classified as an instance of uncertainty, and it is described in equation 5.24. Therefore, this metric evaluates the model capacity to correctly classify uncertainty. *Specificity*, on the other hand, provides the probability of a question associated to certainty being correctly classified and it is illustrated by equation 5.25.

$$Sensitivity = \frac{TP}{TP + FN} \tag{5.24}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5.25}$$

Since the data is imbalanced (there are more certainty events than uncertainty occurrences), the most appropriate measure to evaluate the model performance is *f1 score*, defined in equation 5.26 as the harmonic mean between *precision* and *recall*. *Recall* is a synonym of *sensitivity*, as it is possible to verify in equation 5.27. *Precision*, on its turn, represents the probability of a certainty event being classified as an uncertainty event (see equation 5.28) [80].

$$f1\ score = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{5.26}$$

Where

$$Recall = Sensitivity \tag{5.27}$$

And

$$Precision = \frac{FP}{FP + TN} \tag{5.28}$$

### 5.2.8 Relation between Uncertainty and Personality

The FFM personality scales were measured through the NEO-FFI questionnaire. The NEO-FFI has 60 items, 12 for each personality dimension. Each question possess a standard five-point Likert-like scale as its possible answers, where 1 corresponded to "completely disagree"and 5 to "completely agree". The result of each scale - Neuroticism, Extraversion, Conscientiousness, Agreeableness and Openness to Experience - was accessed by computing the average of the answers related to each scale.

Subsequently, it was analysed the relation between personality and uncertainty. Due to possible data losses and consecutive divergence in questions' total number, the percentage of questions associated to uncertainty by the model was calculated for each participant. To search for a correlation between this measure and each personality scale, the Pearson correlation coefficient was employed.

### 5.2.9 Application of the Model to a Clinical Context

Several physicians answered a survey where the DDI severity in multimorbid patients was evaluated, simulating real-life medical decision making. The structure of this survey was similar to the personality questionnaires and, therefore, the constructed uncertainty model was applied to this context.

The features presented in section 5.2.3 were extracted. In table 4.1, it can be noticed that the multimorbidity interaction matrix includes four distinct scales (severity, probability, priority and confidence), and, hence, all the features were averaged, since the questionnaires mentioned above had only one scale. Uncertainty was, then, predicted for the items of the survey. Subsequently, it was explored if the amount of difficult DDI in each case was consistent with its complexity (i.e., if the low conflict case had the lowest amount of difficult DDI, for example), in order to test its validity across different contexts.

Moreover, the relation between uncertainty and personality was also assessed with this data since personality may manifest differently in totally distinct situations - while answering to questionnaires and during career decision making.

# Chapter 6

# Results and Discussion

In this chapter, the results of the studies that constitute this dissertation are presented and discussed. It comprises the population description of each study, the outcomes of the comparison between eye and mouse trajectories, the performance evaluation of the uncertainty model, the correlations between uncertainty and each personality scale and, lastly, the effectiveness analysis of the uncertainty model in the clinical environment.

## 6.1 Eye and Mouse Movements Comparison

### 6.1.1 Population Description

81 volunteers - 59 female and 22 male - participated in this study, with ages ranging from 16 to 34 years old.

### 6.1.2 Eye Tracking Lost Data

As mentioned in section 5.1.2, part of the data acquired with the eye tracker was lost. In figure 6.1, it is presented a histogram with the mean percentage of eye tracking lost data. The average percentage of lost data was 24.11%. It was thus verified that the eye gaze data is not always satisfactory, in line with [4]. It is fundamental to highlight that there was not loss of mouse tracking data.

Besides the lost data, some eye tracking data was removed. As explained in section 5.1.2, the last step of the eye gaze data correction was the elimination of the trials that, in the end of the calibration process, were still not calibrated. The mean value of discarded trials was 11.37 trials. Figure 6.2b illustrates the number of discarded trials. Figure 6.2a, on its turn, shows a histogram with the number of trials that would be discarded only with the first stage of the calibration. In this case, the mean discarded trials

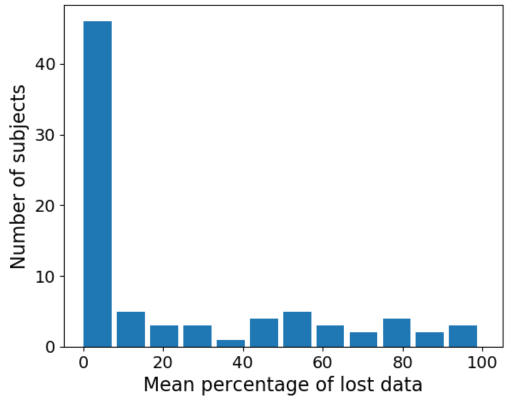



Figure 6.1: Mean percentage of eye tracking lost data.

would be 15.54 trials. Note that in figure 6.2a there are subjects with extremely high numbers of removed trials, while in figure 6.2b the maximum number of eliminated trials is less than 50 trials. This demonstrates a slight progress in the data cleaning due to the second step of the process presented in section 5.1.2.

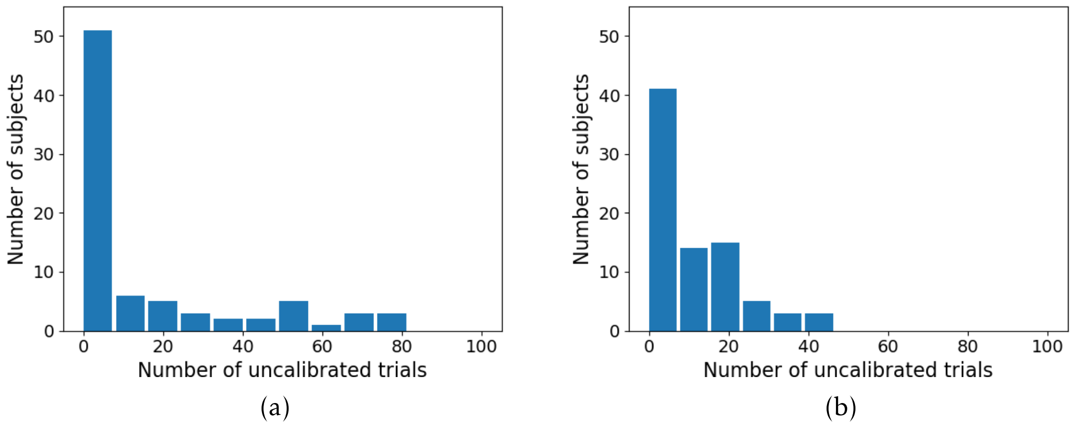
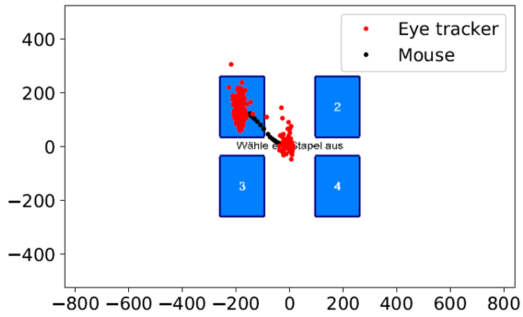


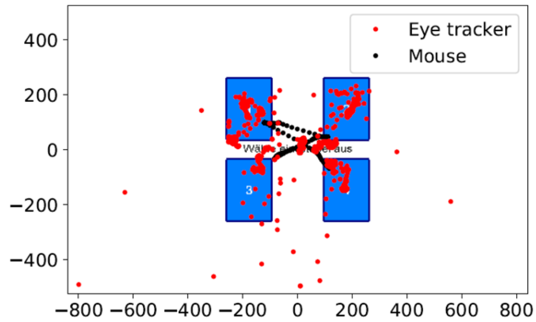

Figure 6.2: a) Number of trials that would be discarded only with the first step of the calibration process and b) the actual number of discarded trials.

### 6.1.3 Data Visualization

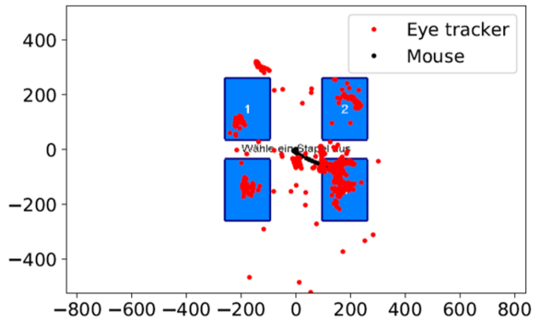
As a first approach to compare both data, eye and mouse movements were visualized. Figure 6.3a shows a trial with the most common behaviour - the mouse moves along a straight line from the centre to the chosen deck and the eyes follow this path. Figure 6.3b presents a trial where the participant waver between all the decks and the behaviour captured by the eye tracker and by the mouse cursor is similar, since all the regions were visited by both. However, there are trials where eye and mouse movements are distinct. For example, in figure 6.3c, the eye gaze visits all the decks and the mouse cursor only visits one.




(a)




(b)




(c)

Figure 6.3: Eye gaze (red) and mouse cursor (black) movements. In a) it is presented a straight path where the eye and the mouse are correlated, b) shows hesitation between decks where the movements are similar and c) displays a trial where the movements are distinct.

### 6.1.4 Comparative Measures

The mean, standard deviation, minimum and maximum values of the comparative measures introduced in section 5.1.3 are shown in table 6.1.

Table 6.1: Variables that quantify the relation between eye and mouse movements.

| | Mean | Standard deviation | Minimum | Maximum |
|---|---|---|---|---|
| **Common regions ratio** | 0.83 | 0.13 | 0.43 | 0.99 |
| **Times correlation** | 0.72 | 0.10 | 0.42 | 0.97 |
| **Intersection ratio** | 0.62 | 0.08 | 0.39 | 0.80 |
| **Alterations correlation** | 0.34 | 0.24 | -0.11 | 0.93 |

The results illustrate a strong relationship between eye gaze and mouse movements.

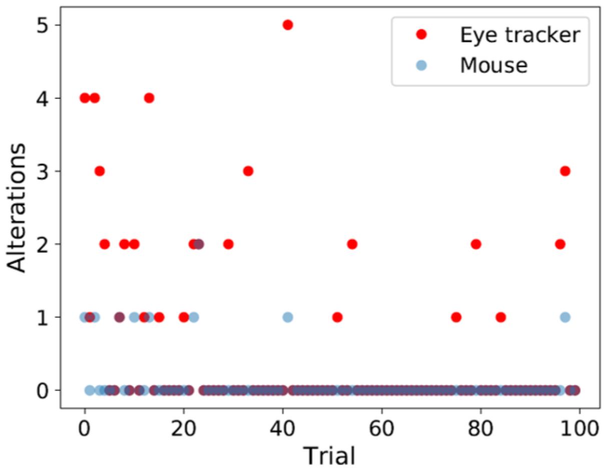



Figure 6.4: Eye (red) and mouse (blue) alterations per trial of a certain participant. When the number of eye and mouse alterations is the same, the dot is purple.

On the one hand, the regions visited by the eye gaze and by the mouse cursor were the same 83% of the times. On the other, the correlation between the time spent in each region by the eye gaze and the mouse cursor was 0.72. And, finally, 62% of the times the eye gaze and the mouse cursor were visiting the same region.

Nevertheless, the correlation between eye and mouse alterations is poor. Therefore, even that the mouse and the eye gaze visit the same regions most of the times, the number of times that the player changes from a deck to another with the mouse and with the eye gaze is not directly related. Generally, the great part of the trials does not have any alterations for both eye gaze and mouse cursor. Moreover, the number of alterations detected with the eye gaze is commonly equal or superior to the number of alterations made with the mouse cursor. It is possible to verify this in figure 6.4, which represents the number of eye (red) and mouse (blue) alterations in each trial of a certain participant. When the number of eye and mouse alterations is the same, the dot is purple.

### 6.1.5 Exploring and Non-exploring Conditions

The comparative measures were computed for non-exploring (A) and exploring (B) conditions and the results are presented in table 6.2 and 6.3, respectively.

As it was expected, the results show a higher correlation between eye and mouse movements in condition A. Nevertheless, although *common regions ratio* shows a decrease of 40% from condition A to condition B, the decline is inferior for *times correlation* (14%) and *intersection ratio* (22%). Accordingly, the time spent by the eye gaze and the mouse cursor in the same regions are comparable in both conditions, but the visited regions diverge in condition B. This suggests that the common regions are the ones where the participant spends the most part of the time, and there are regions that the eye gaze rapidly visits, and the mouse cursor does not visit at all (or the opposite, however, as it

Table 6.2: Variables that quantify the relation between eye and mouse movements in non-exploring condition.

| | Non-exploring condition (A) | | | |
|---|---|---|---|---|
| | **Mean** | **Standard deviation** | **Minimum** | **Maximum** |
| **Common regions ratio** | 0.97 | 0.04 | 0.78 | 1.00 |
| **Times correlation** | 0.78 | 0.09 | 0.43 | 0.93 |
| **Intersection ratio** | 0.64 | 0.10 | 0.35 | 0.83 |

Table 6.3: Variables that quantify the relation between eye and mouse movements in exploring condition.

| | Exploring condition (B) | | | |
|---|---|---|---|---|
| | **Mean** | **Standard deviation** | **Minimum** | **Maximum** |
| **Common regions ratio** | 0.58 | 0.10 | 0.33 | 0.89 |
| **Times correlation** | 0.67 | 0.19 | 0.11 | 1.00 |
| **Intersection ratio** | 0.50 | 0.09 | 0.20 | 0.71 |

was mentioned above, there are, in general, more eye gaze alterations than mouse cursor alterations).

It was demonstrated that, when a participant hesitates between different options, the mouse movements diverge from the eye movements relatively to occasions of certainty. Even so, since eye tracking is associated with losses of calibration and data, taking into account the advantages of using mouse tracking and the overall results, mouse data revealed to be a convenient approximation of eye gaze data.

## 6.2 Uncertainty Model

### 6.2.1 Population Description

The population covered 88 participants. Notwithstanding, the data from 8 of them was excluded as a consequence of using touch screen devices. Furthermore, 1 of the participants was removed from the study due to errors in mouse data. As a result, the study comprised 79 participants, 35 female and 44 male, with ages ranging from 18 to 35 years old.

The violin plot presented in figure 6.5 illustrates the personality questionnaire results of the final 79 participants. Regarding the population distribution, it is possible to verify that there were few individuals classified as high and low in every scale. Neuroticism clearly had the lowest mean, which means that the majority of the population has low

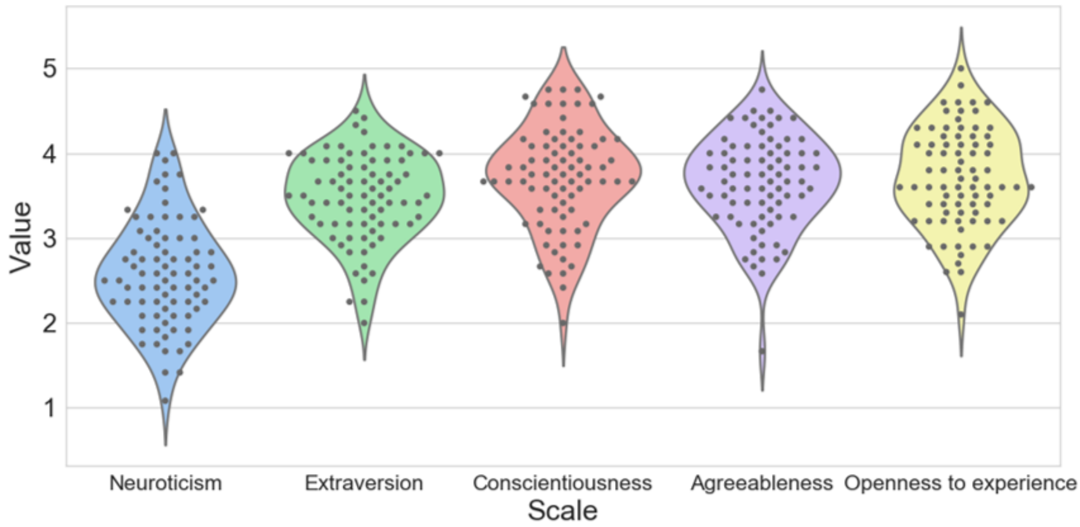



Figure 6.5: Violin plot with the NEO-FFI results. The grey dots represent the results of each person.

scores in this scale.

The mean values and standard deviations were 2.59±0.62 for Neuroticism, 3.48±0.51 for Extraversion, 3.71±0.60 for Conscientiousness, 3.65±0.55 for Agreeableness and, finally, 3.71±0.58 for Openness to Experience.

### 6.2.2 Lost Data

Besides the lost data due to the use of touch screen devices, some mouse data files lost a few samples. The mean percentage of lost samples was 0.11%, which is insignificant.

### 6.2.3 Features Extraction

In this section, some of the temporal, spatial and contextual features computed for the construction of the uncertainty model are illustrated.

#### 6.2.3.1 Temporal Features

Figure 6.6 shows the temporal evolution of a questionnaire completion from the survey of a certain individual. The blue circles represent the moment of entering a question and the numbers inside them are the items' numbers. It is possible to verify that the participant spends a long time in question 1 and 3 when compared with the time spent in the other items.

In figure 6.7, an example of slow and fast movements is displayed. The colour represents velocity, where higher intensity symbolize higher speeds.

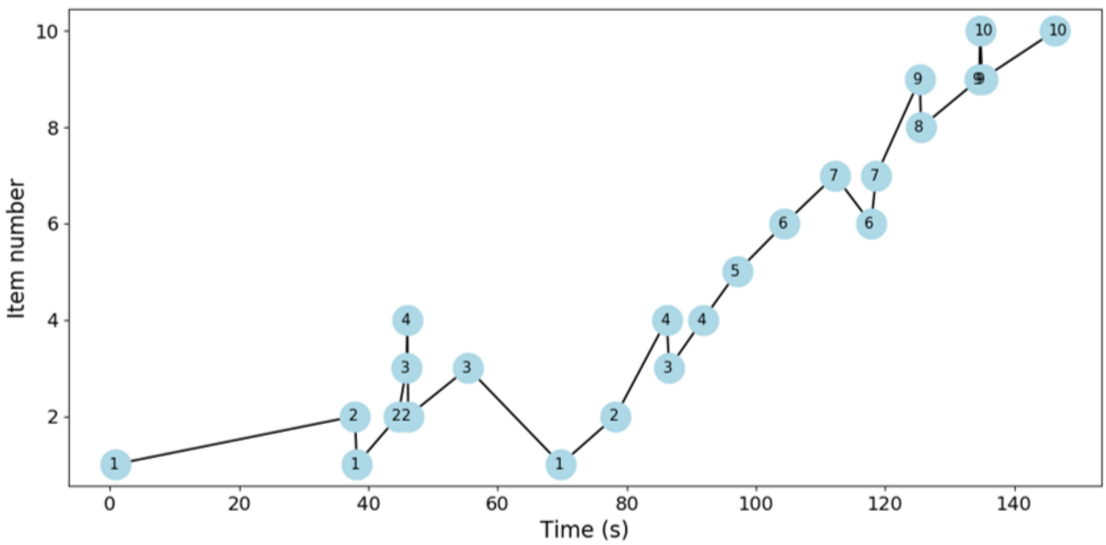
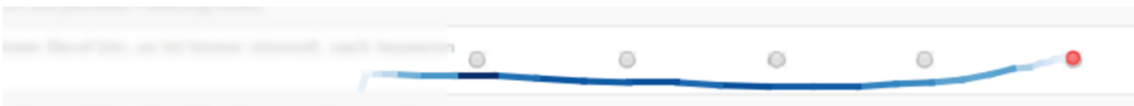



Figure 6.6: Temporal evolution of the first ten questions' response from the survey of a certain participant. Each blue circle represents the moment of entering in an item and the number inside it represents the question number.

(a)



(b)

Figure 6.7: Instance of a) low and b) high *velocity*. Higher colour intensities symbolize higher speeds.

#### 6.2.3.2 Spatial Features

Figure 6.8 illustrates two opposite behaviours. In figure 6.8a, the participant moves the mouse cursor from an answer to the next one in a straight path. Accordingly, the *straightness* is approximately 1. Moreover, the travelled *distance* is short, as well as the average *distance from answer*. In contrast, figure 6.8b shows a movement that could have been exactly the same but turned out to be an extensive path and, therefore, the *straightness* is low.

#### 6.2.3.3 Contextual Features

The number of *interactions* with an item can be inferred by analysing a plot similar to figure 6.6. For example, this participant interacted four times with question 2.

Figure 6.9 displays a visit to a previous question without changing its answer, i.e., a

(a)

(b)

Figure 6.8: Instance of a) high and b) low *straightness*, a) short and b) long *distance* travelled and *distance from answer*. The red dots represent clicks.

Figure 6.9: An example of a revisit. The participant clicked on an item (red dot) and, subsequently, returned to a previous question without changing its answer.

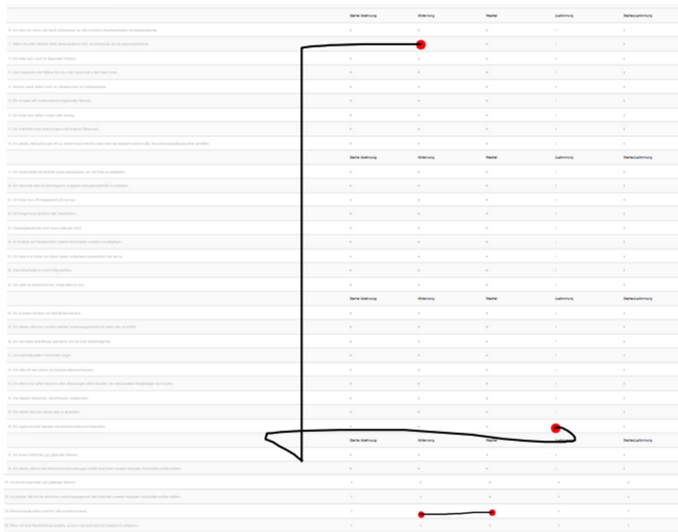


Figure 6.10: Two corrections, a correction between (above) and a correction within (below) item.

*revisit*. Corrections - between and within item - are exhibited in figure 6.10. Finally, in figure 6.11, an instance of <-*turn* is shown. In this image, it is also possible to observe five out of five hovered answers.

#### 6.2.3.4 Features Statistical Analysis

Table 6.4 presents the mean, standard deviation, minimum and maximum of each feature, apart from the normalized features, since they all have a mean value of 0 and a standard deviation of 1 [75].

Figure 6.11: An example of a <-turn.

Table 6.4: Extracted features.

| | **Mean** | **Standard deviation** | **Minimum** | **Maximum** |
|---|---|---|---|---|
| **Temporal features** | | | | |
| **Accumulated time** (s) | 7.46 | 5.24 | 0.63 | 124.29 |
| **Time before click** (s) | 5.90 | 4.77 | 0.01 | 121.48 |
| **Pause before click** (s) | 0.41 | 0.88 | 0.00 | 18.61 |
| **Correction time** (s) | 0.31 | 1.31 | 0.00 | 27.25 |
| **Hover selected answer** | 0.49 | 0.31 | 0.01 | 1.00 |
| **Velocity** (px/s) | 154.71 | 180.04 | 1.26 | 3153.96 |
| **Spatial features** | | | | |
| **Distance** (px) | 423.44 | 364.84 | 5.63 | 9822.23 |
| **Distance from answer** (px) | 112.09 | 84.66 | 2.22 | 652.66 |
| **Straightness** | 0.62 | 0.23 | 0.00 | 1.00 |
| **Contextual features** | | | | |
| **Interactions** | 2.15 | 1.25 | 1 | 13 |
| **Revisits** | 1.13 | 1.24 | 0 | 12 |
| **Corrections within item** | 0.12 | 0.37 | 0 | 4 |
| **Corrections between item** | 0.01 | 0.12 | 0 | 2 |
| **Hovered answers** | 0.61 | 0.22 | 0.20 | 1.00 |
| **<-Turns** | 2.70 | 2.36 | 0 | 77 |

### 6.2.4 Training and Testing Sets

In order to select instances of certainty and uncertainty to train and test the model, mouse movements videos from 6 individuals answering a 60 item questionnaire were visualized. Accordingly, 360 questions were observed, but only 175 were chosen due to the difficulty and subjectivity of the task. 51 items were associated to uncertainty and 124 to certainty.

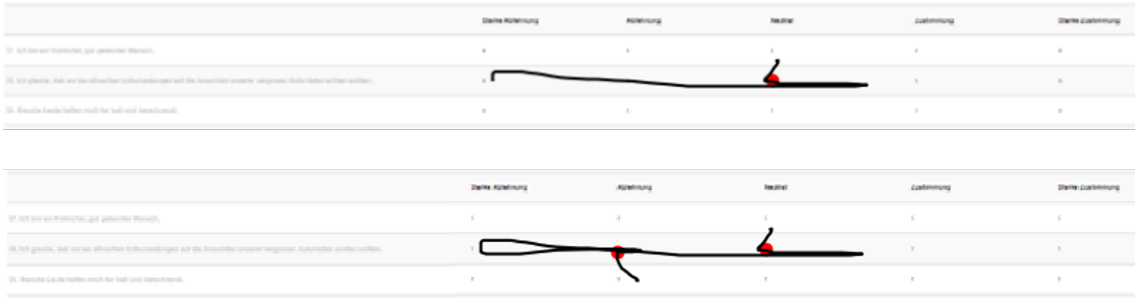


Figure 6.12: Question associated to uncertainty.

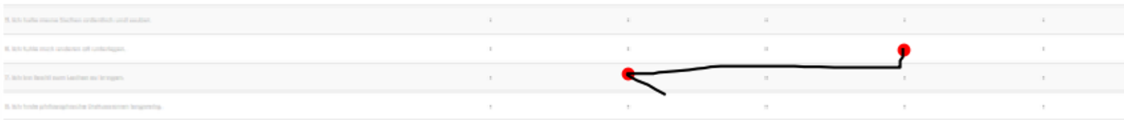


Figure 6.13: Question associated to certainty.

Figure 6.12 shows one of the items selected as an instance of uncertainty. The participant enters the question and immediately selects option 3. Afterwards, the individual moves the mouse cursor towards option 4, but reverses this trajectory until reaching option 1. Subsequently, the direction is inverted and the final answer is option 2. Note that this occurrence comprises a long distance travelled, a low straightness, one correction within item, two <-turns and three out of five hovered answers.

On the contrary, figure 6.13 exhibits an example of certainty, where the mouse moves straightly from the answer of a question to the next one.

### 6.2.5 Features Selection and Relevance

The highly correlated features were removed, as it was explained in section 5.2.4. The features eliminated with this criterion were *time before click*, *hover selected answer*, *straightness normalized*, *revisits*, *revisits normalized* and *hovered answers normalized*. Therefore, the number of final features was 24.

Some features have more importance than others in the classification process. From equation 5.23, it is possible to infer that features with higher regression coefficients are more relevant to the classification. Table 6.5 shows the regression coefficients of the selected features ordered from the highest to the lowest absolute value.

It is possible to verify, from table 6.5, that the number of *<-turns* is the most relevant feature and, with a positive regression coefficient, it increases the probability of detecting an uncertainty event. On that account, when facing uncertainty while interacting with a computer, individuals tend to change the horizontal direction more frequently, probably due to hesitation between consecutive alternatives, in line with [54].

Furthermore, the *distance* travelled produce a great and positive impact on the outcome, suggesting that people move the mouse from a possible answer to another while

Table 6.5: Regression coefficients of the selected features.

| Feature | **Regression coefficient** |
|---|---|
| **<-Turns** | 1.47 |
| **Distance normalized** (px) | 1.23 |
| **Distance** (px) | 1.19 |
| **Distance from answer normalized** (px) | -0.93 |
| **Interactions** | 0.65 |
| **Accumulated time** (s) | 0.61 |
| **Straightness** | -0.49 |
| **Pause before click** (s) | 0.31 |
| **Corrections between item** | -0.31 |
| **Distance from answer** (px) | -0.29 |
| **Hovered answers** | 0.28 |
| **<-Turns normalized** | 0.22 |
| **Pause before click normalized** (s) | -0.18 |
| **Corrections within item normalized** | -0.16 |
| **Correction time normalized** (s) | 0.15 |
| **Hover selected answer** | -0.15 |
| **Velocity** (px/s) | 0.13 |
| **Velocity normalized** (px/s) | -0.13 |
| **Corrections within item** | -0.06 |
| **Corrections between item normalized** | 0.05 |
| **Accumulated time normalized** (s) | 0.03 |
| **Correction time** (s) | 0.01 |

deciding which one to select. *Distance from answer*, on its turn, affects the result negatively, which means that, although individuals travel longer distances during moments of uncertainty, they tend to maintain the mouse cursor closer to the selected alternative. Probably this is influenced by consecutive questions with opposite (or very different) responses. That is, when a person moves directly from option 1 of an item to option 5 of the subsequent question, one of these items is associated to a large mean *distance from answer*. Nonetheless, in a question associated to uncertainty, where the travelled distance is long, this effect is attenuated.

Analysing the regression coefficient of *interactions*, it can be concluded that people visit more often items that arouse uncertainty. In these questions, individuals take longer times to answer (*accumulated time* has a positive and significant regression coefficient) and deviate more from the straight line trajectory between successive answers (*straightness* is associated to a negative coefficient).

It is surprising that the number of corrections influence negatively the result. This means that, when the number of corrections increases, the probability of identifying an uncertainty event decreases. Perhaps the great part of corrections result from distractions, which might be more recurrent in the absence of uncertainty, since the latter evokes more reflection.

### 6.2.6 Model Performance Evaluation

To evaluate the performance of the model, *sensitivity*, *specificity* and *f1 score* were accessed and their results are presented in table 6.6.

The *sensitivity* obtained was 0.78, which means that the instances of uncertainty are correctly classified in 78% of the times. The *specificity*, on its turn, achieved a value of 0.94 and hence the probability of a certainty event being correctly predicted is 94%.

As explained in section 5.2.7, the most appropriate measure to evaluate this model is *f1 score*. Using this metric, the estimated performance of the model was 0.81. Taking into account that uncertainty assessment concerns a subjective evaluation, the performance of the model is very good.

Table 6.6: Model performance evaluation measures.

| Sensitivity | Specificity | f1 score |
|---|---|---|
| 0.78±0.17 | 0.94±0.08 | 0.81±0.15 |

### 6.2.7 Uncertainty Statistical Analysis

Following the application of the model to all participants' questions, the percentage of questions associated to uncertainty was computed, which is illustrated in figure 6.14. The values ranged from 6.36% to 81.08% with a mean value and standard deviation of 28.08 ± 15.15 %.

Figure 6.15 shows the contrast of the mouse movements between the individuals with the minimum and maximum percentages of questions that evoked uncertainty. The behaviours are clearly different, where the travelled distance is much higher in the latter.

### 6.2.8 Relation between Uncertainty and Personality

Regarding the relationship between uncertainty and personality, it was expected a positive correlation between uncertainty and Neuroticism. Neurotic people have a tendency

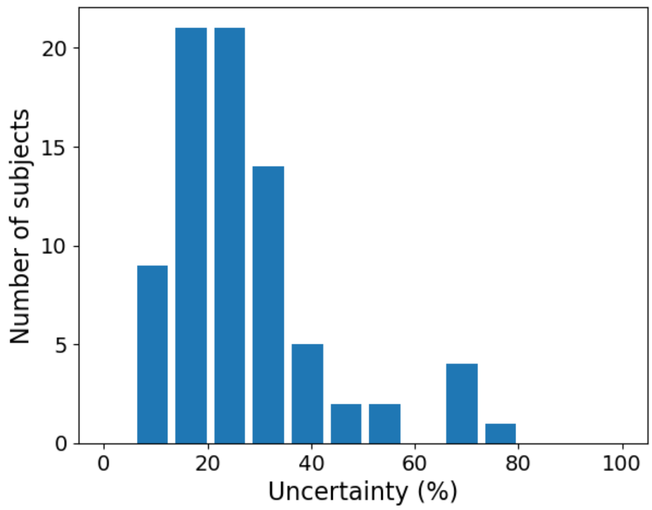
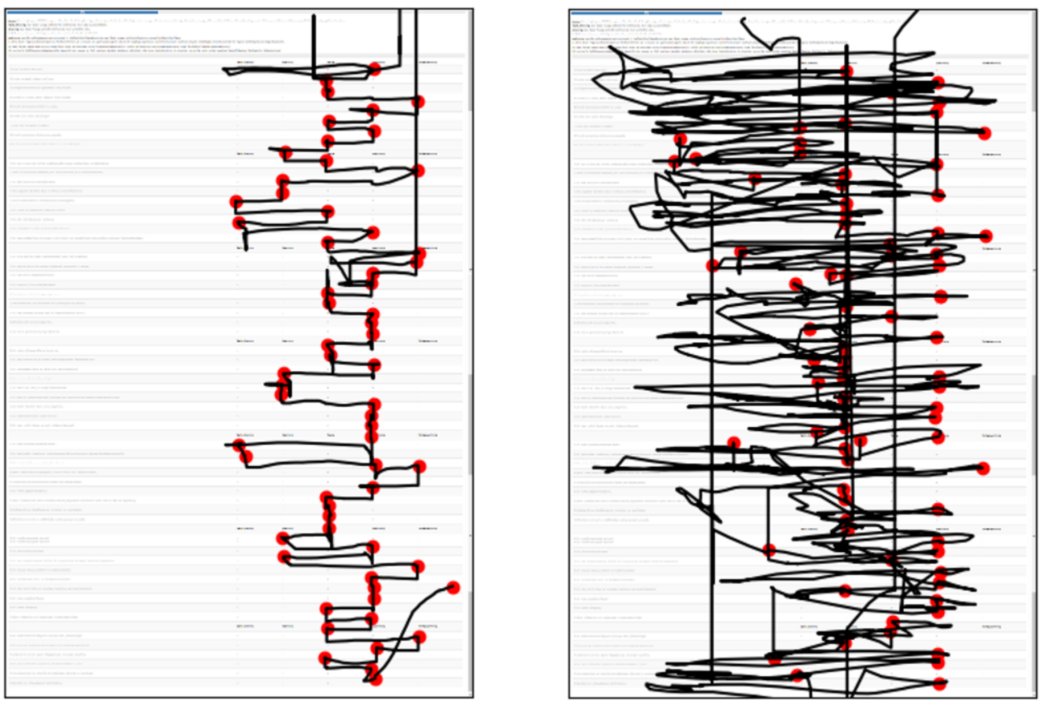



Figure 6.14: Percentage of questions associated to uncertainty.



(a)

(b)

Figure 6.15: Mouse movements of a questionnaire from the person with a) the minimum and b) the maximum percentage of uncertainty items.

to experience negative emotions and, consequently, it is likely that they face them during decision making tasks, motivating procrastination [1]. People with high scores in Neuroticism usually have low self-esteem [44], leading to difficulties in decision making tasks [1]. In addition, one of the traits of Neuroticism is anxiety [38], which affects negatively the decision making behaviour [81].

Extraversion, on its turn, is highly related to assertiveness [38], which might ease the decision making processes and result in a negative correlation between this dimension and uncertainty.

These hypotheses were supported by [1, 49], where it was reported a positive correlation between indecisiveness and Neuroticism and a negative correlation between indecisiveness and Extraversion.

Conscientiousness, on the one hand, is linked to self-discipline and a tendency to complete tasks, which relate negatively to complications during decisional tasks. This theory was presented and supported by [49]. On the other hand, people with high Conscientiousness are rigorous and perfectionists [82], which could contribute to a higher indecisiveness [1]. Consequently, either a positive or a negative correlation would make sense.

People high in Openness to Experience scale are tolerant to uncertainty [83], and this could possibly lead to a negative correlation coefficient.

Speaking of Agreeableness, there are not arguments concerning its association with uncertainty.

The Pearson correlation coefficients between each FFM scale and the percentage of uncertainty events are displayed in table 6.7.

Table 6.7: Correlation between uncertainty percentage and each FFM scale: Neuroticism (N), Extraversion (E), Conscientiousness (C), Agreeableness (A) and Openness to Experience (O).

| Correlation coefficient | **N** | **E** | **C** | **A** | **O** |
|---|---|---|---|---|---|
| **Uncertainty percentage** | -0.03 | -0.02 | 0.08 | 0.06 | 0.08 |

We can observe that the results have very low correlation and, accordingly, they don't support the hypotheses discussed above. This might have happened due to the nature of the task: probably people do not fully manifest their personality (with mouse movements) while responding to personality questionnaires since the task is not demanding enough.

## 6.3 Application to a Clinical Context

### 6.3.1 Population Description

The population covered 8 participants, 6 female and 2 male, with ages ranging from 27 to 33 years old. Nevertheless, only 7 of them answered the NEO-FFI questionnaire.

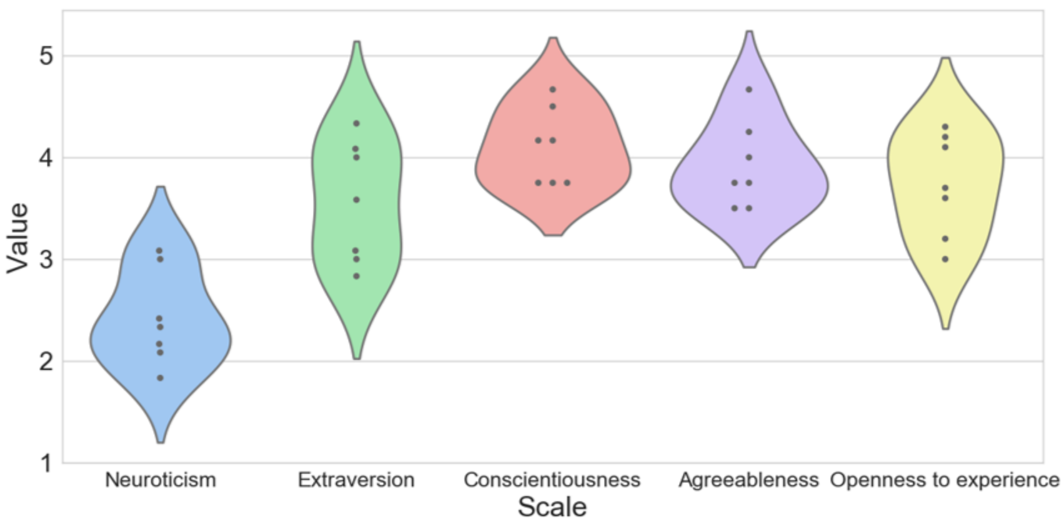



Figure 6.16: Violin plot with the personality questionnaire results. The grey dots represent the results of each person.

The violin plot presented in figure 6.16 illustrates the FFM questionnaire results of the 7 participants who answered it. The mean values and standard deviations were 2.42±0.36 for Neuroticism, 3.56±0.50 for Extraversion, 4.14±0.31 for Conscientiousness, 3.92±0.33 for Agreeableness and, finally, 3.73±0.40 for Openness to Experience.

### 6.3.2 Lost Data

Similarly to section 6.2.2, the mean percentage of lost samples from each case is displayed in table 6.8.

Table 6.8: Mean percentage of lost samples of each medical case.

| | **Lost samples (%)** |
|---|---|
| **Low conflict case** | 1.03 |
| **Moderate conflict case** | 0.04 |
| **High conflict case** | 0.03 |

From table 6.8, it can be noticed that the number of lost samples is negligible.

### 6.3.3 Uncertainty Statistical Analysis

The mean and standard deviation of the percentage of questions that evoked uncertainty was 30.00 ± 7.43 % with values ranging from 18.18% to 43.94%. Although the number of participants is much lower in this experiment, the mean uncertainty percentage is similar to the previous study (28.08%).

### 6.3.4 Comparison between Cases Complexity and Model Outcomes

In order to test the validity of the uncertainty model in this context, the percentage of items that aroused uncertainty (obtained with the model) in each case was compared to the cases complexity. Table 6.9 presents the mean, standard deviation, minimum and maximum values of the percentage of questions associated to uncertainty for each medical case.

Table 6.9: Percentage of items that evoked uncertainty of each medical case.

| | **Uncertainty (%)** | | |
|---|---|---|---|
| | **Low conflict case** | **Moderate conflict case** | **High conflict case** |
| **Mean** | 27.92 | 35.28 | 29.87 |
| **Standard deviation** | 8.10 | 5.81 | 5.26 |
| **Minimum** | 9.09 | 21.21 | 18.18 |
| **Maximum** | 51.51 | 45.45 | 40.91 |

Observing the results, it is possible to verify that the mean percentage of uncertainty items is lower in the high conflict case than in the moderate conflict case. This may seem contradictory, but it matches the anticipated assumptions. As it was exposed in section 4.3.1, the high conflict case is more severe and it comprises the most complex DDI, but the number of difficult items is higher in the moderate conflict case. Since these outcomes include the percentage of uncertainty questions and not their complexity, it was predicted that this percentage would be higher in the moderate conflict case. The minimum values are coherent with this as well.

Regarding the maximum values, the highest value is, surprisingly, from the low conflict case. Analysing all the data, it was verified that this happened due to the unique person who unexpectedly was more uncertain in the low conflict case than in the other two cases. This could be caused by a lower tiredness or a higher availability during the fulfillment of the low conflict case. It is important to note that the first case answered by this outlier was the low conflict case - possibly in the first case the physician took a long time and in the others the participant did not want to spend that amount of time, for example.

Even so, these outcomes support the validity of the model across different contexts. Nevertheless, it is necessary, in the future, to test it again with more participants and in other environments.

### 6.3.5 Relation between Uncertainty and Personality

Similarly to section 6.2.8, it was explored the relationship between the FFM's dimensions and the percentage of items that aroused uncertainty. It is not a good practice to employ the Pearson correlation coefficient in small datasets [84] and, for this reason, table 6.10 displays the results of the personality questionnaire (which range from 1 to 5) as well as the uncertainty percentage, which is ordered from the lowest to the highest value.

Table 6.10: Uncertainty percentage and FFM scales: Neuroticism (N), Extraversion (E), Conscientiousness (C), Agreeableness (A) and Openness to Experience (O).

| **Uncertainty (%)** | **N** | **E** | **C** | **A** | **O** |
|---|---|---|---|---|---|
| 19.19 | 3.00 | 3.00 | 3.75 | 4.00 | 4.10 |
| 20.20 | 1.83 | 4.08 | 4.50 | 4.25 | 3.20 |
| 23.48 | 2.08 | 4.33 | 4.67 | 4.67 | 4.30 |
| 25.76 | 2.17 | 3.08 | 3.75 | 3.50 | 4.20 |
| 28.79 | 2.33 | 3.58 | 4.17 | 3.75 | 3.60 |
| 32.32 | 3.08 | 2.83 | 3.75 | 3.75 | 3.00 |
| 42.42 | 2.42 | 4.00 | 4.17 | 3.50 | 3.70 |

It can be perceived that, apart from the first subject (and the last subject, but with a non-substantial deviation), uncertainty tends to increase with Neuroticism, in line with the arguments exposed in section 6.2.8. However, this relationship is weak and it would be necessary a higher amount of data to confirm it. Regarding the other scales, it can not be discerned any association.

Chapter 7

# Conclusions

This last chapter summarizes the work developed and its main results. Moreover, some suggestions of future work are provided.

## 7.1 General Results

The main objective of this dissertation was to construct a model that detects events of uncertainty using mouse cursor movements acquired during survey responses.

To begin, a comparison study between eye and mouse movements was developed to understand whether mouse tracking could replace eye tracking or not. To do so, it was necessary to calibrate the eye gaze data and, hence, a new calibration methodology was introduced. Regarding the relationship between eye and mouse cursor trajectories, it was demonstrated that, in events without uncertainty, they highly correlate. Nevertheless, when a person hesitates between different options, this correlation decreases. Even so, the benefits of choosing mouse tracking justify its use.

Subsequently, to build the model, several features were extracted from mouse tracking data. Some of them were more relevant to the classification process, and, assessing their importance, it was concluded that, in case of uncertainty while interacting with a computer, individuals increase the number of horizontal direction inversions with the cursor, the distance travelled by the mouse is higher but the mean distance from the selected answer is lower and the number of visits to the items that arouse uncertainty increases, as well as the time spent there.

Uncertainty is a subjective concept and, even so, the estimated performance of the created model was 0.81.

Following the construction of the model, this was applied to a medical decision making task in order to understand if it fits in divergent contexts. To test it, the percentage of events associated to uncertainty was computed for three cases with different degrees of severity and complexity, and the outcomes were consistent with the task design.

In order to explore if personal attributes influence people's reaction to uncertainty, the FFM scales were correlated with the percentage of uncertainty occurrences in both environments - during survey responses and in the clinical context. In the former, the correlations were insignificant, which can be related to the nature of the task. Nevertheless, in the latter, it was recognized a positive (but weak) relationship between uncertainty and Neuroticism. This association can be explained through the existing link between Neuroticism and procrastination [1], low self-esteem [44] and anxiety [38], which affect negatively decision making processes [1, 81].

## 7.2 Future Work

In order to improve and conclude this research work, some suggestions of future work are presented in this section.

First of all, the certainty and uncertainty instances used to train and test the model were obtained through visualization and subsequent subjective evaluation. Taking into account the available possibilities, it was the best solution. Nonetheless, an approach where the participants reveal their moments of uncertainty would be more correct and should be attempted in the future. Possibly, with this procedure, a broader classification could be applied - a three point scale instead of a binary classification, for example - in order to approximate the model to reality.

Additionally, the application of the model to a clinical context involved few participants and, accordingly, it should be extended to more subjects. Although the results matched the expected outcomes, it can not be concluded that the model is suitable in different contexts since the dataset was small. Furthermore, it would be interesting to validate the relationship between uncertainty and Neuroticism.

Considering that Indecisiveness Scale is widely used in psychology experiments, it would be relevant to search for a relation between uncertainty (obtained with the model) and indecisiveness.

Lastly, to better understand how uncertainty affects our daily lives, a relation between uncertainty and performance (in a professional decision making task, for example) should be explored.

# Bibliography


[1] A. Di Fabio, L. Palazzeschi, L. Asulin-Peretz, and I. Gati. "Career Indecision Versus Indecisiveness: Associations With Personality Traits and Emotional Intelligence." In: *Journal of Career Assessment* 21.1 (2013), pp. 42–56. doi: 10.1177/1069072712454698.

[2] A. Patalano, B. Juhasz, and J. Dicke. "The Relationship Between Indecisiveness and Eye Movement Patterns in a Decision Making Informational Search Task." In: *Journal of Behavioral Decision Making* 23 (2010), pp. 353–368. doi: 10.1002/bdm.661.

[3] R. O. Frost and D. L. Shows. "The Nature and Measurement of Compulsive Indecisiveness." In: *Behaviour Research and Therapy* 31.7 (1993), pp. 683–692. doi: 10.1016/0005-7967(93)90121-A.

[4] M. Chen, J. R. Anderson, and M. Sohn. "Correlation of eye / mouse movements on web browsing." In: *CHI '01 Extended Abstracts on Human Factors in Computing Systems*. 2001, pp. 281–282. doi: 10.1145/634067.634234.

[5] K. Rodden and X. Fu. "Exploring how mouse movements relate to eye movements on web search results pages." In: *Workshop on Web Information Seeking and Interaction at SIGIR*. 2007, pp. 29–32. url: http://www.rodden.org/kerry/wisi2007.pdf.

[6] D. Gassmann, M. Cheetham, K. Siebenhuener, B. M. Holzer, C. Meindl-Fridez, F. F. Hildenbrand, V. Virgini, M. Martin, and E. Battegay. "The multimorbidity interaction severity index (MISI) A proof of concept study." In: *Medicine* 96.8 (2017). doi: 10.1097/MD.0000000000006144.

[7] A. Dix. "Human-Computer Interaction." In: *Encyclopedia of Database Systems*. Ed. by L. Liu and M. T. Özsu. Boston, MA: Springer US, 2009, pp. 1327–1331. isbn: 978-0-387-39940-9. doi: 10.1007/978-0-387-39940-9_192.

[8] G. M. Olson and J. S. Olson. "Human-Computer Interaction: Psychological Aspects of the Human Use of Computing." In: *Annual Review of Psychology* 54.1 (2003), pp. 491–516. issn: 0066-4308. doi: 10.1146/annurev.psych.54.101601.145044.

[9] C. Voeffray. "Emotion-sensitive Human-Computer Interaction (HCI): State of the art - Seminar paper." In: *Seminar emotion recognition*. 2012, pp. 1–4. URL: https://diuf.unifr.ch/main/diva/sites/diuf.unifr.ch.main.diva/files/T6.pdf.

[10] G Sinha, R Shahi, and M Shankar. "Human Computer Interaction." In: *3rd International Conference on Emerging Trends in Engineering and Technology*. 2010, pp. 1–4. DOI: 10.1109/ICETET.2010.85.

[11] K. E. Pocius. "Personality factors in human-computer interaction: a review of the literature." In: *Computers in Human Behavior* 7 (1991), pp. 103–135. DOI: 10.1016/0747-5632(91)90002-I.

[12] *This is eye tracking*. URL: https://www.tobii.com/group/about/this-is-eye-tracking/ (visited on 08/13/2018).

[13] J. Zhu, E. Zhang, and K. Rio-Tsonis. "Eye Anatomy." In: *eLS* (2012). DOI: 10.1002/9780470015902.a0000108.pub2CITATIONS.

[14] *Diagram of the Eye - Lions Eye Institute*. URL: https://www.lei.org.au/services/eye-health-information/eye-diagram/ (visited on 09/21/2018).

[15] A. Poole and L. J. Ball. "Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects." In: *Encyclopedia of Human-Computer Interaction* (2005), pp. 211–219. URL: https://pdfs.semanticscholar.org/92bc/546258e9b6560cea225ca9f6745fa636ae6a.pdf.

[16] K. Harezlak, P. Kasprowski, and M. Stasch. "Towards accurate eye tracker calibration -methods and procedures." In: *Procedia Computer Science* 35.C (2014), pp. 1073–1081. DOI: 10.1016/j.procs.2014.08.194.

[17] *How do Tobii Eye Trackers work? - Learn more with Tobii Pro*. URL: https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/how-do-tobii-eye-trackers-work/ (visited on 08/13/2018).

[18] K. Rayner. "Eye movements in reading and information processing: 20 years of research." In: *Psychol Bulletin* 124.3 (1998), pp. 372–422. DOI: 10.1037//0033-2909.124.3.372.

[19] K. Krafka, A. Khosla, P. Kellnhofer, and H. Kannan. "Eye Tracking for Everyone." In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 2176–2184. DOI: 10.1109/CVPR.2016.239.

[20] T. W. Ng. "The optical mouse as a two-dimensional displacement sensor." In: *Elsevier* 107.1 (2003), pp. 21–25. ISSN: 09244247. DOI: 10.1016/S0924-4247(03)00256-5.

[21] N. N. A. Charniya and S. V. Dudul. "Simple low-cost system for thickness measurement of metallic plates using laser mouse navigation sensor." In: *IEEE Transactions on Instrumentation and Measurement* 59.10 (2010), pp. 2700–2705. ISSN: 00189456. DOI: 10.1109/TIM.2010.2045555.

[22] A. A. Calvo and S. Perugini. "Pointing Devices for Wearable Computers." In: *Advances in Human-Computer Interaction* (2014). DOI: http://dx.doi.org/10.1155/2014/527320.

[23] E. Hehman, R. M. Stolier, and J. B. Freeman. "Advanced mouse-tracking analytic techniques for enhancing psychological science." In: *Group Processes and Intergroup Relations* 18.3 (2015), pp. 384–401. DOI: 10.1177/1368430214538325.

[24] E. Arroyo, T. Selker, and W. Wei. "Usability tool for analysis of web designs using mouse tracks." In: *CHI '06 Extended Abstracts on Human Factors in Computing Systems*. 3. 2006, pp. 484–489. DOI: 10.1145/1125451.1125557.

[25] H. Gamboa and A. Fred. "A behavioral biometric system based on human computer interaction." In: *Proceedings of SPIE - The International Society for Optical Engineering* 5404 (2004), pp. 381–392. DOI: 10.1117/12.542625.

[26] D. Chudá and P. Krátky. "Usage of computer mouse characteristics for identification in web browsing." In: *Proceedings of the 15th International Conference on Computer Systems and Technologies - CompSysTech '14*. 2014, pp. 218–225. DOI: 10.1145/2659532.2659645.

[27] A. A. E. Ahmed and I. Traore. "A New Biometric Technology Based on Mouse Dynamics." In: *IEEE Transactions on Dependable and Secure Computing* 4.3 (2007), pp. 165–179. DOI: 10.1109/TDSC.2007.70207.

[28] M. Pusara and C. E. Brodley. "User re-authentication via mouse movements." In: *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security - VizSEC/DMSEC '04* (2004), pp. 1–8. DOI: 10.1145/1029208.1029210.

[29] I. H. Witten and E. Frank. *Data Mining - Practical Machine Learning Tools and Techniques*. Ed. by J. Gray. Second. San Francisco: Elsevier, 2005. ISBN: 0120884070.

[30] B. W. Balleine. "The Neural Basis of Choice and Decision Making." In: *The Journal of Neuroscience* 27.31 (2007), pp. 8159–8160. DOI: 10.1523/JNEUROSCI.1939-07.2007.

[31] K. Gok and N. Atsan. "Decision-making under stress and its implications for managerial decision-making: a review of literature." In: *International Journal of Business and Social Research* 06.03 (2016), pp. 38–47. DOI: 10.1504/IJEM.2003.003297.

[32] A. Riabacke. "Managerial Decision Making Under Risk and Uncertainty." In: *IAENG International Journal of Computer Science* 32.4 (2006). URL: http://www.iaeng.org/IJCS/issues{_}v32/issue{_}4/IJCS{_}32{_}4{_}12.pdf.

[33] K. Starcke and M. Brand. "Decision making under stress: A selective review." In: *Neuroscience and Biobehavioral Reviews* 36 (2012), pp. 1228–1248. DOI: 10.1016/j.neubiorev.2012.02.003.

[34] E. Rassin, P. Muris, I. Franken, M. Smit, and M. Wong. "Measuring general indecisiveness." In: *Journal of Psychopathology and Behavioral Assessment* 29 (2007), pp. 61–68. DOI: 10.1007/s10862-006-9023-z.

[35] G. Saucier and S. Srivastava. "What makes a good structural model of personality? Evaluating the big five and alternatives." In: *Handbook of Personality and Social Psychology: Vol 4. Personality Processes and Individual Differences*. Ed. by M. Mikulincer and P. Shaver. 2015. Chap. 13, pp. 283–305. ISBN: 9781433816994.

[36] B. W. Roberts and D. Mroczek. "Personality trait change in adulthood." In: *Current Directions in Psychological Science* 17.31 (2008), pp. 31–35. DOI: 10.1111/j.1467-8721.2008.00543.x.

[37] C. Kandler, J. Zimmermann, and D. A.N. P. Mcadams. "Core and Surface Characteristics for the Description and Theory of Personality Differences and Development." In: *European Journal of Personality* 28 (2014), pp. 231–243. DOI: 10.1002/per.1952.

[38] R. R. McCrae and O. P. John. "An introduction to the five-factor model and its applications." In: *Journal of Personality* 60.2 (1992), pp. 175–215. DOI: 10.1111/j.1467-6494.1992.tb00970.x.

[39] S. Rothmann and E. P. Coetzer. "The big five personality dimensions and job performance." In: *Personnel Psychology* 44.1-26 (1991). DOI: 10.1111/j.1744-6570.1991.tb00688.x.

[40] P. J. Corr and G. Matthews, eds. *The Cambridge Handbook of Personality Psychology*. Cambridge University Press, 2009, pp. 148–161. ISBN: 9780521862189.

[41] P. T. Costa and R. R. McCrae. "Four ways five factors are not basic." In: *Personality and Individual Differences* 13.6 (1992), pp. 653–665. DOI: 10.1016/0191-8869(92)90236-I.

[42] B. Esteves. "Personality assessment based on biosignals during a decision-making task." Master Thesis. Faculdade de Ciências e Tecnologia - Universidade Nova de Lisboa, 2017, p. 12.

[43] B. Rammstedt and O. P. John. "Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German." In: *Journal of Research in Personality* 41 (2007), pp. 203–212. DOI: 10.1016/j.jrp.2006.02.001.

[44] O. P. John, R. W. Robins, and L. A. Pervin. *Handbook of Personality Theory and Research*. Ed. by O. P. John, R. W. Robins, and L. A. Pervin. Third Edit. New York: The Guilford Press, 2008, pp. 159–181. ISBN: 9781593858360.

[45] L. Parks-Leduc, G. Feldman, and A. Bardi. "Personality Traits and Personal Values : A Meta-Analysis." In: *Personality and Social Psychology Review* 19.1 (2015), pp. 3–29. DOI: 10.1177/1088868314538548.

[46] A. J. Hornof and T. Halverson. "Cleaning up systematic error in eye-tracking data by using required fixation locations." In: *Behavior Research Methods, Instruments, and Computers* 34.4 (2002), pp. 592–604. DOI: 10.3758/BF03195487.

[47] P. Blignaut, M. Nystrom, K. Holmqvist, and R. Dewhurst. "Improving the accuracy of video-based eye-tracking in real- time through post-calibration regression." In: *Current Trends in Eye Tracking Research* (2014). DOI: 10.1007/978-3-319-02868-2_5.

[48] Q. Guo and E. Agichtein. "Towards predicting web searcher gaze position from mouse movements." In: *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems - CHI EA '10.* 2010, pp. 3601–3606. DOI: 10.1145/1753846.1754025.

[49] V. Germeijs and K. Verschueren. "Indecisiveness and Big Five personality factors: Relationship and specificity." In: *Personality and Individual Differences* 50 (2011), pp. 1023–1028. DOI: 10.1016/j.paid.2011.01.017.

[50] J. Page, M. A. Bruch, and R. F. Haase. "Role of perfectionism and Five-Factor model traits in career indecision." In: *Personality and Individual Differences* 45 (2008), pp. 811–815. DOI: 10.1016/j.paid.2008.08.013.

[51] J. A. Watson. "Indecisiveness: Measuring Individual Differences With Mouse Tracking Software." Doctoral dissertation. Ball State University, 2015.

[52] I. K. Schneider, F. van Harreveld, M. Rotteveel, S. Topolinski, J. van der Pligt, N. Schwarz, and S. L. Koole. "The path of ambivalence: tracing the pull of opposing evaluations using mouse trajectories." In: *Frontiers in Psychology* 6.996 (2015), pp. 1–12. DOI: 10.3389/fpsyg.2015.00996.

[53] C. Cepeda, J. Rodrigues, M. C. Dias, D. Oliveira, D. Rindlisbacher, M. Cheetham, and H. Gamboa. "Mouse tracking measures and movement patterns with application for online surveys." In: *Cross Domain Conference for Machine Learning and Knowledge Extraction (CD-MAKE 2018)* 11015 (2018), pp. 28–42. DOI: https://doi.org/10.1007/978-3-319-99740-7_3.

[54] M. Zushi, Y. Miyazaki, and K. Norizuki. "Web application for recording learners' mouse trajectories and retrieving their study logs for data analysis." In: *Knowledge Management and E-Learning: An International Journal* 4.1 (2012), pp. 37–50. ISSN: 20737904. URL: https://www.researchgate.net/publication/268064257{_}Web{_}application{_}for{_}recording{_}learners'{_}mouse{_}trajectories{_}and{_}retrieving{_}their{_}study{_}logs{_}for{_}data{_}analysis.

[55] F. G. Conrad, M. F. Schober, and T. Coiner. "Bringing Features of Human Dialogue to Web Surveys." In: *Applied Cognitive Psychology* 21 (2007), pp. 165–187. DOI: 10.1002/acp.1335. arXiv: NIHMS150003.

[56] R. Horwitz, F. Kreuter, and F. Conrad. "Using Mouse Movements to Predict Web Survey Response Difficulty." In: *Social Science Computer Review* 35.3 (2016). DOI: 10.1177/0894439315626360.

[57] L. Pilárik and I. Sarmány-Schuller. "Personality predictors of decision-making of medical rescuers." In: *Studia Psychologica* 53.2 (2011), pp. 175–184. URL: http://www.scopus.com/inward/record.url?eid=2-s2.0-79960022307{\&}partnerID=tZOtx3y1.

[58] M. T. Buelow and J. A. Suhr. "Construct validity of the Iowa gambling task." In: *Neuropsychology Review* 19 (2009), pp. 102–114. DOI: 10.1007/s11065-009-9083-4.

[59] *Eye Tracking Solutions by SMI*. URL: https://www.smivision.com/ (visited on 08/10/2018).

[60] *Neurobehavioral Systems*. URL: https://www.neurobs.com/ (visited on 08/10/2018).

[61] *Python*. URL: https://www.python.org/ (visited on 08/10/2018).

[62] *PyCharm: Python IDE for Professional Developers by JetBrains*. URL: https://www.jetbrains.com/pycharm/ (visited on 01/30/2018).

[63] *NumPy*. URL: http://www.numpy.org/ (visited on 08/10/2018).

[64] *Matplotlib: Python plotting — Matplotlib 2.2.2 documentation*. URL: https://matplotlib.org/ (visited on 08/10/2018).

[65] *SciPy library*. URL: https://www.scipy.org/scipylib/index.html (visited on 08/11/2018).

[66] *Python Data Analysis Library — pandas*. URL: https://pandas.pydata.org/ (visited on 08/10/2018).

[67] *BeGaze 2.3 Manual*. 2009.

[68] P. Sedgwick. "Pearson's correlation coefficient." In: *BMJ* (2012). DOI: 10.1136/bmj.e4483.

[69] *scikit-learn: machine learning in Python — scikit-learn 0.19.2 documentation*. URL: http://scikit-learn.org/stable/ (visited on 08/11/2018).

[70] *seaborn: statistical data visualization — seaborn 0.9.0 documentation*. URL: https://seaborn.pydata.org/ (visited on 08/29/2018).

[71] J. Huang and R. White. "User See, User Point: Gaze and Cursor Alignment in Web Search." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2012)*. 2012, pp. 1341–1350. URL: https://jeffhuang.com/Final{_}GazeCursor{_}CHI12.pdf.

[72] N. Zheng, A. Paloski, and H. Wang. "An efficient user verification system via mouse movements." In: *Proceedings of the 18th ACM conference on Computer and communications security - CCS '11* (2011), pp. 139–150. DOI: 10.1145/2046707.2046725.

[73] H. S. Hou and H. C. Andrews. "Cubic Splines for Image Interpolation and Digital Filtering." In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26.6 (1978), pp. 508–517. ISSN: 0096-3518. DOI: 10.1109/TASSP.1978.1163154.

[74] L. A. Shalabi, Z. Shaaban, and B. Kasasbeh. "Data Mining: A Preprocessing Engine." In: *Journal of Computer Science* 2.9 (2006), pp. 735–739. ISSN: 1549-3636. DOI: 10.3844/jcssp.2006.735.739.

[75] P. K. Tan, T. J. Downey, E. L. Spitznagel, P. Xu, D. Fu, D. S. Dimitrov, R. A. Lempicki, B. M. Raaka, and M. C. Cam. "Evaluation of gene expression measurements from commercial microarray platforms." In: *Nucleic Acids Research* 31.19 (2003), pp. 5676–5684. DOI: 10.1093/nar/gkg763.

[76] S. Sperandei. "Understanding logistic regression analysis." In: *Biochemia Medica* 24.1 (2014), pp. 12–18. DOI: 10.11613/BM.2014.003.

[77] H. A. Park. "An introduction to logistic regression: From basic concepts to interpretation with particular attention to nursing domain." In: *Journal of Korean Academy of Nursing* 43.2 (2013), pp. 154–164. DOI: 10.4040/jkan.2013.43.2.154.

[78] C. R. Shalizi. *Advanced Data Analysis from an Elementary Point of View*. 2018, pp. 254–258. URL: http://www.stat.cmu.edu/{~}cshalizi/ADAfaEPoV/ADAfaEPoV.pdf.

[79] T. Saito and M. Rehmsmeier. "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets." In: *PLOS ONE* 10.3 (2015), pp. 1–21. DOI: 10.1371/journal.pone.0118432.

[80] Y. Sun, M. S. Kamel, A. K. Wong, and Y. Wang. "Cost-sensitive boosting for classification of imbalanced data." In: *Pattern Recognition* 40 (2007), pp. 3358–3378. DOI: 10.1016/j.patcog.2007.04.009.

[81] K. A. Byrne, C. D. Silasi-Mansat, and D. A. Worthy. "Who chokes under pressure? The Big Five personality traits and decision-making under pressure." In: *Personality and Individual Differences* 74 (2015), pp. 22–28. DOI: 10.1016/j.paid.2014.10.009.

[82] J. Stoeber, K. Otto, and C. Dalbert. "Perfectionism and the Big Five: Conscientiousness predicts longitudinal increases in self-oriented perfectionism." In: *Personality and Individual Differences* 47.4 (2009), pp. 363–368. DOI: 10.1016/j.paid.2009.04.004.

[83] N. Nicholson, E. Soane, M. Fenton-O'Creevy, and P. Willman. "Personality and domain-specific risk taking." In: *Journal of Risk Research* 8.2 (2005), pp. 157–176. ISSN: 1466-4461. DOI: 10.1080/1366987032000123856.

[84] D. G. Bonett and T. A. Wright. "Sample size requirements for estimating Pearson, Kendall and Spearman correlations." In: *Psychometrika* 65.1 (2000), pp. 23–28. DOI: 10.1007/BF02294183.

# Appendix A

# Publications

This appendix includes two publications.

The first is entitled "Mouse tracking measures and movement patterns with application for online surveys", and it was submitted during this research work. This article was accepted to "Cross Domain Conference for Machine Learning and Knowledge Extraction"(CD-MAKE 2018).

# Mouse tracking measures and movement patterns with application for online surveys

Catia Cepeda ✉[1,2], Joao Rodrigues[2], Maria Camila Dias[2], Diogo Oliveira[2], Dina Rindlisbacher[1,3], Marcus Cheetham[1,3] and Hugo Gamboa[2]

[1] Department of Internal Medicine, University Hospital Zurich, Zurich, Switzerland
[2] LIBPhys (Laboratory for Instrumentation, Biomedical Engineering and Radiation Physics), Faculdade de Cincias e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal
c.cepeda@campus.fct.unl.pt
[3] University Research Priority Program "Dynamics and Healthy Aging", University of Zurich, Zurich, Switzerland

**Abstract.** There is growing interest in the field of human-computer interaction in the use of mouse movement data to infer e.g. user's interests, preferences and personality. Previous work has defined various patterns of mouse movement behavior. However, there is a paucity of mouse tracking measures and defined movement patterns for use in the specific context of data collection with online surveys. The present study aimed to define and visualize patterns of mouse movements while the user provided responses in a survey (with questions to be answered using a 5-point Likert response scale). The study produced a wide range of different patterns, including new patterns, and showed that these can easily be distinguished. The identified patterns may - in conjunction with machine learning algorithms - be used for further investigation toward e.g. the recognition of the user's state of mind or for user studies.

**Keywords:** Knowledge extraction · Mouse behavior patterns · Mouse tracking · Human-computer interaction · User · Survey

## 1 Introduction

A multitude of human factors influences human-computer interaction (HCI) (e.g., [18]). The influence on HCI of individually stable patterns of thinking, feeling and behavior is of longstanding interest (e.g., [9, 21]), as this often reflects underlying interests, preferences and personality. Making decisions is a complex cognitive and affective process [8, 13]. Understanding user behavior in the context of decision-making has increasingly attracted attention in HCI research [10, 19].

Pointer tracking refers to the recording of users' mouse cursor positions, used, for example, to capture the mouse movement trajectories for the purpose of further analysis. Data acquisition of mouse cursor positions has the advantage of being cheap, easy to implement and is already integrated in the use of the computer.

The present study aimed to identify patterns of mouse movements while the users give input in an online survey. These mouse movement patterns are potentially relevant as a means to understanding the user, such as in terms of the user's patterns of decision uncertainty.

Given the relative paucity of mouse tracking measures and mouse movement patterns in the literature, we present a new set of mouse behavior patterns that could potentially be combined with machine learning algorithms as a means to capturing information [14] about stable patterns of thinking, feeling and behavior of the user.

### 1.1 Related work

Eye tracking systems are used in HCI research since mid-1970s [22]. The data structure is similar to that of mouse movements (x and y positions in screen over time). In fact, a wide range of eye movement behaviors have been associated with mouse movements behaviors. There is also multimodal data acquisition devices available, such as *Tobii* and *SensorMotoric Instruments* (SMI) systems, that allow concurrent measurement of eye and mouse movement behavior.

For instance, *Tobii* permits eye tracking and analysis of eye sampling behavior while the user observes and interacts with web pages [4]. This system also enables concurrent acquisition of video, sound, key-strokes and mouse clicks. Analyses include a range of measures such as mouse movement velocity, and can visualize results using various methods, such as heat maps. The analyses of different modalities may also be combined in order to assess, for example, the time from the first fixation to a particular target until the user clicks on the same target (or the number of clicks on the target).

SMI [2] also provides behavioral and gaze analysis software for research in the fields of reading research, psychology, cognitive neuroscience, marketing research and usability testing. While this system only processes eye and head tracking data, it has the advantage of allowing the analyzes of several subjects simultaneously. This permits analysis, for example, of the hit ratio, that is the relative number of subjects in the sample that fixated at least once on the target.

Although eye tracking systems have a comparatively long history, the field of mouse tracking had developed several interesting approaches for mouse movements analysis. This largely relates to web pages usability testing in order to improve the user experience [1, 3, 5–7, 15], but others extract data from the mouse coordinates, such as path distance, time measures and mouse clicks in order to study user's behavior rather than the web design itself.

For instance, *Revett et al.* and *Hugo et al.* [11, 23] propose the biometric identification of the user based only on mouse or pointer movements. Another approach, led by *Khan et al.* [17], related the mouse behavior patterns with personality. In *Pimenta et al.* mental fatigue has been detected by means of mouse movements [20], while *Hibbel et al.* related movements to emotions [12, 26].

Other measures and movements patterns have also been used in behavior studies. In 2006, *Arroyo et al.* described mouse behaviors in the context of websites, reporting user behavior that consists of a long pause next to text or a blank space, followed by a fast and direct movement towards a link [6]. *Arroyo et al.* also examined hesitation patterns and random movements, while *Huang et al.* compared clicks and hover distributions, unclicked hovers and abandonments [15].

*Seelye et al.* used the deviation of the movement in relation to a straight path and the time between the two targets to distinguish older adults with and without mild cognitive impairment (MCI). They found that more curved or looped mouse movements and less consistency over time are more closely correlated with MCI subjects [24].

*Yamauchi et al.* focused on two trajectory measures from mouse cursor to detect user emotions. They defined attraction as the area under the curve from the starting position to the end position and zigzag as the number of direction changes during the movement. A statistical model build with these trajectory measures could predict nearly 10%-20% of the variance of positive affect and attentiveness ratings [25].

*Arapakis et al.* used a large number of measures to predict user engagement, as indicated by, for example, attention and usefulness [5]. The set of measures included the most common distance and time measures and also measures related to the target, for instance, the number of movements toward and away from the target, or the number of hovers over the target compared with around the target.

More recently, *Katerina et al.* used a wide set of measures, including mouse and keyboard measures [16]. Their objective was to examine the relationship between the measures extracted from mouse and keystroke and end-user behavioral measures. Two examples of measures examined in terms of mouse movements are the number of mouse long pauses and the number of clicks in the end of direct mouse movements. From keystroke dynamics one example of a measure done was the time elapsed between key press and key release.

To the best of our knowledge, no previous studies have reported mouse movements during data collection using online surveys.

## 2 Study Design

### 2.1 Participants and procedure

N=119 volunteers recruited via a pool of test participants and students of the University of Zurich and of the ETH Zurich participated in this study. The participants were between 20 and 52 years old (M=25.4; SD=5.4; 18 male). All participants were native or fluent speakers of Standard German. Written informed consent was obtained before participation, according to the guidelines of the Declaration of Helsinki.

### 2.2 Data Acquisition Architecture

In this study, the data resulted from the interaction of the user with the web browser while completing an online survey, which was programmed to send the data to a server machine via AJAX, where it is finally recorded as a file in a data base.

The results of the survey are also saved on the database, although in this case via the Survey Management System using *PHP*. Therefore, if needed, these results could be accessed as well.

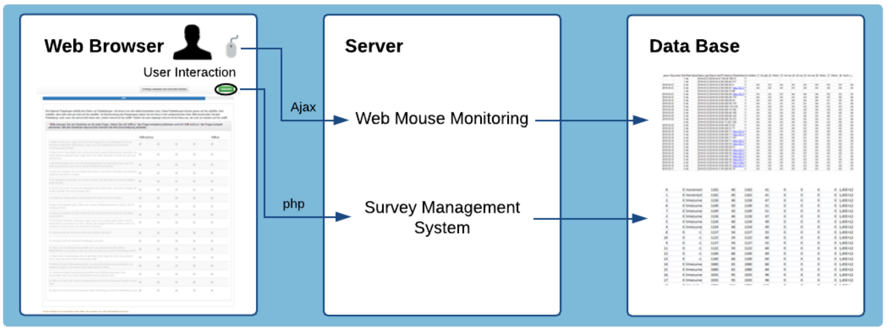



**Fig. 1.** Architecture.

### 2.3 Data Collection

The pointer movement is recorded by a server, which creates a report file with relevant recorded data: frame number; event type (represented by 0 when a movement is verified and 1 when the mouse button is pressed down); question number if hovered; answer number if hovered; x and y screens position (in pixels) and time stamp. The name of the file includes the IP address, the survey ID and the step of the questionnaire.

The online survey is constructed using a freely available software survey tool on the web. The online survey presents a sequence of statements and the answers are 5-point Likert-type scale. The results from the survey could be returned to a csv file.

### 2.4 Data cleaning

To ensure correct formatting and processing of data from the server file, a validation procedure is applied as a first step. This validation procedure ignores data acquired with touch screen devices, reorder the data by time, join different files from the same questionnaire and detects how many samples are lost.

## 3 Behavioral Patterns Description

The data acquired with the LimeSurvey contains information about the mouse position with and without scroll in pixels. This data is first interpolated with equal time interval between samples in order to retrieve the correct information from it. With the mouse position pre-processed and the other information delivered by the LimeSurvey, several measures from temporal, spatial and contextual domains can be derived.

In this study, these measures are essential to compute several of the behavioral patterns described further.

### 3.1 Overview Pattern

A behavior that can be found in some subjects in participating/answering surveys regards getting an overall idea of the number of questions, the length of the survey or the types of questions. This behavior is characterized by, at the beginning of the survey, scrolling the cursor over a wide area in direction to the bottom of the survey getting an overview of it. In figure 2 it is represented the mouse y coordinate represented over time, which makes it easy to observe this behavior. The first question are at the top of the plot (small y values) and, moving forward through the next questions, the y increases. At the beginning of the questionnaire, this subject goes to the end of the survey and then comes back to the first questions. This behavior also occurs after one minute and two minutes of interaction, but never so far as the first time.

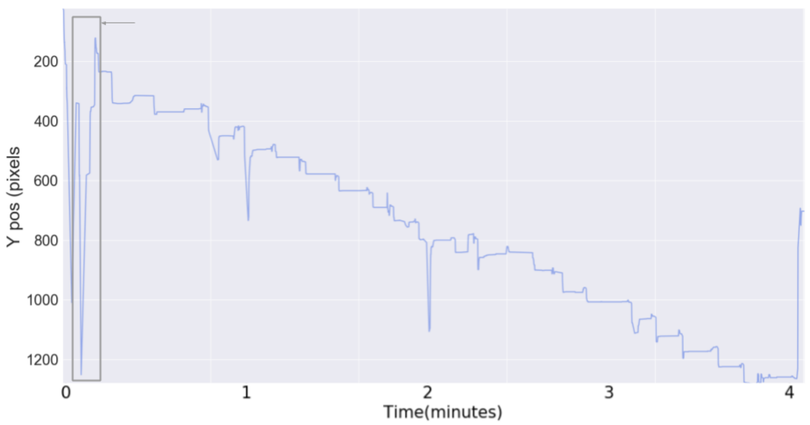



**Fig. 2.** Representation of the y-axis mouse movement over time. The rectangle area corresponds to an overview pattern.

### 3.2 Fast Decision Pattern

While some people take a long time to answer the questions, others are very fast. It is possible to find both behaviors, that we call Fast Decision Patterns, which are represented in figure 3. Both plots represent the question where the mouse is located over time and, as it is possible to observe, the subject at the top is much faster than the bottom subject, taking one and a half less minutes to answer the same questionnaire.

The work of Arroyo et al. [6] analyzed fast movements towards a target.

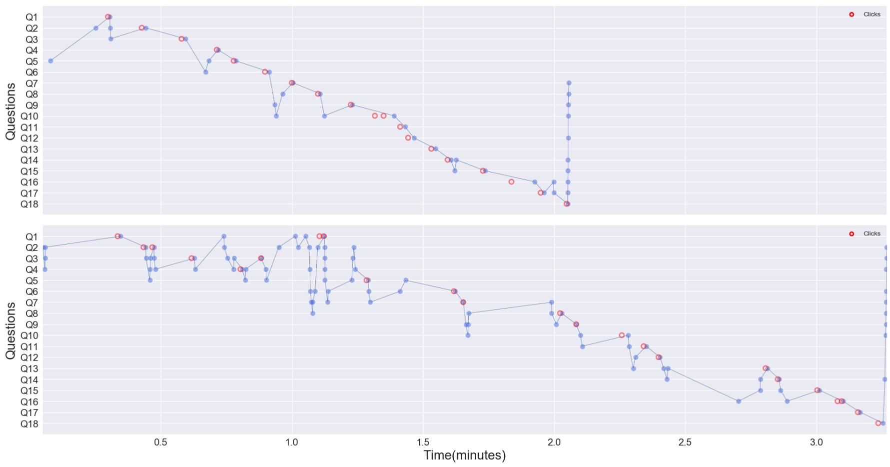



**Fig. 3.** Representation of the questions where the mouse is located over time for two different subjects. The subject at the top takes around two minutes to answer the questionnaire while the subject at the bottom almost four minutes.

### 3.3 Revisit Pattern

A typical behavior of the subject that can be found in the survey context, is to revisit prior questions after some time of having answered. In Figure 4 the user has revisited a prior answer (from question 14 to question 3) which was at the top of the survey. Interestingly, after answering the first time to the question 3, this subject responded to question 4 and came back to question 3, having changed three times the option previously answered. The revisit was around three minutes after these changes.

The analysis done by SMI [2] considers a similar metric with eye movements for a group evaluation: the average number of glances into the target.

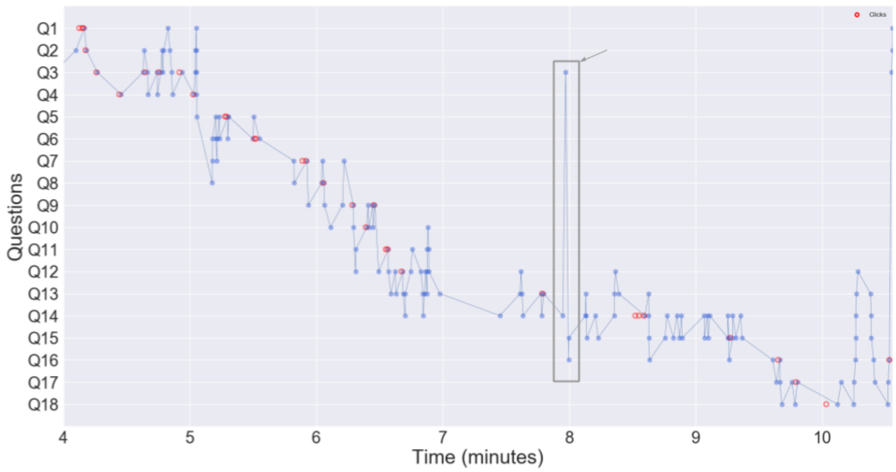



**Fig. 4.** Representation of the questions where the mouse is located over time. The red circles represent the mouse clicks. The rectangle area corresponds to a revisit behavior.

### 3.4 Skips Pattern

When answering the survey some subjects would not have a linear behavior of following the natural order of questions. In fact, some subjects would skip questions and answer in an unnatural order. In Figure 5, it is represented the questions answered over time. It is observed that the user does not take a linear approach in completing the survey, after answering question two, the subject starts to answer from question 14 to the previous questions. When the user is back to question 3, goes again to the end and answer question 18 until question 15.

### 3.5 Hover Pattern

In the context of the survey, a typical behavior found on certain users is hovering multiple available options before selecting their final answer. In Figure 6, two different users are compared in their survey completion. The flow chart indicates the way each user behave by indicating in which options they kept their mouse. Each blue circle is a selectable option to answer the corresponding question, which the user hovered. The size of the circle is proportional to the time spent on that option.

As can be seen on Figure 6, the user on the right (b) has more hovered areas (specially highlighted areas) than the user represented on the left (a).

Although *Tobii* [4] is an eye tracking system, it considers the number of fixations before fixating on the target, which is similar to what we are suggesting

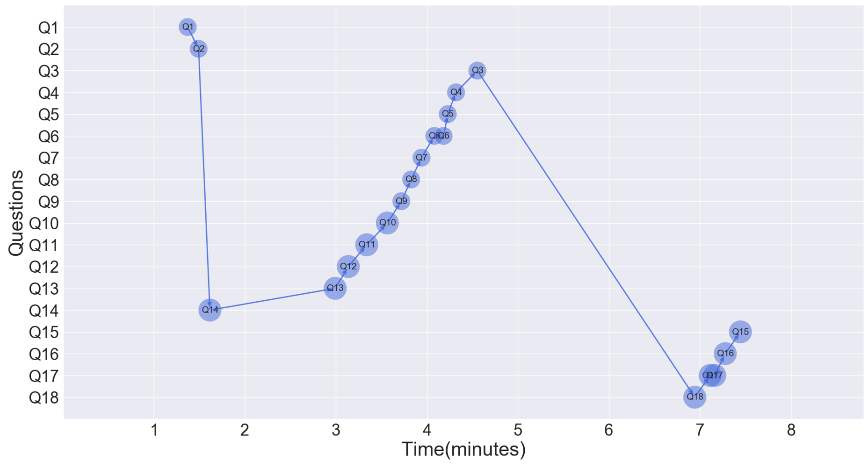



**Fig. 5.** Representation of the questions answered over time. This user is an example of skips behavior.

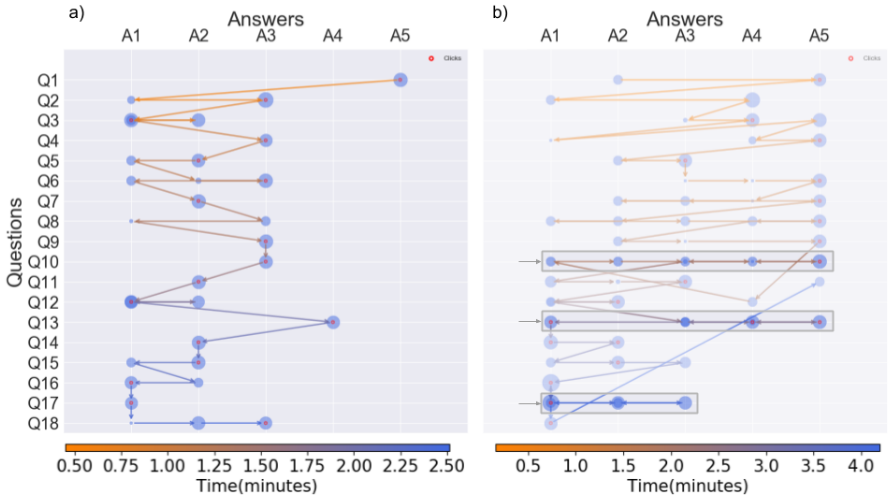



**Fig. 6.** Chart flow of two different users in answering the survey. The y-axis represents the question number and the x-axis is the option answer number. Clicks are depicted as red circles. The color-bar shows the flow of time used by the arrows that point the flow of the user's behavior. a) The user has a less representative hovering behavior. b) The highlighted areas show the hovering behavior of the user.

here. Previous studies also includes hover patterns in mouse movements analysis. *Katerina et al.* [16] considered the number of mouse hovers that turned into mouse clicks and *Arapakis et al.* [5] compared between hovering the area of interest in relation to other areas. Also *Huang et al.* [15] analyzed the hover distributions and clicks to verify the number of search results hovered before the user clicks.

### 3.6 Hover Reading Pattern

During the completion of questionnaires, the questions have some text in the left border which can be read in several ways. We found two distinct patterns: some people move the mouse to the text area, while reading the question, while others just move the mouse around the answers area. One example of each behavior are shown in figures 7 and 8, for the first it is evident that for each item the subject is hovering the text of the question before choosing an answer. That is not verified in the second, that only moves the mouse around the answers.

The computational process of this behavior is quite easy, the survey software has a tool in which the width of text of the question can be defined. Knowing that, the x mouse coordinates can be associated to questions or answers area.

**Fig. 7.** Representation of the mouse movements (in blue) of a subjects that moves the mouse to the question text while reading it. The mouse clicks are represented by a red circle.

### 3.7 Inter Item Pattern

The distance and time taken between the answered choice and enter the next question could be different from person to person. The time and distance are highly correlated and define the same kind of behaviors. However, some more specific patterns can be highlighted, for instance, the subject can take more time because it was moving slower, or because it was moving a lot, even if quickly. Therefore it is important to individualize these measures. In figure 9 it is

**Fig. 8.** Representation of the mouse movements (in blue) of a subjects that keeps the mouse around the answers area, even when reading the question. The mouse clicks are represented by a red circle.

presented four possible behaviors. Considering that the color intensity depends on the velocity (more intense for higher speeds), the a) and b) present short distance inter items, being b) much faster than a), while c) and d) present long distances inter item, being d) much faster than c). Here although a) and c) have very different distances, the speed of movement is similar. The same is true to b) and d).

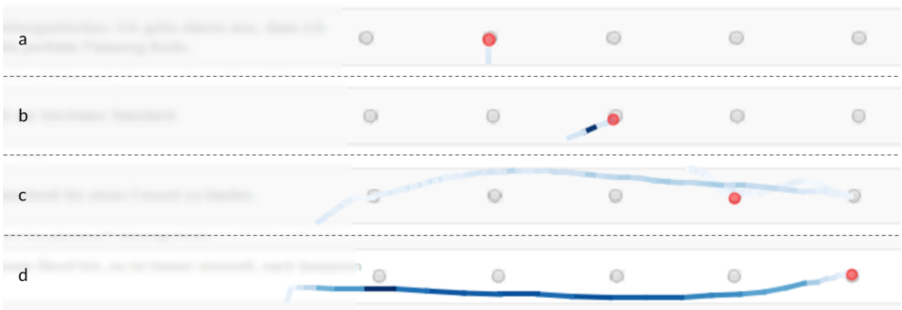



**Fig. 9.** Representation of the mouse movement in the survey context considering only the inter item interval. The color of the line corresponds to the velocity of the movement (color more intense for higher velocity). In a) there is an example of short distance but low speed, in b) short distance and high speed, in c) long distance and low speed and d) long distance and high speed.

### 3.8 Long Pauses Pattern

Long pauses correspond to mouse movements at the same place (x and y coordinates) for a long period of time. This can be observed in figure 10 in which orange circles represent long pauses while answering the survey questions. The longer the pauses, the larger the circles.

Multiple studies considered the number and time of long pauses [6, 16, 24]

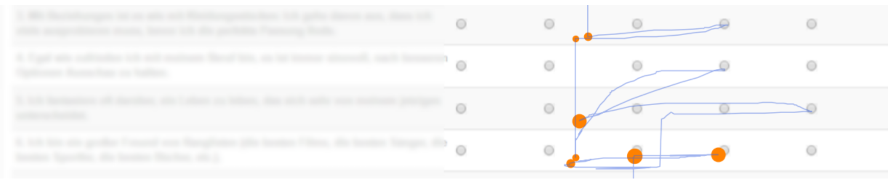


**Fig. 10.** Representation of long pauses pattern in mouse movement. In orange are presented circles that are larger according to the time paused.

### 3.9 Straight and Curvy Pattern

Straight patterns are characterized by a direct or straight line in direction to a target. This pattern indicates that a target has been spotted and the subject decided to move the cursor towards it. The opposite behavior is the curvy pattern, characterized by more curved movements. Comparing figure 11 with figure 12 it is possible to detect a huge differences in the way they move the mouse.

The studies from *Katerina et al.* [16] and *Seelye et al.* [24] had these patterns into consideration, having compared more straight or curved movements.

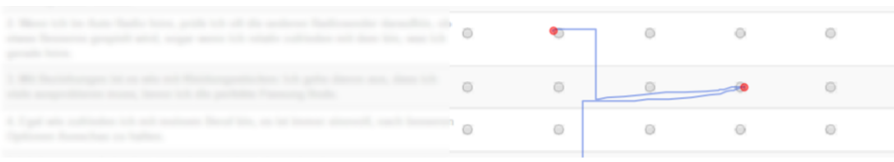


**Fig. 11.** Representation of straight patterns with the mouse. The red circles correspond to mouse clicks.

### 3.10 <-turn Pattern

While making a decision, sometimes the mouse movement nearly inverts its direction, this pattern has been called <-turn. Figure 13 presents this behavior

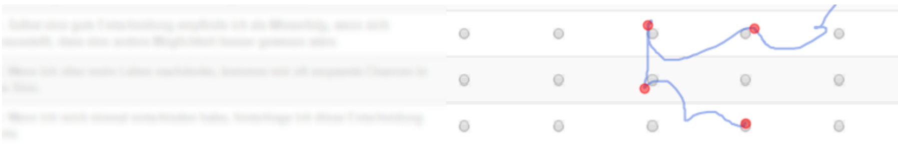


**Fig. 12.** Representation of curvy patterns with the mouse. The red circles correspond to mouse clicks.

two times during the choice of a question answer. To compute this behavior it should be detected angles close to 180 degrees in change of direction within the same item.

*Yamauchi et al.* [25] analyzed a similar pattern considering direction change.

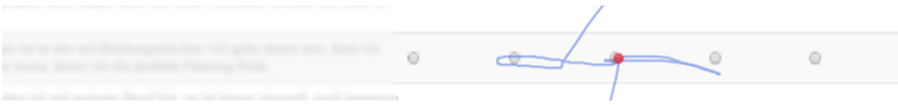


**Fig. 13.** Representation of <-turn pattern. In blue is presented the mouse movement and in red the mouse click.

### 3.11 Random Movements

While some movements are spontaneous and have an inner purpose, others might just be unconscious and have no specific intention. The latter patterns are described as random movement patterns and are characterized by a large number of movements confined in a non-interest area for a short time, as shown in figure 14.

This behavior was briefly described by [6], however they do not present a visualization example or a way to compute random movements.

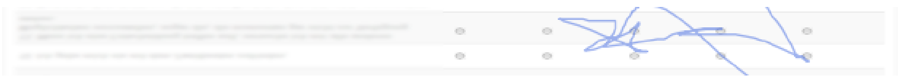


**Fig. 14.** Representation of a random movement made by the mouse cursor.

### 3.12 Loop Pattern

In the category of random movements, a pattern that can be found is characterized by a turn of more than 360, which can be defined as a loop and observed in figure 15. This behavior was previously considered by *Seelye et al.* [24] that calculated the number of looped mouse movements.

**Fig. 15.** Representation of a loop pattern.

## 4 Conclusion

This study demonstrates the use of mouse tracking measures and movement patterns in the specific context of online survey-based data collection. The survey consisted of several questions, each to be answered using a 5-point Likert response scale. Using only the mouse movements data, we show that it is possible to extract a wide range of different behaviors. The results also show the behavior patterns can easily be distinguished by mere visual inspection.

Although some of the behavioral patterns have already been reported in other studies (e.g., [6, 16, 24]), none were used in the context of surveys. Given that this is a completely different task situation with different task requirements, the proposed patterns require a different interpretation. This work delivered also new patterns of movement that were not reported in the previous literature, contributing therefore to the current state of the art.

It is possible to group several of these patterns according to their potential explanation. There are patterns that might be associated with personality traits, decision confidence or decision difficulty, but this awaits further investigation. For example, overview, fast decision, skips, straight and curvy, inter items intervals and long pauses could indicate personal characteristics and some users would follow the questions in an orderly and sequential manner, while others would first get an overall picture of the survey questions and then answer (overview pattern). Fast decisions could be related to confidence, and decision difficulty could be associated with hover pattern and <-turn.

Concerning the hover reading pattern, the users that move the mouse to the question text while reading it are less goal-oriented than those who just move the cursor directly to the next question. Whether this is so requires further investigation. If this is the case, it is also possible the first group of users could reveal a higher correlation between mouse and eye movements.

## 5 Future Work

As a first step after this work, it would be interesting to create metrics that express each of the patterns extracted. Consistent with other studies, we will progress in order to apply machine learning techniques to infer personality and states of mind from mouse movements data.

The recognition of these patterns in more complex contexts could be applied to improve the usability of websites and create an adapted design and contents according with user preferences.

Another application area is clinical/ergonomics field, for example to recognize mental fatigue or even mental diseases by studying the cognitive state of the subject given that users state of mind could be directly associated with a conjunction of behaviors.

## References

1. Clicktale, https://www.clicktale.com/
2. Eye Tracking Solutions by SMI, https://www.smivision.com/
3. Inspectlet - Website Heatmaps, Session Recording, Form Analytics, https://www.inspectlet.com/
4. Eye tracking technology for research - Tobii Pro (2015), https://www.tobiipro.com/
5. Arapakis, I., Leiva, L.A.: Predicting User Engagement with Direct Displays Using Mouse Cursor Information. In: Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval - SIGIR '16. pp. 599–608. ACM Press, New York, New York, USA (2016). https://doi.org/10.1145/2911451.2911505, http://dl.acm.org/citation.cfm?doid=2911451.2911505
6. Arroyo, E., Selker, T., Wei, W.: Usability tool for analysis of web designs using mouse tracks. In: CHI '06 extended abstracts on Human factors in computing systems - CHI EA '06. p. 484. ACM Press, New York, New York, USA (2006). https://doi.org/10.1145/1125451.1125557, http://dl.acm.org/citation.cfm?doid=1125451.1125557
7. Atterer, R., Wnuk, M., Schmidt, A.: Knowing the user's every move. In: Proceedings of the 15th international conference on World Wide Web - WWW '06. p. 203. ACM Press, New York, New York, USA (2006). https://doi.org/10.1145/1135777.1135811, http://portal.acm.org/citation.cfm?doid=1135777.1135811
8. Damasio, A.R.: Descartes' error : emotion, reason, and the human brain. New York (1994)

9. Dillon, A., Watson, C.: User analysis in HCI - The historical lessons from individual differences research. International Journal of Human Computer Studies **45**(6), 619–637 (dec 1996). https://doi.org/10.1006/ijhc.1996.0071, https://www.sciencedirect.com/science/article/pii/S1071581996900713
10. Djamasbi, S., Tulu, B., Loiacono, E., Shitefleet-Smith, J.: Can a Reasonable Time Limit Improve the Effective Usage of a Computerized Decision Aid? Communications of the Association for Information Systems **23**(22), 393–408 (2008), http://digitalcommons.wpi.edu/uxdmrl-pubs http://digitalcommons.wpi.edu/uxdmrl-pubs/27 http://aisel.aisnet.org/cais/vol23/iss1/22
11. Gamboa, H.F.S.: Multi-modal behavioral biometrics based on HCI and electrophysiology (April 2008), 1–216 (2008), http://www.lx.it.pt/ afred/pub/thesisHugoGamboa.pdf
12. Hibbeln, M., Jenkins, J.L., Schneider, C., Valacich, J.S., Weinmann, M.: How Is Your User Feeling? Inferring Emotion Through Human-Computer interaction Devices. MIS Quarterly **41**(1), 1–21 (2017). https://doi.org/10.25300/MISQ/2017/41.1.01
13. Hodgkinson, G.P., Bown, N.J., Maule, A.J., Glaister, K.W., Pearman, A.D.: Breaking the frame: An analysis of strategic cognition and decision making under uncertainty. Strategic Management Journal **20**(10), 977–985 (oct 1999). https://doi.org/10.1002/(SICI)1097-0266(199910)20:10¡977::AID-SMJ58¿3.0.CO;2-X
14. Holzinger, A.: Human-Computer Interaction and Knowledge Discovery (HCI-KDD): What is the benefit of bringing those two fields to work together? In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). vol. 8127 LNCS, pp. 319–328. Springer, Berlin, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40511-2_22, http://link.springer.com/10.1007/978-3-642-40511-2_22
15. Huang, J., White, R.W., Dumais, S.: No clicks, no problem: Using cursor movements to understand and improve search (2011). https://doi.org/10.1145/1978942.1979125
16. Katerina, T., Nicolaos, P.: Mouse behavioral patterns and keystroke dynamics in End-User Development: What can they tell us about users' behavioral attributes? Computers in Human Behavior **83**, 288–305 (jun 2018). https://doi.org/10.1016/j.chb.2018.02.012, https://www.sciencedirect.com/science/article/pii/S0747563218300700
17. Khan, I.A., Brinkman, W.P., Fine, N., Hierons, R.M.: Measuring personality from keyboard and mouse use. In: Proceedings of the 15th European conference on Cognitive ergonomics the ergonomics of cool interaction - ECCE '08. p. 1 (2008). https://doi.org/10.1145/1473018.1473066
18. Olson, G.M., Olson, J.S.: Human-Computer Interaction: Psychological Aspects of the Human Use of Computing. Annual Review of Psychology **54**(1), 491–516 (feb 2003). https://doi.org/10.1146/annurev.psych.54.101601.145044
19. Payne, J.W., Bettman, J.R., Johnson, E.J.: Adaptive strategy selection in decision making. Journal of Experimental Psychology: Learning, Memory, and Cognition **14**(3), 534–552 (1988). https://doi.org/10.1037/0278-7393.14.3.534, http://doi.apa.org/getdoi.cfm?doi=10.1037/0278-7393.14.3.534
20. Pimenta, A., Carneiro, D., Novais, P., Neves, J.: Monitoring mental fatigue through the analysis of keyboard and mouse interaction patterns. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial In-

telligence and Lecture Notes in Bioinformatics) **8073 LNAI**, 222–231 (2013). https://doi.org/10.1007/978-3-642-40846-5_23
21. Pocius, K.E.: Personality factors in human-computer interaction: A review of the literature. Computers in Human Behavior **7**(3), 103–135 (jan 1991). https://doi.org/10.1016/0747-5632(91)90002-I, https://www.sciencedirect.com/science/article/pii/074756329190002I
22. Rayner, K.: Eye movements in reading and information processing: 20 years of research. Psychological bulletin **124**(3), 372–422 (nov 1998), http://www.ncbi.nlm.nih.gov/pubmed/9849112
23. Revett, K., Jahankhani, H., de Magalhães, S.T., Santos, H.M.D.: A Survey of User Authentication Based on Mouse Dynamics. In: Global E-Security, pp. 210–219. Springer, Berlin, Heidelberg (2008). https://doi.org/10.1007/978-3-540-69403-8_25
24. Seelye, A., Hagler, S., Mattek, N., Howieson, D.B., Wild, K., Dodge, H.H., Kaye, J.A.: Computer mouse movement patterns: A potential marker of mild cognitive impairment. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring **1**(4), 472–480 (dec 2015). https://doi.org/10.1016/j.dadm.2015.09.006, https://www.sciencedirect.com/science/article/pii/S2352872915000792 http://linkinghub.elsevier.com/retrieve/pii/S2352872915000792
25. Yamauchi, T., Xiao, K.: Reading Emotion From Mouse Cursor Motions: Affective Computing Approach (nov 2017). https://doi.org/10.1111/cogs.12557
26. Zimmermann, P., Guttormsen, S., Danuser, B., Gomez, P.: Affective computinga rationale for measuring mood with mouse and keyboard. International Journal of Occupational Safety and Ergonomics **9**(4), 539–551 (jan 2003). https://doi.org/10.1080/10803548.2003.11076589

The last publication is entitled "Predicting Response Uncertainty in Online Surveys: A Proof of Concept"and it covers the main achievements of this dissertation. The article was accepted to the "12th International Joint Conference on Biomedical Engineering Systems and Technologies"(BIOSTEC 2019).

# Predicting Response Uncertainty in Online Surveys: A Proof of Concept

Maria Camila Dias[1], Catia Cepeda[1,2], Dina Rindlisbacher[2,3], Edouard Battegay[2,3], Marcus Cheetham[2,3*] and Hugo Gamboa[1*]

[1] *LIBPhys (Laboratory for Instrumentation, Biomedical Engineering and Radiation Physics), Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal*

[2] *Department of Internal Medicine, University Hospital Zurich, Zurich, Switzerland*

[3] *Center of Competence Multimorbidity, University of Zurich, Zurich, Switzerland*

*mc.dias@campus.fct.unl.pt, c.cepeda@campus.fct.unl.pt, Dina.Rindlisbacher@usz.ch, Edouard.Battegay@usz.ch, Marcus.Cheetham@usz.ch, hgamboa@fct.unl.pt*

[*] *These authors have made an equal contribution*

Keywords: Uncertainty, Human-Computer Interaction, Signal Processing, Machine Learning.

Abstract: Online questionnaire research is growing at a fast pace. Mouse tracking methods enable capture of overt online behaviour while respondents fill out questionnaires and can give insight into perceptual, cognitive and affective processes. The present work aimed to examine mouse cursor movements in terms of features that are thought to indicate events of uncertainty and to generate a machine learning model to detect these events. N=79 participants completed a survey while mouse data was tracked. A range of features were extracted from the data and selected for model training and testing. Using Logistic Regression and k-fold cross-validation, the model achieved an estimated performance of 81 %. The features showed that, during uncertain events, the number of horizontal direction inversions increase and the mouse cursor moves greater distances. Items that evoked uncertainty were associated with longer interaction times and a higher number of visits. Future work should be developed to validate these methods further.

## 1 INTRODUCTION

Self-report questionnaires are the main methods of personality assessment (Boyle and Helmes, 2009). The assessed personality constructs (e.g. extraversion) are often complex, with several facets (Mcgrath, 2005). These facets are typically captured using a number of questionnaire items. For example, extraversion can include the facets gregariousness, warmth, positive emotions, activity, assertiveness, and excitement seeking (Mccrae and Costa, 2015). For each item, a number of response alternatives are provided (often in a Likert response format) that the respondent may use to provide his or her answer (Paulhus and Vazire, 2007).

Typically, the item asks whether a particular statement about the respondent's psychological states or traits is true (e.g., I feel anxious and uneasy in emergency situations). The respondents may sometimes experience difficulty responding to questions that they had not previously thought about, retrieving all information that is directly relevant for answering the question, selecting the response alternatives that most closely matches the respondent's subjective views, feelings, and judgments about his or her own "truth", or dealing with the many items and the consistency of the answers to these (Schwarz and Hippler, 1991; Dunning et al., 2004).

These difficulties might influence a respondent's overt behaviour while completing a questionnaire, e.g. how long a person dwells over a question, how quickly a response is given, or whether a response is revisited and corrected. There are other multiple possible factors that might influence a respondent's overt behaviour while completing a questionnaire, ranging from difficulties related to the construction of the questionnaire itself to factors related to the respondent, such as self-uncertainty or indecisiveness (Rassin, 2006; Paulhus and Vazire, 2007).

The aim of this work was to explore respondent's overt behaviour while completing online questionnaires. Online survey research is growing at a great pace (Paulhus and Vazire, 2007). Mouse tracking, that is, the collection of cursor positions, is a relatively recent method that can provide information about respondent's overt behaviour and underlying perceptual, cognitive and affective processes (Hehman et al., 2015). The present work examined mouse cursor movements in terms of features that are thought to indicate events of uncertainty and to generate a machine learning model that detects these events.

### 1.1 Related Work

Mouse tracking has been used in a few psychological studies concerning indecisiveness, ambivalence and response difficulty.

Watson (2015) analysed the validity of measuring indecisiveness with mouse tracking, which was not supported. Several features were computed and correlated with indecisiveness and it was only found one significant correlation - the number of vertical direction changes. Surprisingly, it was not obtained an association between indecisiveness and response times.

Schneider et al. (2015) investigated the effect of ambivalence on mouse cursor trajectories by assessing response times and the maximum deviation from the idealized straight line trajectory toward the unchosen answer. It was concluded that, in case of uncertainty, the maximum deviation is higher. In agreement with Watson (2015), it was not found a relation between response times and uncertainty.

Zushi et al. (2012) developed a software that tracks mouse movements of students during their learning activities in order to help teachers understand their students' behaviours. It was verified that mouse trajectories become unstable (e.g. excessive number of horizontal direction inversions) when learners are hesitant. It was also perceived that response times and the number of horizontal direction inversions have a negative and strong correlation with the ratio of correct answers. That probably means that horizontal direction inversions and response times are good predictors of uncertainty (which contradict the findings of Watson (2015) and Schneider et al. (2015) related to response times). Accordingly, Conrad et al. (2007) used response times and age to detect response difficulty. Conversely, response times do not specify the cause of the delay. Slow responses can be related to multitasking, as answering a call, or the answer could involve mental arithmetic, among several causes (Horwitz et al., 2016).

Due to the disadvantages of using only response times as a predictor to response difficulty, Horwitz et al. (2016) used mouse cursor trajectories and age to predict it, achieving an accuracy of 74.28% (with temporal information, the accuracy rose to 79.11%). Hover the question text for more than 2s, mark a response option text for more than 2s and horizontal directional inversions turned out to be significant predictors of uncertainty. The main objective of this research work was similar to ours, however, the approach was different. Firstly, the extracted features were not the same. Moreover, Horwitz considered the features as binary (indicating whether each movement occured or not) or in a 3-point scale ("0" if the movement did not occur, "1" if the movement occured once and "2" if it occured more than once), which may provide less information than the actual number of occurrences or the total time of a certain pattern. Taking these relevant differences into account, the current study is innovative and, accordingly, the presented model might achieve a higher accuracy than the already existing model.

## 2 METHODS

### 2.1 Participants and Procedure

N = 79 volunteers (35 female) with ages ranging from 18 to 35 years old participated. The participants were recruited from the University of Zurich via flyers, and all of them were healthy, native or fluent speakers of standard German, with normal, or corrected-to-normal, vision, without a medical history of neurological or psychiatric illnesses and no current use of medication or drugs. They were paid 20 Swiss Francs, or the equivalent credit point. Written informed consent was obtained before participation in accordance with the guidelines of the Declaration of Helsinki.

### 2.2 Data Acquisition

Participants were seated in a quiet room while completing the online survey. The responses to an online questionnaire and related mouse movement data were collected. This data was used to construct the uncertainty model.

The mouse cursor data contained the frame number, event type (0 during movement, 1 when the mouse button is pressed down in the beginning of a click and 4 when the button is released in the end of a click), question number if hovered, answer number if hovered, x and y cursor's position (in pixels) and time.

### 2.3 Technological Materials

The Python packages used were NumPy (Bressert, 2012), SciPy (Blanco-Silva, 2013), Pandas (McKinney, 2011), Scikit-learn (Pedregosa et al., 2011) and Seaborn (Hernández et al., 2017).

## 2.4 Data Pre-processing

To ensure a correct processing of data from the mouse file, a cleaning procedure was applied as a first step. This omitted data acquired with touch screen devices, reordered the data by time and joined different files from the same questionnaire of the same person.

## 2.5 Features Extraction

Several features related to uncertainty behaviour were computed for each question of the survey. With these variables, it was created a model that detects the difficult questions for an individual. In this section, the temporal, spatial and contextual features are presented.

### 2.5.1 Temporal Features

Firstly, to access the temporal information, it was necessary to remove the time associated to abandon events. Sometimes, due to external factors (e.g. receiving an e-mail or answering a call), an individual may abandon the survey. Without correction, the questions where the abandons occur could be associated to uncertainty as a result of the time spent there. Therefore, the abandon events are identified - when the mouse cursor is not moving for more than 10 times the mean question time - and removed.

Short times in questions are also ignored. They can be caused by quick visits to the question above or below since the question height is small, or by scroll. These events occur when the time spent in a question is lower than 100 ms (Huang and White, 2012).

The temporal features are *accumulated time*, *time before click*, *pause before click*, *correction time*, *hover selected answer* and *velocity*.

The *accumulated time* is the total time in an item, i.e., the sum of all time intervals in a question, as expressed in equation 1, where $\Delta t_{qi}$ represents, hence, a time interval spent in question $i$.

$$Accumulated\ time = \sum \Delta t_{qi} \tag{1}$$

*Time before click* is the sum of all time intervals in a question until the first click, as shown in equation 2. For example, if a participant enters a question for the first time at t = 20s, stays in the item for 10s without clicking, abandon the question, comes back at t = 45s and clicks for the first time at t = 50s, the time before click is 15s.

$$Time\ before\ click = \sum_{enter}^{1^{st} click} \Delta t_{qi} \tag{2}$$

The *pause before click*, i.e., the time interval that an individual remains stopped before clicking an answer, was also computed, based on Zheng et al. (2011). If the participant clicks more than once in a single question (to correct a previous answer), this value is averaged.

*Correction time* is the sum of all time intervals in a question from the first click to the last click (last correction), as it is indicated in equation 3. If there is not any correction, the result is zero.

$$Correction\ time = \sum_{1^{st} click}^{last click} \Delta t_{qi} \tag{3}$$

*Hover selected answer* is the ratio between the sum of the time intervals spent hovering the selected answer of a certain question and the total hover time in that question. It was based on a feature extracted by Horwitz et al. (2016) and Cepeda et al. (2018). In this study, when an individual is in the response area, i.e., close to one of the possible answers, it is considered that he is hovering that answer. This feature is described in equation 4, where $\Delta t_{hover\ sel\ ans,\ qi}$ represents a time interval spent hovering the selected answer of question $i$ and $\Delta t_{hover,qi}$ is a time interval spent hovering the answers of question $i$.

$$Hover\ selected\ answer = \frac{\sum \Delta t_{hover\ sel\ ans,\ qi}}{\sum \Delta t_{hover,qi}} \tag{4}$$

The mean *velocity* was also calculated. To compute this variable, in order to have equal temporal intervals proportional to the mean time variance, it was applied a cubic spline interpolation. Using this method, a series of unique cubic polynomials are adjusted between the data points, resulting in a smooth continuous curve (Hou and Andrews, 1978).

### 2.5.2 Spatial Features

Firstly, it was applied a cubic spline interpolation to smooth the spatial signal, producing intervals equal to the mean distance variance. Subsequently, the spatial features *distance*, *distance from answer* and *straightness* were computed.

The total *distance* is the sum of the distances travelled in every visit to a specific question.

The *distance from answer*, i.e., distance from the path inside a question to the selected answer, was also computed. This variable is illustrated in equation 5, where $x_{ans}$ and $y_{ans}$ are the $x$ and $y$ coordinates of the question's last click and $n$ is the number of samples. For the construction of the model, it was calculated the mean *distance from answer*.

Figure 1: An example of a revisit. The participant clicked on an item (red dot) and, subsequently, returned to a previous question without changing its answer.

$$Distance\ from\ answer = \sqrt{(x_i - x_{ans})^2 + (y_i - y_{ans})^2}, \quad i = 1, \ldots, n-1 \tag{5}$$

*Straightness* is the ratio between the Euclidean distance from the moment of entering in a question until leaving it and the total distance travelled in that question (Gamboa and Fred, 2004). It is defined in equation 6. The mean *straightness* over all the visits to a specific question was used.

$$Straightness = \frac{\sqrt{(x_1 - x_n)^2 + (y_1 - y_n)^2}}{\sum_{i=1}^{n-1} \sqrt{\Delta x_i{}^2 + \Delta y_i{}^2}} \tag{6}$$

#### 2.5.3 Contextual Features

The contextual features comprise the number of *interactions* with each question (i.e., the number of times in each question) as well as the number of *revisits*, which is the event of going back to a previous question without changing its answer. An instance of a revisit is illustrated in figure 1.

The number of corrections was also calculated. There are two types of corrections - *corrections within item* and *corrections between item*. The first occurs when an individual selects an answer, remains in the same question and changes the option, while the latter happens when a person selects an answer, moves forward to next questions and, after answering at least one more question, goes back and changes the previous answer. These corrections are displayed in figure 2.

The number of <-*turns*, i.e., horizontal direction changes (Zushi et al., 2012; Horwitz et al., 2016; Cepeda et al., 2018), was extracted by computing horizontal trajectory derivative changes from positive to negative values or vice-versa. This feature is exemplified in figure 3.

Lastly, the relative number of *hovered answers* was computed and it is illustrated in equation 7.

$$Hovered\ answers = \frac{Number\ of\ hovered\ answers}{Total\ number\ of\ answers} \tag{7}$$

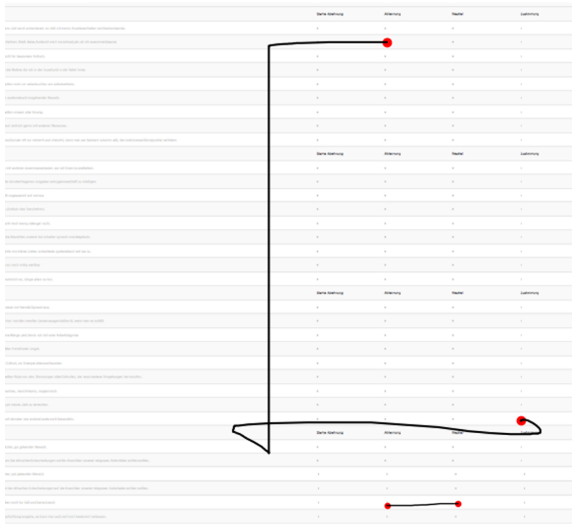


Figure 2: Two corrections, a correction between (above) and a correction within (below) item.

Figure 3: An example of a <-turn.

#### 2.5.4 Features Normalization

Distinct people express uncertainty differently. For example, maybe the time spent in a difficult question by a fast person is equal to the time spent in an easy question by a slower individual. Accordingly, the features were normalized for each person separately using the formula presented in equation 8, where $z_i$ represents the sample $x_i$ after normalization, $\bar{x}$ and $\sigma$ are the mean and standard deviation of the samples, respectively. This normalization is known as z-score (Shalabi et al., 2006). Applying this transformation, the samples are reshaped so that its mean and standard deviation become 0 and 1, respectively (Tan et al., 2003).

$$z_i = \frac{x_i - \bar{x}}{\sigma} \tag{8}$$

Nonetheless, with all the features normalized, it is only possible to identify the most difficult questions for each individual. In the hypothetical case of uncertainty in all questions (or a great part of them), this would be a problem. Therefore, the original values of each feature were also used to construct the model. Taking this into account, 30 features were used - 15 normalized and 15 not normalized.

Subsequently, all the features from all the participants were concatenated and normalized in order to standardize the range of the variables.

## 2.6 Features Selection

There is a negative effect of using irrelevant features in machine learning systems. Some classifiers are not sensible enough to detect the influence of relevant features in the presence of many variables (Sperandei, 2014). Taking this into account, it is advantageous to precede learning with a feature selection stage (Witten and Frank, 2005).

Accordingly, the highly correlated features were eliminated (Witten and Frank, 2005), since the information they provide is almost the same. The Pearson correlation coefficient was accessed and, if two features had an absolute coefficient higher than 0.9, one of them was left out.

## 2.7 Model Training and Testing

In order to train the uncertainty model and, afterwards, test it, several examples of uncertainty and certainty while answering survey questions were needed. Since the mouse cursor data has both spatial (vertical and horizontal directions) and temporal information, numerous mouse movements videos were observed and those needed examples were selected. Clearly, detecting uncertainty involves a subjective evaluation, which could be a barrier to construct an accurate model. To escape this problem, 3 individuals made the analysis separately and 1 of them had few information about the study. A question was selected (for certainty or uncertainty) only if at least 2 people had chosen it.

Mouse movements videos from 6 individuals answering a 60 item questionnaire were visualized. Accordingly, 360 questions were observed, but only 175 were chosen due to the difficulty and subjectivity of the task. 51 items were associated to uncertainty and 124 to certainty.

Figure 4 shows one of the items selected as an instance of uncertainty. The participant enters the question and immediately selects option 3. Afterwards, the individual moves the mouse cursor towards option 4, but reverses this trajectory until reaching option 1. Subsequently, the direction is inverted and the final answer is option 2. Note that this occurrence comprises a long distance travelled, a low straightness, one correction within item, two <-turns and three out of five hovered answers. It is important to refer that we can not access the temporal information through the image, which might also be extremely relevant.

On the contrary, figure 5 exhibits an example of certainty, where the mouse moves straightly from the answer of a question to the next one.

With the instances of certainty and uncertainty, the

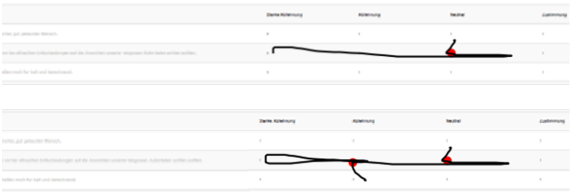


Figure 4: Question associated to uncertainty.

Figure 5: Question associated to certainty.

model could be trained and tested. It is recurrent to use, for example, two-thirds of the data for training and the remaining one-third to test the model. Conversely, the training or testing sets might not be representative. This problem may be soluted by repeating the process of training and testing various times with different samples. Taking this into consideration, it was used the *10-fold cross validation*. In this procedure, the data is divided in ten approximately equal partitions, where one of them is used for testing, while the remaining nine are used for training, and the process is repeated ten times. In each iteration, the datasets change and, accordingly, every instance is used for both training and testing, and exactly once for testing. Finally, the ten estimated accuracies are averaged to obtain the overall accuracy. The number of folds might have been different, but a considerable amount of tests has led to the conclusion that ten is the number that reaches the best estimate of error. Even though these results are questionable, the 10-fold cross-validation has become the conventional practice (Witten and Frank, 2005).

## 2.8 Classification

The applied classification method was *Logistic Regression*, due to its effectiveness when the outcome variable is dichotomous (in this case, the outcome might be certainty or uncertainty). In this technique, the probability of occurrence of an event is estimated by fitting the data to a logistic curve. Accordingly, non-linear relationships between the input features and the outcome variable can be handled (Park, 2013).

The fundamental mathematical concept underlying *Logistic Regression* is the logit. The logit is the natural logarithm of odds ratio, which is the ratio between the probability of occurrence of an event (in this case, uncertainty) and the probability of non-occurence of the same event. The logistic model has the form presented in equations 9 and 10, where $p$ represents the probability of an event, $\beta_i$ illustrates the regression coefficients and $x_i$ are the input features (Sperandei, 2014).

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n \quad (9)$$

Solving for $p$,

$$p = \frac{1}{1+e^{-(\beta_0+...+\beta_n x_n)}} \quad (10)$$

When $p \geqslant 0.5$ it is predicted $Y = 1$ (uncertainty), otherwise, $Y = 0$, where $Y$ is the outcome variable (Shalizi, 2018). From equation 10, it is possible to verify that a positive $\beta_i$ increases (and a negative $\beta_i$ decreases) the probability of $Y = 1$.

## 2.9 Model Evaluation

In binary classification, data is constituted by two opposite classes, positives and negatives. Accordingly, the possible outcomes comprise True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). In this study, the positives are the questions linked to uncertainty.

The true positive rate, or *sensitivity*, and the true negative rate, or *specificity*, were computed (Witten and Frank, 2005). In this case, the *sensitivity* represents the probability of a question that evokes uncertainty being classified as an instance of uncertainty, and it is described in equation 11. *Specificity*, on the other hand, provides the probability of a question associated to certainty being correctly classified and it is illustrated by equation 12.

$$Sensitivity = \frac{TP}{TP+FN} \quad (11)$$

$$Specificity = \frac{TN}{TN+FP} \quad (12)$$

Since the data is imbalanced (there are more certainty events than uncertainty occurrences), the most appropriate measure to evaluate the model performance is *f1 score*, defined in equation 13 as the harmonic mean between *precision* and *recall*. *Recall* is a synonym of *sensitivity*, as it is possible to verify in equation 14. *Precision*, on its turn, represents the probability of a certainty event being classified as an uncertainty event, as shown in equation 15 (Sun et al., 2007).

$$f1\ score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (13)$$

Where

$$Recall = Sensitivity \quad (14)$$

And

$$Precision = \frac{FP}{FP+TN} \quad (15)$$

# 3 RESULTS

## 3.1 Features Selection and Importance

The highly correlated features were removed, as it was explained in section 2.6. The features eliminated with this criterion were *time before click*, *hover selected answer*, *straightness normalized*, *revisits*, *revisits normalized* and *hovered answers normalized*. Therefore, the number of final features was 24.

Some features have more importance than others in the classification process. From equation 10, it is possible to infer that features with higher regression coefficients are more relevant to the classification. Table 1 shows the regression coefficients of the selected features ordered from the highest to the lowest absolute value.

## 3.2 Model Evaluation

The model evaluation measures - *sensitivity*, *specificity* and *f1 score* - are presented in table 2.

## 3.3 Uncertainty Results

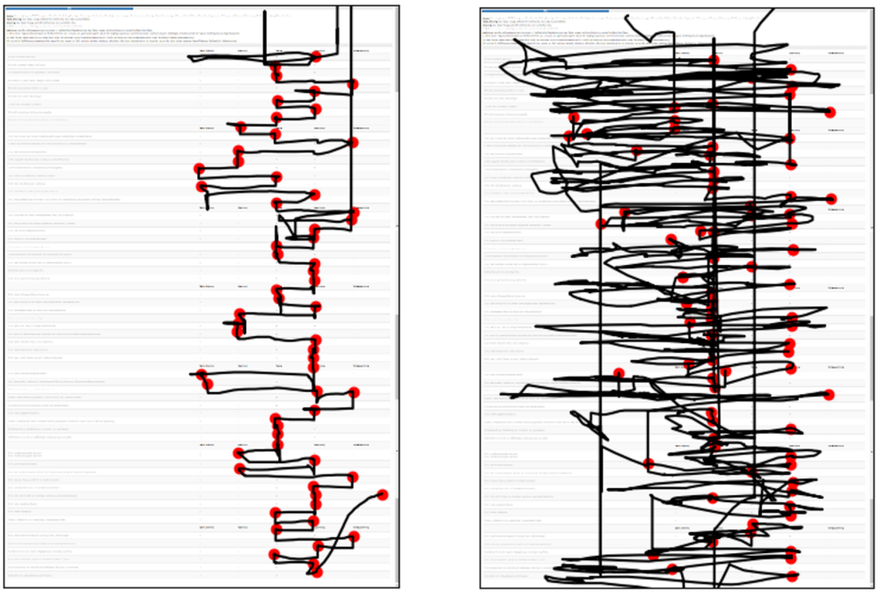
Following the application of the model to all participants' questions, the percentage of questions associated to uncertainty was computed. The values ranged from 6.36% to 81.08% (M = 28.08%; SD = 15.15%).

Figure 6 shows the contrast of the mouse movements between the individuals with the minimum and maximum percentages of questions that evoked uncertainty.

# 4 DISCUSSION

Regarding the construction of a model, it is important to assess its relevant features, performance evaluation and outcomes.

Firstly, from table 1, it can be analysed which were the most important features for the construction of the model. It is possible to verify that the number of *<-turns* is the most relevant feature and, with a positive regression coefficient, it increases the probability of detecting an uncertainty event. On that account, when facing uncertainty while interacting with a computer, individuals tend to change the horizontal



(a) (b)

Figure 6: Mouse movements of a questionnaire from the person with a) the minimum and b) the maximum percentage of uncertainty items.

direction more frequently, probably due to hesitation between consecutive alternatives, in line with Zushi et al. (2012).

Furthermore, the *distance* travelled produce a great and positive impact on the outcome, suggesting that people move the mouse from a possible answer to another while deciding which one to select. *Distance from answer*, on its turn, affects the result negatively, which means that, although individuals travel longer distances during moments of uncertainty, they tend to maintain the mouse cursor closer to the selected alternative. Probably this is influenced by consecutive questions with opposite (or very different) responses. That is, when a person moves directly from option 1 of an item to option 5 of the subsequent question, one of these items is associated to a large mean *distance from answer*. Nonetheless, in a question associated to uncertainty, where the travelled distance is long, this effect is attenuated.

Analysing the regression coefficient of *interactions*, it can be concluded that people visit more often items that arouse uncertainty. In these questions, individuals take longer times to answer (*accumulated time* has a positive and significant regression coefficient) and deviate more from the straight line trajectory between successive answers (*straightness* is associated to a negative coefficient).

It is surprising that the number of corrections influence negatively the result. This means that, when the number of corrections increases, the probability of identifying an uncertainty event decreases. Perhaps the great part of corrections result from distractions, which might be more recurrent in the absence of uncertainty, since the latter evokes more reflection.

Regarding the model evaluation, the *sensitivity* obtained was 0.78, which means that the instances of uncertainty are correctly classified in 78% of the times. The *specificity*, on its turn, achieved a value of 0.94 and hence the probability of a certainty event being correctly predicted is 94%. Using *f1 score*, the estimated performance of the model was 0.81. Taking into account that uncertainty assessment concerns a subjective evaluation, the performance of the model is very good.

Following the construction of the model, the percentage of instances that evoked uncertainty was accessed. As already mentioned, figure 6 illustrates the mouse movements from the person with minimum percentage and from the participant with the maximum percentage, and the behaviours are clearly different, where the distance travelled is much higher in the latter.

Table 1: Regression coefficients of the selected features.

| Feature | Regression coefficient |
|---|---|
| **<-Turns** | 1.47 |
| **Distance normalized** (px) | 1.23 |
| **Distance** (px) | 1.19 |
| **Distance from answer normalized** (px) | -0.93 |
| **Interactions** | 0.65 |
| **Accumulated time** (s) | 0.61 |
| **Straightness** | -0.49 |
| **Pause before click** (s) | 0.31 |
| **Corrections between item** | -0.31 |
| **Distance from answer** (px) | -0.29 |
| **Hovered answers** | 0.28 |
| **<-Turns normalized** | 0.22 |
| **Pause before click normalized** (s) | -0.18 |
| **Corrections within item normalized** | -0.16 |
| **Correction time normalized** (s) | 0.15 |
| **Hover selected answer** | -0.15 |
| **Velocity** (px/s) | 0.13 |
| **Velocity normalized** (px/s) | -0.13 |
| **Corrections within item** | -0.06 |
| **Corrections between item normalized** | 0.05 |
| **Accumulated time normalized** (s) | 0.03 |
| **Correction time** (s) | 0.01 |

## 5 CONCLUSIONS

The objective of this study was to construct a model that detects events of uncertainty using mouse cursor movements acquired during survey responses.

To build the model, several features were extracted from mouse tracking data. Some of them were more relevant to the classification process, and, assessing their importance, it was concluded that, in case of uncertainty while interacting with a computer, individuals increase the number of horizontal direction inversions with the cursor, the distance travelled by the mouse is higher but the mean distance from the selected answer is lower and the number of visits to the item that aroused uncertainty increases, as well as the time spent there.

Table 2: Model performance evaluation measures.

| Sensitivity | Specificity | f1 score |
|---|---|---|
| 0.78 | 0.94 | 0.81 |

The estimated performance of the created model was 0.81.

Despite the good results, some improvements should be made in the future. First of all, the uncertain events were defined by visualization and subjective evaluation of independent raters. In a next step, the actual participants should provide feedback as to their own experience of uncertain events.

Based on this, future work should validate this method further in other contexts, such as during a career decision making task. Accordingly, the developed model may be used to help to identify confusing items in a questionnaire or to provide real-time help for difficult questions.

## REFERENCES


Blanco-Silva, F. J. (2013). *Learning SciPy for Numerical and Scientific Computing*. Birmingham.

Boyle, G. J. and Helmes, E. (2009). Methods of personality assessment. *Humanities & Social Sciences papers*, (327).

Bressert, E. (2012). *SciPy and NumPy*. O'ReillyMedia, first edition.

Cepeda, C., Rodrigues, J., Dias, M. C., Oliveira, D., Rindlisbacher, D., Cheetham, M., and Gamboa, H. (2018). Mouse tracking measures and movement patterns with application for online surveys. *Cross Domain Conference for Machine Learning and Knowledge Extraction (CD-MAKE 2018)*, 11015:28–42.

Conrad, F. G., Schober, M. F., and Coiner, T. (2007). Bringing Features of Human Dialogue to Web Surveys. *Applied Cognitive Psychology*, 21:165–187.

Dunning, D., Heath, C., and Suls, J. M. (2004). Flawed Self-Assessment: Implications for Health, Education, and the Workplace. *Psychological Science in the Public Interest*, 5(3):69–106.

Gamboa, H. and Fred, A. (2004). A behavioral biometric system based on human computer interaction. *Proceedings of SPIE - The International Society for Optical Engineering*, 5404:381–392.

Hehman, E., Stolier, R. M., and Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes and Intergroup Relations*, 18(3):384–401.

Hernández, C. X., Harrigan, M. P., Sultan, M. M., and Pande, V. S. (2017). MSMExplorer: Data Visualizations for Biomolecular Dynamics. *The Journal of Open Source Software*.

Horwitz, R., Kreuter, F., and Conrad, F. (2016). Using Mouse Movements to Predict Web Survey Response Difficulty. *Social Science Computer Review*, 35(3).

Hou, H. S. and Andrews, H. C. (1978). Cubic Splines for Image Interpolation and Digital Filtering. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(6):508–517.

Huang, J. and White, R. (2012). User See, User Point: Gaze and Cursor Alignment in Web Search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2012)*, pages 1341–1350.

Mccrae, R. R. and Costa, P. T. (2015). Rotation to Maximize the Construct Validity of Factors in the NEO Personality Inventory. *Multivariate Behavioral Research*, 24(1):107–124.

Mcgrath, B. (2005). Hepatoprotective and antioxidant activity of methanolic extract of vetiveria zizanioides roots against paracetamol-induced liver damage in rats. *Journal of Personality Assessment*, 85(2):112–124.

McKinney, W. (2011). pandas: a Foundational Python Library for Data Analysis and Statistics. *Python for High Performance and Scientific Computing*, pages 1–9.

Park, H. A. (2013). An introduction to logistic regression: From basic concepts to interpretation with particular attention to nursing domain. *Journal of Korean Academy of Nursing*, 43(2):154–164.

Paulhus, D. L. and Vazire, S. (2007). The Self-Report Method. In Robins, R. W., Fraley, R. C., and Krueger, R., editors, *Handbook of research methods in personality psychology*, chapter 13, pages 224–239. Guilford, New York.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Rassin, E. (2006). A psychological theory of indecisiveness. *Netherlands Journal of Psychology*, 63(1):2–13.

Schneider, I. K., van Harreveld, F., Rotteveel, M., Topolinski, S., van der Pligt, J., Schwarz, N., and Koole, S. L. (2015). The path of ambivalence: tracing the pull of opposing evaluations using mouse trajectories. *Frontiers in Psychology*, 6(996):1–12.

Schwarz, N. and Hippler, H.-J. (1991). Response Alternatives: The Impact of Their Choice and Presentation Order. In Biemer, P., Groves, R., Lyberg, L., Mathiowetz, N., and Sudman, S., editors, *Measurement error in surveys*, chapter 3, pages 41–56. Wiley, Chichester.

Shalabi, L. A., Shaaban, Z., and Kasasbeh, B. (2006). Data Mining: A Preprocessing Engine. *Journal of Computer Science*, 2(9):735–739.

Shalizi, C. R. (2018). *Advanced Data Analysis from an Elementary Point of View*.

Sperandei, S. (2014). Understanding logistic regression analysis. *Biochemia Medica*, 24(1):12–18.

Sun, Y., Kamel, M. S., Wong, A. K., and Wang, Y. (2007). Cost-sensitive boosting for classification of imbalanced data. *Pattern Recognition*, 40:3358–3378.

Tan, P. K., Downey, T. J., Spitznagel, E. L., Xu, P., Fu, D., Dimitrov, D. S., Lempicki, R. A., Raaka, B. M., and Cam, M. C. (2003). Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Research*, 31(19):5676–5684.

Watson, J. A. (2015). *Indecisiveness: Measuring Individual Differences With Mouse Tracking Software*. PhD thesis, Ball State University.

Witten, I. H. and Frank, E. (2005). *Data Mining - Practical Machine Learning Tools and Techniques*. Elsevier, San Francisco, second edition.

Zheng, N., Paloski, A., and Wang, H. (2011). An efficient user verification system via mouse movements. *Proceedings of the 18th ACM conference on Computer and communications security - CCS '11*, pages 139–150.

Zushi, M., Miyazaki, Y., and Norizuki, K. (2012). Web application for recording learners' mouse trajectories and retrieving their study logs for data analysis. *Knowledge Management and E-Learning: An International Journal*, 4(1):37–50.