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*

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

Abstract: Online questionnaire-based research is growing at a fast pace. Mouse-tracking methods provide a potentially important data source for this research by enabling the capture of respondents' online behaviour while answering questionnaire items. This behaviour can give insight into respondents' perceptual, cognitive and affective processes. The present work focused on the potential use of mouse movements to indicate uncertainty when answering questionnaire items and used machine learning methods as a basis to model these. N=79 participants completed an online questionnaire while mouse data was tracked. Mouse movement features were extracted and selected for model training and testing. Using logistic regression and k-fold cross-validation, the model achieved an estimated performance accuracy of 89%. The findings show that uncertainty is indicated by an increase in the number of horizontal direction inversions and the distance covered by the mouse and by longer interaction times with and a higher number of revisits to questionnaire items that evoked uncertainty. Future work should validate these methods further.

1 INTRODUCTION

Self-report questionnaires are the main method of personality assessment (Boyle and Helmes, 2009). The assessed personality constructs (e.g. extraversion) may have several facets (Mcgrath, 2005). For example, extraversion can include the facets gregariousness, warmth, positive emotions, activity, assertiveness and excitement seeking (Mccrae and Costa, 2015). Each facet is typically captured by measuring a person's responses to a number of questionnaire items. Typically, an item asks whether a particular statement about the respondent is true (e.g., I feel anxious and uneasy in emergencies). For each item, a rating scale with a number of response alternatives is provided (often in a Likert response format) (Paulhus and Vazire, 2007) that allows the respondent to confirm the degree to which the statement is true or false.

Sometimes a respondent may find it difficult to make a rating about a statement. The respondent may not have thought about the statement previously, have difficulty retrieving from memory all relevant information, feel unsure about which response alternative best matches the respondent's subjective point of view, or find it difficult dealing with many similar statements in a questionnaire (Schwarz and Hippler, 1991; Dunning et al., 2004). The respondent may also tend to be self-uncertain or indecisive (Rassin, 2006; Paulhus and Vazire, 2007).

We considered whether uncertainty in processing and responding to questionnaire items might be detectable in concomitant mouse movement behaviour. Mouse tracking, i.e., 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). Indicators of response uncertainty might include how long a person hovers with the mouse over a question, how quickly a response is given, or whether a person revisits an items or corrects the previous response to it.

The aim of the present work was to create a machine-learning model that identifies events of re-

*These authors have made an equal contribution.

Dias, M., Cepeda, C., Rindlisbacher, D., Battegay, E., Cheetham, M. and Gamboa, H.
Predicting Response Uncertainty in Online Surveys: A Proof of Concept.
DOI: 10.5220/0007381801550162
In *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2019)*, pages 155-162
ISBN: 978-989-758-353-7

sponse uncertainty. To do so, features of mouse movement behaviour were extracted while respondents processed and answered questionnaire items.

The developed model may be used to identify confusing items in a survey. Moreover, it may help physicians understand their main difficulties, since the practice of medicine is characterized by complex situations that arouse uncertainty, which has implications for the quality and costs of healthcare.

1.1 Related Work

Human-Computer Interaction (HCI) often reflects the users' mental processes. Accordingly, understanding human behaviour through the analysis of HCI has been an area of interest for many years. Mouse tracking is a powerful, cheap and easy to implement tool to assess HCI (Cepeda et al., 2018). Mouse tracking has been used for online survey research (Cepeda et al., 2018; Horwitz et al., 2016) and for examining response difficulty (Schneider et al., 2015; Zushi et al., 2012; Horwitz et al., 2016).

Schneider et al. (2015) investigated the effect of ambivalence (a concept similar in some ways to uncertainty) on mouse cursor trajectories by assessing response times and the maximum deviation from the idealized straight line trajectory toward the answer that was not chosen. The authors concluded that deviation was greater with greater ambivalence.

Zushi et al. (2012) tracked mouse movements during students' learning activities in order to help teachers understand their students' behaviours. It was shown that mouse trajectories are unstable (e.g. excessive number of horizontal direction inversions) when learners are hesitant. It was also reported that response times and number of horizontal direction inversions have a strong negative correlation with the ratio of correct answers. This suggests that response times and horizontal direction inversions may be predictive of uncertainty.

Conrad et al. (2007) also used response times to detect response difficulty. But response times do not specify the reason for the delay. Slower responses can have several causes, such as when multitasking and distracted by another task (Horwitz et al., 2016). Horwitz et al. (2016) used therefore mouse cursor trajectories to predict response difficulty, achieving a performance accuracy of between 74.28% and 79.11%. Significant predictors of uncertainty were horizontal directional inversions, hovering the mouse cursor over a question for more than 2s, and marking a response option for more than 2s.

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. All 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 use of medication or drugs. They received 20 Swiss Francs or the equivalent credit point for participation. 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 an online questionnaire. The questionnaire responses and all mouse movement data were collected. The mouse data included the frame number, x and y cursor's position (in pixels), time, event type (0 during movement, 1 when pressing down the mouse button and 4 when the button is released), the question number if the mouse hovered over it, and the number of the response alternative if the mouse hovered over it.

2.3 Technological Materials

The LimeSurvey web application was used to conduct the questionnaire. The Python packages NumPy (Bressert, 2012) and Pandas (McKinney, 2011) were used for all data processing and analyses, SciPy (Blanco-Silva, 2013) was used to extract features from the mouse tracking data and Scikit-learn (Pedregosa et al., 2011) for model training, testing and classification.

2.4 Data Pre-processing

To ensure a correct processing of data from the mouse file, a cleaning procedure was applied to omit data acquired with touch screen devices, to reorder the data by time, and to join different files from the same questionnaire of the same person.

2.5 Features Extraction

Several features related to uncertainty behaviour were computed for each item of the questionnaire. Based on these features, a machine learning model was generated to detect items that evoked uncertainty. 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 Δt_{qi} represents, hence, a time interval spent in question i.

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

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.

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

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 certain 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.

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

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.

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

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 is 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*.

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

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.

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

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.

Where

$$\Delta x_i = x_{i+1} - x_i \qquad \Delta y_i = y_{i+1} - y_i \qquad i = 1, \ldots, n-1 \tag{7}$$

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.

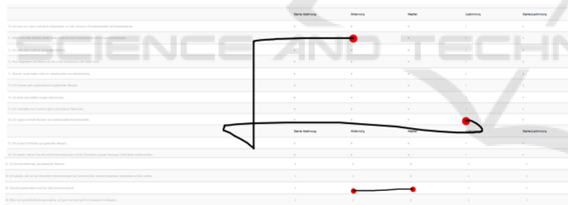


Figure 2: Two corrections, a correction between (above) and a correction within (below) item.

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.

Figure 3: An example of a <-turn.

Lastly, the relative number of *hovered answers* was computed and it is illustrated in equation 8.

$$\text{Hovered answers} = \frac{\text{Number of hovered answers}}{\text{Total number of answers}} \tag{8}$$

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 9, where z_i represents the sample x_i after normalization, $\bar{x}$ and σ 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).

$$z_i = \frac{x_i - \bar{x}}{\sigma} \tag{9}$$

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 each feature was individually normalized in order to standardize the range of the variables for all the participants.

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 and test the uncertainty model, several examples of items showing response uncertainty and certainty were needed. These examples comprise a combination of features and a respective outcome (certainty or uncertainty). However, it was not known which items evoked, or not, uncertainty. To solve

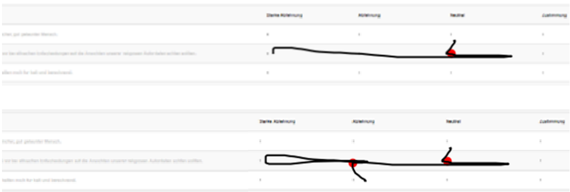


Figure 4: Question associated to uncertainty.

Figure 5: Question associated to certainty.

this problem, mouse movement videos of 6 individuals answering a 60 item questionnaire (360 questions in total) were observed and rated by three raters in terms of uncertainty or certainty. The final examples of items for training and testing were selected only if rated as uncertainty or certainty by at least 2 of the raters. In the end, 51 items were rated as uncertainty and 124 as 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. On the contrary, figure 5 shows an example of certainty, where the mouse moves straight from the answer of a question to the next one.

10-fold cross validation was applied for model training and testing. In this procedure, the data is divided in ten approximately equal partitions, where one partition is used for testing and the other nine for training. This process is repeated ten times. In each iteration, the datasets change and, accordingly, every partition is used for both training and testing, and exactly once for testing. Finally, the ten estimated accuracies are averaged to obtain the overall accuracy.

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 could 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-occurrence of the same event. The logistic model has the form presented in equations 10 and 11, where p represents the probability of an event, β_i illustrates the regression coefficients and x_i are the input features (Sperandei, 2014).

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \qquad (10)$$

Solving for p,

$$p = \frac{1}{1+e^{-(\beta_0+\ldots+\beta_n x_n)}} \qquad (11)$$

When $p \geqslant 0.5$ it is predicted $Y = 1$ (uncertainty), otherwise, $Y = 0$, where Y is the outcome variable (Shalizi, 2018). From equation 11, it is possible to verify that a positive β_i increases (and a negative β_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 12. *Specificity*, on the other hand, provides the probability of a question associated to certainty being correctly classified and it is illustrated by equation 13.

$$Sensitivity = \frac{TP}{TP+FN} \qquad (12)$$

$$Specificity = \frac{TN}{TN+FP} \qquad (13)$$

To estimate the performance of the model, accuracy was accessed. Accuracy is the ratio between the correct classifications and all the classifications (Witten and Frank, 2005), as it is shown in equation 14.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (14)$$

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 11, 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 ten most relevant features ordered from the highest to the lowest absolute value.

Table 1: Regression coefficients of the ten most relevant 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

3.2 Model Evaluation

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

Table 2: Model performance evaluation measures.

Sensitivity	Specificity	Accuracy
0.78±0.17	0.94±0.08	0.89±0.08

3.3 Uncertainty Results

Following the application of the model to all participants' questions, the percentage of questions associated to uncertainty was computed. 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

Table 1 shows the most important features for the construction of the model. The number of *<-turns* is the most relevant feature and, with a positive regression coefficient, it increases the probability of detecting an uncertainty event. Individuals thus tend to change the horizontal direction more frequently during a moment of uncertainty, probably due to hesitation between consecutive alternatives. This is in line with Zushi et al. (2012).

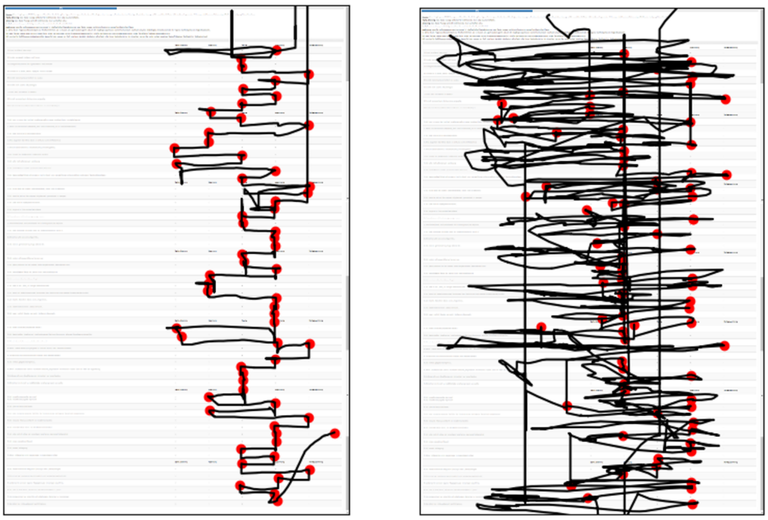
The *distance* travelled has a strong positive impact on the outcome, suggesting that respondents move the mouse from a possible answer to another while deciding which one to select. *Distance from answer* affected the result negatively, meaning that, although individuals travel longer distances during moments of uncertainty, they tend to maintain the mouse cursor closer to the selected alternative.

Analysing the regression coefficient of *interactions*, it can be concluded that people visit items that arouse uncertainty more often. In these items, individuals take longer 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.

Regarding the model evaluation, the *sensitivity* obtained was 0.78, which means that the instances of uncertainty were correctly classified in 78% of the times. The *specificity* was 0.94 (i.e. the probability of a certainty event being correctly predicted is 94%). The classification of certainty versus uncertainty was correct in 89% of the cases. The estimated performance of the model was therefore better than that of Horwitz et al. (2016). This improvement might relate to the choice of features used to indicate uncertainty.

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.



(a) (b)

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

5 CONCLUSIONS

This study aimed to assess respondent's mouse cursor movements in terms of uncertainty while processing and answering items of an online questionnaire. The estimated performance accuracy of the created model for uncertainty detection was 89%.

As a proof of concept, the uncertain events were defined by subjective evaluation of independent raters. As a next step, the actual participants should provide feedback as to their own experience of uncertain events. Future work could validate this method further in other contexts, such as during an online career assessment task or to identify moments of uncertainty as a basis for providing real-time online help for difficult items or 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.

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.

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.

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.