Research article

# Knowledge extraction from pointer movements and its application to detect uncertainty ☆

Catia Cepeda [a,b,*], Maria Camila Dias [b], Dina Rindlisbacher [a], Hugo Gamboa [b,1], Marcus Cheetham [a,1]

[a] *Department of Internal Medicine, University Hospital Zurich, Zurich, Switzerland*
[b] *LIBPhys (Laboratory for Instrumentation, Biomedical Engineering and Radiation Physics), Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal*

## ARTICLE INFO

## ABSTRACT

Pointer-tracking methods can capture a real-time trace at high spatio-temporal resolution of users' pointer interactions with a graphical user interface. This trace is potentially valuable for research on human-computer interaction (HCI) and for investigating perceptual, cognitive and affective processes during HCI. However, little research has reported spatio-temporal pointer features for the purpose of tracking pointer movements in on-line surveys. In two studies, we identified a set of pointer features and movement patterns and showed that these can be easily distinguished. In a third study, we explored the feasibility of using patterns of interactive pointer movements, or micro-behaviours, to detect response uncertainty. Using logistic regression and k-fold cross-validation in model training and testing, the uncertainty model achieved an estimated performance accuracy of 81%. These findings suggest that micro-behaviours provide a promising approach toward developing a better understanding of the relationship between the dynamics of pointer movements and underlying perceptual, cognitive and affective psychological mechanisms.

## 1. Introduction

Human-computer interaction (HCI) is a multidisciplinary field of study on the design, implementation and evaluation of interactive systems (Dix, 2004). Understanding users' patterns of behaviour while interacting with a graphical user interface is important, for example, to assess usability and user experience (e.g. Dillon and Watson, 1996; Pocius, 1991). While eye tracking technology has been the main approach to tracking these patterns (Rayner, 1998), it also requires a special monitoring device and the physical presence of the user to enable acquisition of eye movements.

Pointer (or mouse) tracking provides a low-cost, highly scalable alternative approach to acquiring data on patterns of user behaviour while interacting with a graphical user interface (Chen et al., 2001; Rodden and Fu, 2007). Pointer tracking usually entails the use of proprietary software to collect a trace of the pointer (or cursor) positions as guided by the user's mouse movements (Dix, 2004). Pointer tracking data is typically used to test the usability of web pages and improve user experience (Atterer et al., 2006; Arroyo et al., 2006; Huang et al., 2011; Digital Experience Analytics, 2018; Inspectlet, 2018).

A considerable number of temporal and spatial measures, or features, can be extracted from pointer data for analysing the dynamics of pointer movements. The most common temporal features are velocity and acceleration. Spatial features typically include distance travelled, angle of direction, curvature and straightness (Gamboa and Fred, 2004; Chudá and Krátky, 2014; Ahmed and Traore, 2007; Pusara and Brodley, 2004; Arroyo et al., 2006). More complex features are possible, including hovering patterns (Tzafilkou and Protogeros, 2018; Arapakis and Leiva, 2016; Huang et al., 2011), long pauses (Tzafilkou and Protogeros, 2018; Arroyo et al., 2006; Seelye et al., 2015) and directional changes (Yamauchi and Xiao, 2018) that require more intricate analysis.

There has been much research on user interaction based on pointer data, especially using features of low-granularity such as mouse clicks and the number of pointer movements (e.g. Goecks and Shavlik (2000)). These have been used as indicators of user interest, hesitation, user engagement, and abandonment, mainly for testing web page usability (e.g. Digital Experience Analytics (2018); CrazyEgg (2018); Hotjar (2018)).

Later work focused on the use of pointer movements for user authentication (Revett et al., 2008; Gamboa, 2008), while more recent developments have applied pointer analytics as a behavioural methodology to relate user experience to underlying psychological processes. Arapakis and Leiva (2016) used machine learning to predict user engagement on the basis of temporal and spatial mouse movement features. Tzafilkou and Protogeros (2018) extracted mouse patterns (e.g. random movements or hovers) and found a relationship between these patterns and perceived ease use, perceived usefulness, self-efficacy, learning behaviour, and risk perception.

Pimenta et al. (2013) found a decrease in cognitive performance in relation to a reduction in mouse acceleration and velocity. Seelye et al. (2015) used just mouse movement variables to distinguish older adults with and without mild cognitive impairment. Some authors also found an association between pointer movements and emotions (Hibbeln et al., 2017; Zimmermann et al., 2003; Yamauchi and Xiao, 2018). Hibbeln et al. (2017) concluded that negative emotion can be inferred with an accuracy of 81.7%, based on increasing distances covered by and decreasing speed of the mouse cursor across an interface.

The general aim of the studies of this paper was to delineate a range of spatio-temporal pointer features and explore the feasibility of using these to detect response uncertainty when respondents answer questions presented in an online survey. Study 1 aimed to computer a first set of spatial and temporal features and validate the pre-processing procedures for pointer tracking analysis. The resulting features were then applied in Study 2, in which we defined pointer movements that specifically related to contextual features of the online survey interface (e.g. sequence and layout of survey questions and response scales). The initial selection of micro-behaviours was guided by the results of a previous study (Cepeda et al., 2018), by metrics that have been applied in eye tracking studies (see, e.g. Rayner (1998) and by observation of patterns of potential micro-behaviours in Study 1. This resulted in a new set of spatio-temporal features, referred to as micro-behaviours. Most of these micro-behaviours have not been reported in previous work. In Study 3, we integrated these features in an uncertainty model to explore their use to automatically detecting respondents' hesitation when answering questions. Following the approach of a previous study (Dias et al., 2019), this model was tested using the data set acquired in Study 2.

## 2. Study 1: Extraction of spatio-temporal features

### 2.1. Participants

A sample of 119 volunteers recruited via a pool of students of University of Zurich participated in this study. The participants, or subjects, were aged between 20 and 52 years old (M=25.4; SD=5.4; 18 male), native or fluent speakers of Standard German, consistently right-handed (Annett, 1970), healthy with normal, or corrected-to-normal, vision, no record of neurological or psychiatric illness, and no current medication use.

### 2.2. Procedure

Following recruitment and written and verbal information about the study, the participants received an email with a link to an online survey. After accessing the welcome page of the survey, participants provided informed consent, and self-report questions about their personality. The participant could decide when and where to complete the survey but was required to complete it in one session. After completion of the survey, each participant was informed that all data, including pointer movements, would be analysed and that they could withdraw their data from the study if they wished. Participants also had the option to provide their name and email address or phone number if they wished to receive any feedback about their answers in the survey, irrespective of whether they agreed to the use of the data.

### 2.3. Data acquisition

Acquisition of pointer movement data was implemented in LimeSurvey (LimeSurvey, 2018). This is a free and open-source survey web app that is easy to edit in HTML and enables coding in JavaScript. A JavaScript code was integrated to record mouse movements (see Fig. 1). This data was then sent to a server machine via AJAX for storage. AJAX is a client-side technique to asynchronously send and retrieve data from a server.

### 2.4. Pre-processing data

Some of the original pointer files had data collection errors that needed to go through a pre-processing phase to enable analysis.

#### 2.4.1. Server file correction

In terms of the mouse raw file acquired in the server, the existing problems and corrections implemented were:

1. Two different data lines are together, a paragraph is done between the two;
2. The counter of the data lines are not in the correct order;
3. Files without the number of frame column are identified;
4. If different files are from the same and subjects, they are concatenated;
5. Repeated positions $(x, y)$ in consecutive data lines are removed;
6. Repeated timestamp in consecutive data lines are removed;
7. Data lines with $Not\ A\ Number$ values are removed.

#### 2.4.2. Device identification

The usage of an online survey requires the identification of device type because pointer movement is absent from touch screen devices, the data from which is not further considered. To identify these devices, given that there are no pointer movements (represented as 0 in the pointer file), the predominant *EventCode* in touch devices files will be 1, as it is the identification of clicks. The file is not considered if the ratio between the events where the pointer is moving ($event = 0$) and the events where the mouse is down ($event = 1$) is less than 2:

$$\frac{\#events = 0}{\#events = 1} < 2$$

#### 2.4.3. Data validation

There are some possible errors that need correcting and applying to each server file before extracting information from the data. To this end, an output CSV file was generated with the following details described.

- Original number of subjects
  (#*original subjects*);
- Number of subjects with correct mouse data
  (#*subjects*);
- Number of subjects using touch screen devices
  (#*touch screen files*);
- Percentage of files with missing samples
  (%*files without samples*);
- Mean of percentage of samples missing
  (%*lost samples*);
- Percentage of files that are split
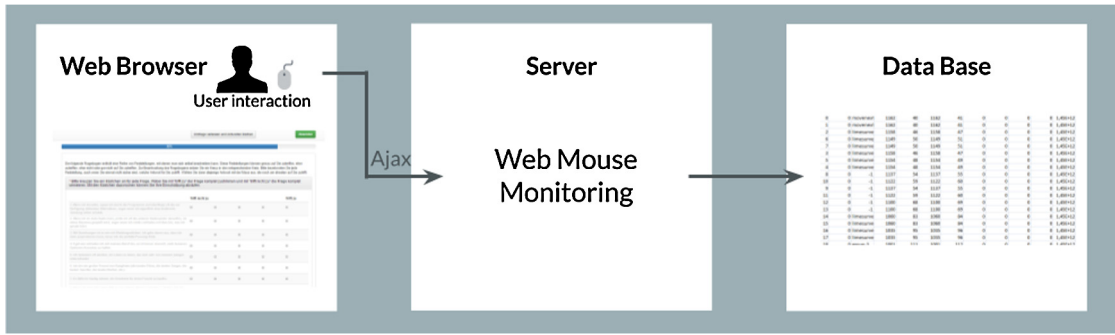  (%*concatenated files*).

**Fig. 1.** Architecture of dataflow from online surveys to database.

**Table 1**. Validation file output from Study 1.

| Result | Output |
|---|---|
| # original subjects | 119 |
| # subjects | 89 |
| # touch screen | 24 |
| % files without samples | 26 |
| % lost samples | 0.0087 |
| % concatenated files | 1.7 |

The result of the validation file from this study is presented in Table 1.

Of the original subjects, 24 used touch devices, meaning that no mouse data is available in these subjects for further analysis. Six subjects completed the survey but pointer movements were not reported. This might be because no mouse movements were detected around or within each survey question. These subjects were disregarded for further analysis.

### 2.5. Spatial information

In the spatial domain, the pointer movement in the survey (or questionnaire) is analysed as a single path, from the beginning until the end of the survey.

To smooth the spatial signal, a spatial vector $s$ representing the cumulative length along the path, between two mouses positions, is calculated:

$$s_i = \sum_{k=1}^{i} \sqrt{\Delta x_k^2 + \Delta y_k^2}, \qquad i = 1...n-1,$$

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

A cubic spline interpolation is then applied in order to produce a curve signal with a space interval equal to the mean value of the length variance:

$$\alpha = \overline{\Delta s},$$

$$\Delta s_i = s_{i+1} - s_i$$

An example of this pre-processing procedure is shown in Fig. 2, illustrating a fraction of a movement and a comparison between the original signal and the interpolated result.

The *survey path length* is easily extracted from s, $\bar{s} = s_{n-1}$ and expressed in pixels. To generalize this measure to other surveys, $\bar{s}$ is divided by the number of items.

To calculate the *angle* vector we used the following expression:

$$\theta = arctan\left(\frac{\Delta y}{\Delta x}\right) \qquad rad,$$

however, to avoid the radian phase discontinuities near $-\pi$ and $\pi$, we added multiples of $\pm 2\pi$:

$$\theta_i = arctan\left(\frac{\Delta y_1}{\Delta x_1}\right) + \sum_{j=1}^{i} min\left\{\Delta \, arctan\left(\frac{\Delta y_i}{\Delta x_i}\right) + 2k\pi\right\},$$

$$k \in Z$$

The *curvature* is inversely proportional to the circle's radius created at the tangent point of the path in study. The curvature is expressed by:

$$c = \frac{x'y'' - y'x''}{(x'^2 + y'^2)^{3/2}}$$

We have interest only in the absolute values of angles and curvatures. The rate of change in curvature is given by the expression:

$$c' = \frac{\Delta c}{\Delta s}$$

Fig. 3 represents the angle and curvature results from the path represented in 2.

From each stroke defined, we can calculate its *length* and *straightness*, which is defined as the ratio of the Euclidean distance between the start and end of the stroke and the total distance travelled (length):

$$straightness = \frac{\sqrt{(x_1 - x_n)^2 + (y_1 - y_n)^2}}{s_{n-1}}$$

The tremors in the user movements were measured by *jitter*, which corresponds to the ratio between the original path length and the smoothed path length:

$$jitter = \frac{s'_{n'}}{s_{n-1}}$$

Table 2 summarizes and describes the final set of spatial features. This includes a representative symbol, the feature, a short name, its unit, distribution of density probability of the existing values and the observed ranges in our population. The range of values for every feature is within expected and acceptable values, that confirms the correct application of pre-processing tools.

### 2.6. Temporal information

In the temporal domain, the interaction of the pointer with the survey was not considered as a whole but as a set of strokes, depending on the interval between sequential movements. When a subject takes more than $k_{stroke}$ seconds to move the pointer, the movements before and after that pause were considered strokes.

In order to have a signal with equal temporal spaces, we applied a cubic spline interpolation to each stroke. The interval delimited is proportional to the mean variance of time:

$$\alpha = k_{time\,interp} \times \overline{\Delta t},$$

$$\Delta t_i = t_{i+1} - t_i$$

Fig. 4 shows an example of a curve interpolated from the x and y original signals. It is possible to identify different strokes, pauses and discretely distinguishable intervals of interpolations for each stroke.

The *total time* of the survey ($t\_total$) was easily extracted from t, $\bar{t} = t_{n-1}$ and expressed in seconds. To generalize this measure to other surveys, $\bar{t}$ was divided by the number of items.
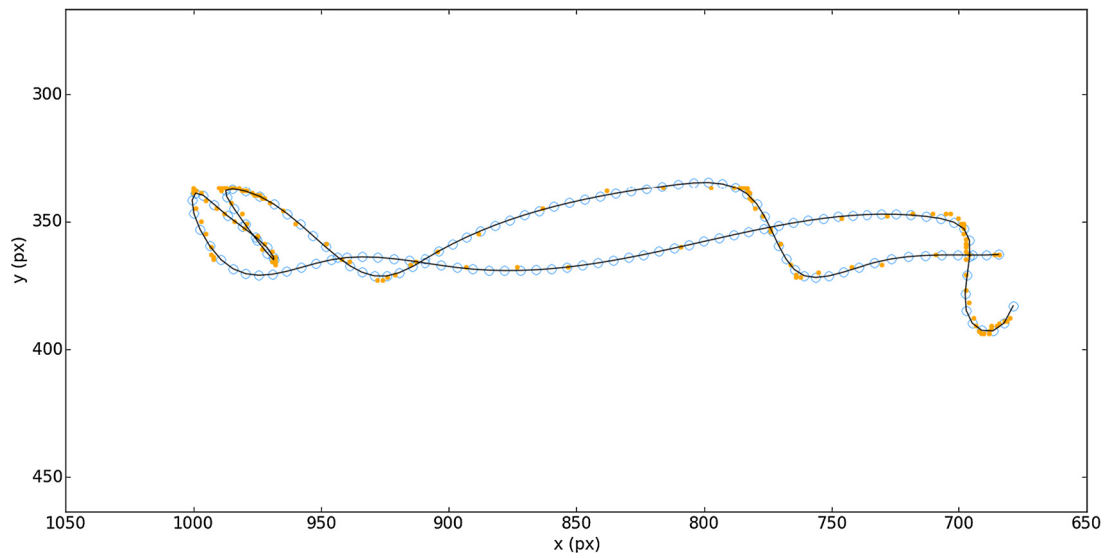
**Fig. 2.** Spatial signal x-y representation in pixels. The orange dots (·) represent the signal extracted from the pointer movement and the blue dots (o) with constant spacing and the black line (-) represent the interpolated curve of these movements.
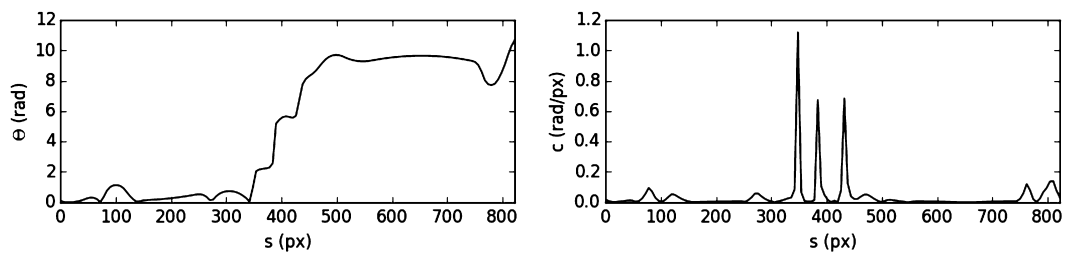


**Fig. 3.** The left panel shows the angles in radians over distance in pixels and the right panel shows the curvature in radians/pixels over distance in pixels.

**Table 2**. Details of features extracted in the spatial domain.

| Symbol | Feature | Name | Unit | Distribution | Range |
|---|---|---|---|---|---|
| | Length | $\bar{s}$ | px | | $[2.8 \times 10, 4.0 \times 10^3]$ |
| | Angle | $\theta$ | rad | | $[-3.1 \times 10, 6.2 \times 10]$ |
| | Curvature | $c$ | $rad.px^{-1}$ | | $[-8.2 \times 10^2, 6.6 \times 10^2]$ |
| | Variation Curvature | $c'$ | $rad.px^{-2}$ | | $[-4.3 \times 10^5, 8.3 \times 10^5]$ |
| | Strokes Length | $s\_stroke$ | px | | $[2.2, 9.0 \times 10^3]$ |
| | Straightness | $straightness$ | - | | $[8.1 \times 10^{-3}, 1.0]$ |
| | Jitter | $jitter$ | - | | $[2.8 \times 10^{-1}, 9.8 \times 10^{-1}]$ |

4
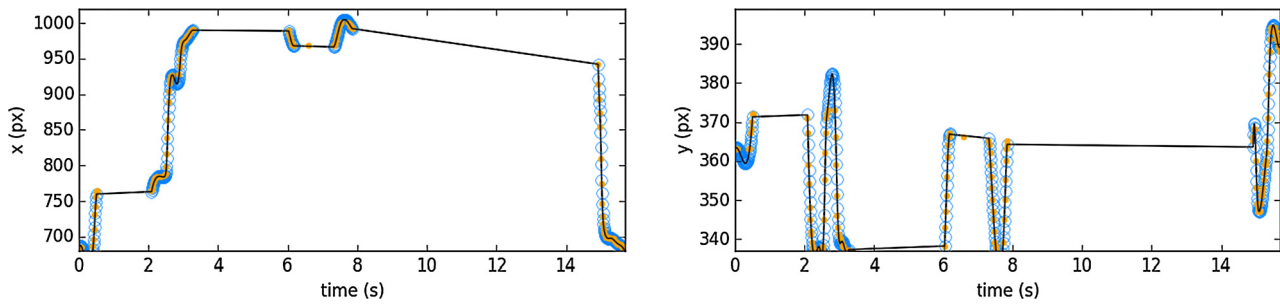
**Fig. 4.** Temporal signal x-y representation in pixels. The orange dots (·) represent the signal extracted from the pointer movement and the blue dots (○) with constant and defined spacing and the black line (-) represent the interpolated curve of these movements.

**Table 3**. Details of features extracted in the temporal domain.

| Symbol | Feature | Name | Unit | Distribution | Range |
|---|---|---|---|---|---|
| | Total time | $t\_total$ | $s$ | | $[3.5, 2.8 \times 10]$ |
| | Velocity | $v\_t$ | $px/s$ | | $[0.0, 1.7 \times 10^5]$ |
| | Horizontal Velocity | $v\_x$ | $px/s$ | | $[0.0, 1.1 \times 10^5]$ |
| | Vertical Velocity | $v\_y$ | $px/s$ | | $[0.0, 1.6 \times 10^5]$ |
| | Angular Velocity | $w$ | $rad/s$ | | $[-9.1 \times 10^1, 1.1 \times 10^2]$ |
| | Acceleration | $a$ | $px/s^2$ | | $[-1.7 \times 10^8, 1.7 \times 10^8]$ |
| | Jerk | $jerk$ | $px/s^3$ | | $[-1.7 \times 10^{11}, 1.7 \times 10^{11}]$ |

To correctly calculate the velocity of the pointer movement, a vector which includes the velocity values when the mouse moves and consider the velocity zero when no movement need to be computed. It is named *velocity* ($v\_t$).

From temporal information we also extract the *horizontal velocity* ($v\_x$), *vertical velocity* ($v\_y$), *acceleration* ($a$), *jerk* ($a'$) and *angular velocity* ($w$):

$$v_x = \frac{\Delta x}{\Delta t} \quad v_y = \frac{\Delta y}{\Delta t} \quad a = \frac{\Delta v_t}{\Delta t}$$

$$a' = \frac{\Delta a}{\Delta t} \quad w = \frac{\Delta \theta}{\Delta t}$$

Table 3 summarizes and describes the final set of temporal features. This includes a representative symbol, the feature, a short name, its unit, distribution of density probability of the existing values and the observed ranges in our population. The range of values for every feature is within expected and acceptable values, that confirms the correct application of pre-processing tools.

## 3. Study 2: Extraction of micro-behavioural features

### 3.1. Participants

A sample of 88 volunteers was recruited for this study at University of Zurich via flyers. The participants were aged between 18 and 35 years old (44 male) and received 20 Swiss Francs or credit points for participation. All participants were native or fluent speakers of Standard German, consistently right-handed (Annett, 1970), healthy, with normal, or corrected-to-normal, vision, no record of neurological or psychiatric illness and no current medication use.

### 3.2. Procedure

Each participant was tested individually by a Master's student in a small, sound-attenuated, dimly lit experimental room. First, informed consent and demographic data were collected via paper and pencil. Then, 60 items of personality assessment were collected in an online survey.
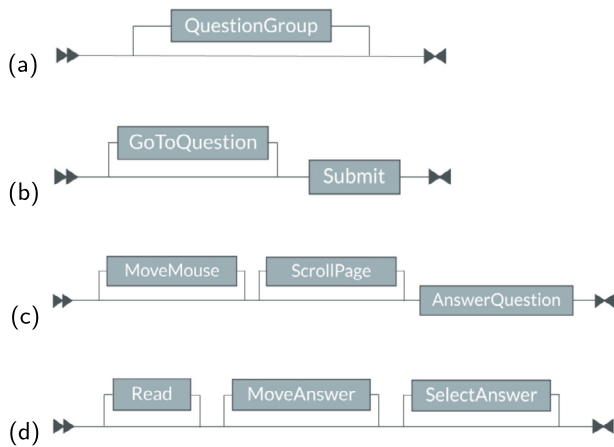
### 3.3. Interactions between pointer and content

To the best of our knowledge, no previous studies have analysed pointer interaction in the specific context of on-line surveys. We explored the use of different features that could be used in combination to describe patterns of pointer movements while a subject processed the online survey and answered survey items. We refer to these as micro-behavioural features as these represent distinguishable sets of spatio-temporal position and movement behaviours in relation to the structure of contextual features (e.g. item sequence, layout of item question, position of and response scale for each item) in the survey. The following describes the approach to building up a set of features that could be extracted from pointer interactions.

The first step is to enter the survey. It is possible to close the webpage and leave the survey or to go to the first question group and then to the next group (see Fig. 5a).

Within the set, or group, of questions (see Fig. 5b), the subject can respond to every question in any order. The participant can submit the survey once all questions are answered.

The subject can move the pointer to a question and then provide an answer, as indicated by movement of the pointer or scrolling to the question (Fig. 5c). It is possible to extract as features the number of

**Fig. 5.** Some interactions between pointer movement and survey content are straightforward. These include entering the survey and moving the pointer to a page with or a group of questions (panel a), or submitting (i.e. leaving) the survey at some point (panel b). Other interactions are more complex and allow extraction of interactions with contextual features of the survey, such as moving the pointer or scrolling to and answering the next question (panel c) and reading a question, changing and selecting (or re-selecting) an answer (panel d).

scrolls conducted and the number of items to which scrolling was performed.

The sequence of steps involved in answering a questionnaire item (Fig. 5d) is the most complex and allows extraction of more contextual features. The most efficient approach is to read the item, move the pointer to the response option and to select that option. However, if the subject then moves the pointer to and selects an alternative option, this is considered a response *correction within an item*. If, after having answered another item or items, the subject returns to and changes the response to an item that the subject already answered, this is considered an *correction between items*. After proceeding to further items, the subject could revisit a previously answered question without changing the answer. This event is defined as *revisiting previous item*.

### 3.4. Micro-behaviours

Micro-behaviours include what we refer to as *overview (# overviews)*. This is characterized by a scrolling of the cursor over a wide area across the interface of the survey. This often occurs in order to develop an overview of the survey structure and content (e.g. length of the survey, number and type of questions) (see Fig. 6). At the beginning of the survey, this particular subject navigated to the end of the survey and then returned to the first question. Computationally, an overview was defined as when the subject crossed more than one-quarter of the items of the whole survey.

The second observed behaviour was the *skip pattern* (#*skips*). When answering the survey some subjects would not follow the natural order of questions but skip questions and answer the items in an arbitrary order. This behaviour is represented in Fig. 7, which shows a subject responding to 14 after answering questions one and two. To compute this feature as a skip, we verified the order of the items answered, without considering further corrections, and coded if the user moves back to a previous question. The final feature corresponds to the number of skipped items and, given that it relates to the number of items of the survey, we normalized this feature dividing by the total number of items.

Three features relating to pauses in movement were considered: time of interaction, time of pauses and number of strokes. The *time of pauses* ($t\_pauses$) is a vector with the interval of times that people remain without interaction more than $k_{pause}$ seconds between strokes and the *number of strokes* (#*strokes*) corresponds to the number of times that

people move the mouse between pauses. The *time of interaction* is the total time that people are moving the mouse, excluding from the total time the time of pauses. We also identified pauses ($k_{pause} = 1\,second$) between pointer movements (or strokes) (see Fig. 8). The number of strokes and time of interaction are normalized according to the length of the survey, and therefore, divided by the total number of items.

The following features are used verify if the subjects are not wandering around the page with the pointer but remain in the same question or if they scroll down the page and come back to the same item. We referred to such wandering as a zapping event and, despite being considered as features, we re-moved them from the signal for further analysis. Based on the time that people remain inside an item area, a zapping item is classified when this interval of time is less than $k_{zapp}$ seconds. When a subject is scrolling, many zapping items are detected and, therefore, we classified this event as a zapping event and considered the items that were crossed by the pointer. From the whole survey we extracted the *number of zapping events* (#*zapp*) and the *number of zapped items* in all zapping events (#*items_zapped*). The last feature was normalized, as divided by the total items in the survey.

The spent *time in each item* is kept in a vector ($t\_item$), which with the zapping corrections gives us more realistic information, ignoring short times caused, for example, by scroll actions.

Sometimes, external factors could interrupt the subject focus on the survey, or the subject could abandon the survey for a while to answer the telephone call, for example. We considered the *number of abandons* (#*abandon*), defined by:

$$Abandon = k_{abandon} \times mean(t_{item})$$

The sum of the *time of abandons* ($t\_abandon$) was also considered as feature. When an abandon event is identified, the respective time value is removed from $t\_item$. This way, this vector will contemplate only the time spent to answer a question, having just normal situation values.

The $t\_item$ consider only the time between the mouse enter the item and leaving it. To consider the total duration that the mouse was inside each item, we calculated the *accumulated time* ($t\_accum$).

When the mouse is inside a question, we identified two different ways of reading the question: moving the pointer to the area of text containing the question while reading the question or moving the pointer around the area of the response options. To identify this so-called *hovering over the text* (#*hover_text*), the x mouse coordinates are associated with question or response area, after defining the width of the text of the question. The associated feature was the accumulative time that the subject dwelled with the pointer in the question area. To be independent of the question area (as these can vary in size deepening on question length), this feature was divided by the number of items.

Further hover features were computed. The *number of hovered answers* ($answer\_hover$) corresponds to the number of hovered answers divided by the total number of possible answers. The *selected answer ratio* ($answer\_ratio$) is defined by the duration of hovering the final answer, in relation to the total time in the answers' area.

Click behaviour was considered. The *time before click* ($t\_bef\_click$) after having entered the area of the corresponding item question was calculated as the sum duration of time spent in a question until the first click. The *pause before click* ($pause\_bef\_click$) as the time taken between the last pointer movement in the area of an item and the click on a response option for that item. If the participant clicks more than once in a single question (to correct a previous answer), this value is averaged. The time between a click in and click out when selecting an item response was also calculated (*time click* ($t\_click$)).

Because some users move the mouse around the final answer, the distance from the path inside a question to the selected answer was also computed. This *distance from answer* ($distance\_answer$) is given by the equation (1), where $x_{answer}$ and $y_{answer}$ are the x and y coordinates of the question's last click.
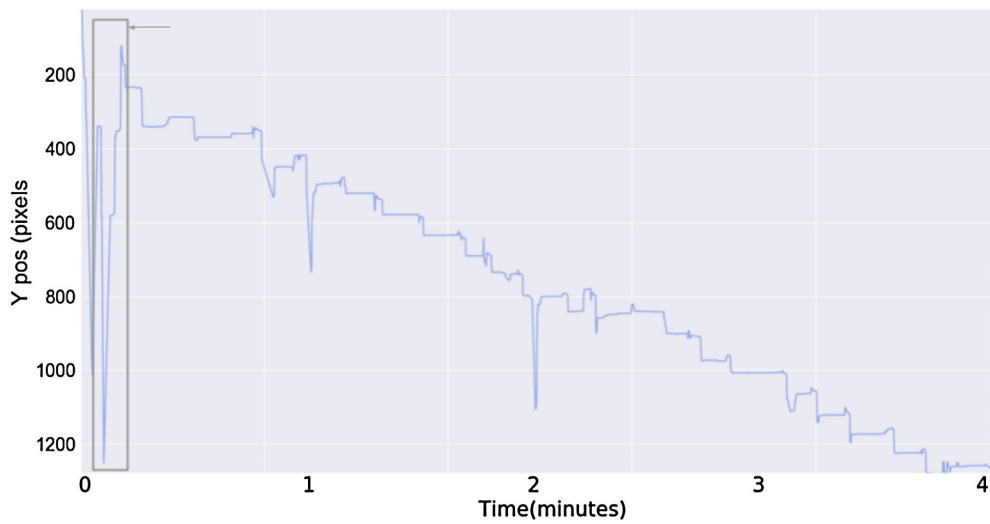
**Fig. 6.** Representation of a subject's pointer movement over time, with movement relating to the first question at the top left of the figure (smallest y value) and the last question on the far right. The rectangular panel indicates a pointer movement pattern of the subject scrolling across the interface of the survey.
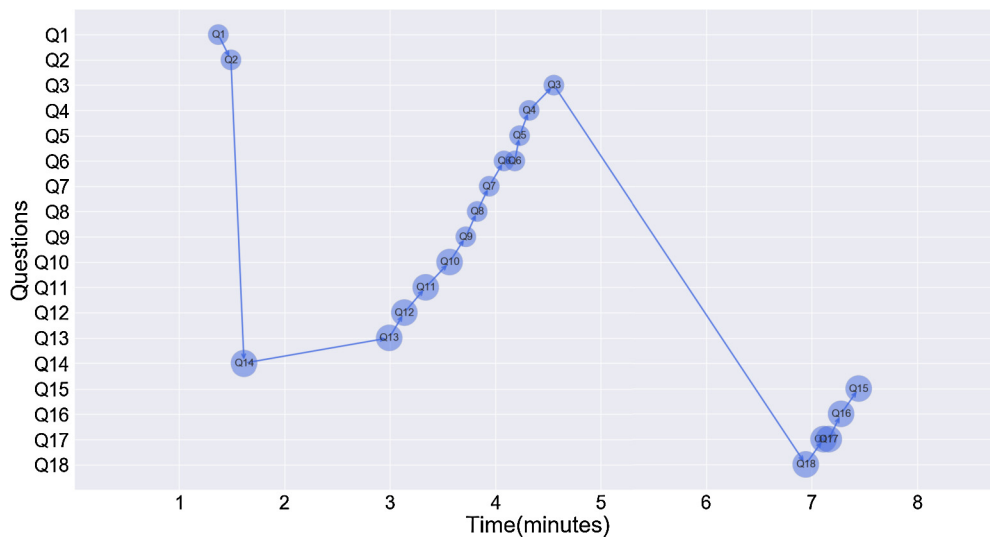


**Fig. 7.** Illustration of a subject's pointer movement behaviour, showing that this subject answered questions 1 and 2, skipped forward to question 14, and then worked backwards from question 14 to question 3.
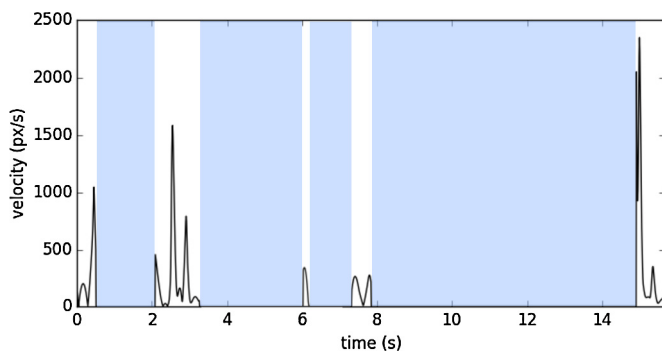


**Fig. 8.** Illustration shows the velocity of pointer movements over the course of the survey. The blue shaded panels show the pauses between pointer movements (or strokes). In this example, there are a total of 5 strokes and 4 pauses.

$$Distance\,from\,answer =$$

$$\sqrt{(x_i - x_{answer}) ** 2 + (y_i - y_{answer}) ** 2},$$

$$i = 1, ..., n - 1 \quad (1)$$



**Fig. 9.** Illustration of a subject's <-turn pattern. The blue line depicts the pointer movement and the red dot shows the mouse click.

While thinking about the answer, some subjects change the mouse horizontal direction, which we called $<-turns$ (see Fig. 9). This was calculated by the horizontal trajectory's derivative changes from positive to negative values or vice-versa.

When the individual selects one option, but keeps interacting inside the item and decides to change the option selected to another answer, we defined this behaviour as *correction within item* (#$correc\_within\_item$).

Before leaving the question, the interval of time spent between click and go to the next item was calculated and named *inter-item interval* (#$inter-item\_interval$).

A different approach to correcting the previous answer is the *correction between item* (#$correc\_between\_item$). In this case, the person selects an answer, move forward to the next questions, and after answering at least one more question, returns to and changes the previous response.
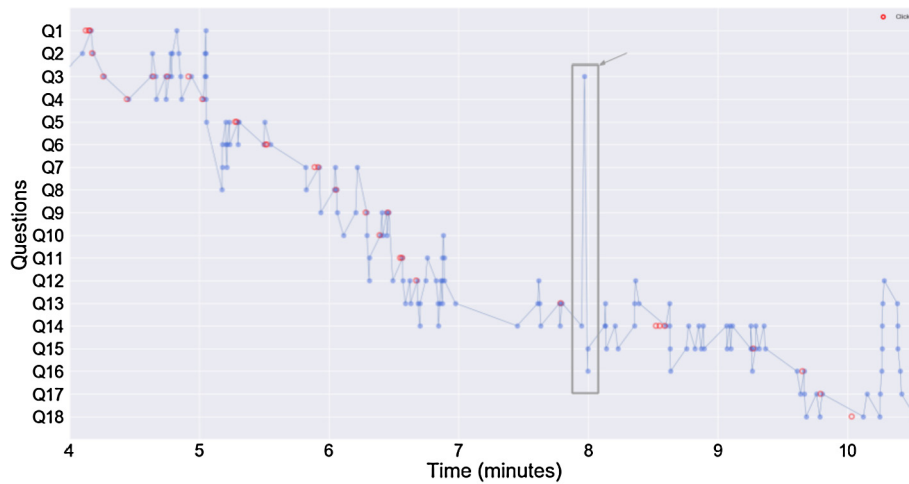
**Fig. 10.** Illustration of the sequence of visits to the survey questions. The red dot depicts mouse clicks. In this case, the subject revisited and re-answered question 3 after attending question 14 (as highlighted in the rectangle).

Instead of changing the previous response, the subject simply returns to a previous item response, a so-called *re-visit* (#*revisits*). This is illustrated in Fig. 10, in which the subject revisits a prior answer (moving from question 14 to question 3). Interestingly, after answering item 3 the first time, this subject then changed the response to question 4 and then returned to question 3. This revisit took place around three minutes after the initial responses.

The *correction time* was calculated by the sum of all the time intervals in a question from the first click until the last click (last correction). This value is considered as zero if there is no correction.

Table 4 summarizes the final features and classifies them. This includes a representative symbol, the feature, a short name, its unit, distribution of density probability of the existing values and the observed ranges in our population. The range of values for every feature is within expected and acceptable values, that confirms the correct computation of features.

## 4. Study 3: Applicability of features to detect uncertainty

While identifying pointer movement features, we considered whether behaviours such as the duration of hover over a question, the time that elapses to provide a response, or whether items or their responses are revisited or corrected might be indicators of response uncertainty.

Although previous studies consistently used the response time as an indicator of response difficulty (Conrad et al., 2007; Schneider et al., 2015; Zushi et al., 2012), others have confirmed the influence of the pointer movement trajectory in predicting response uncertainty. The trajectory has been assessed in terms of horizontal direction inversions (Zushi et al., 2012) and deviation from the idealized straight-line trajectory (Schneider et al., 2015). More recently, Horwitz et al. (2017), used mouse cursor trajectories to predict response difficulty, achieving a performance accuracy of between 74% and 79%. 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 (Horwitz et al., 2017).

Study 3 aimed to create a machine-learning model that identifies events of response uncertainty, using some of the previously described features of mouse movement while subjects processed and answered survey items. The participants and procedure are the same as in Study 2.

### 4.1. Methods

#### 4.1.1. Features extraction

To focus on item-specific uncertainty events, we selected features that relate to an item instead of the whole survey (such as the *overview* feature). Both temporal, spatial and contextual features were used to detect items that could be associated with uncertainty:

- Length;
- Straightness;
- Velocity;
- Accumulated time;
- Hovered answers;
- Selected answer ratio;
- Time before click;
- Pause before click;
- Distance from answer;
- <-turn;
- Correction within item;
- Revisit;
- Correction time;
- Interactions.

To extract the features the constants were defined: $k_{pause} = 1\ second$; $k_{zapp} = 0.1\ second$; $k_{aband} = 10 \times mean(t_{item})\ seconds$. The last feature, *interactions*, is the only new feature, which corresponds to the number of interactions with each question (i.e., the number of times in each question).

To account for general differences in processing time by different subjects, the features were normalized for each person separately using the formula presented in equation (2), where $z_i$ represents the sample $x_i$ after normalization, $\overline{x}$ and $\sigma$ are the mean and standard deviation of the samples, respectively (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).

$$z_i = \frac{x_i - \overline{x}}{\sigma} \qquad (2)$$

With all the features normalized, it is only possible to identify the most difficult questions for each individual. 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 to standardize the range of the variables for all the participants.

**Table 4**. Features of micro-behaviours.

| Symbol | Feature | Name | Unit | Distribution | Range |
|---|---|---|---|---|---|
| | Overview | #overviews | - | | $[0, 2]$ |
| | Skip | #skips | - | | $[0, 1.6 \times 10^{-1}]$ |
| | Time of Pauses | t_pauses | s | | $[1.0, 3.8 \times 10^{1}]$ |
| | Strokes | #strokes | - | | $[0.4, 2.5]$ |
| | Time of Interaction | t_interaction | s | | $[0.1, 8.1]$ |
| | Zapping Events | #zapp | - | | $[0, 46]$ |
| | Items Zapped | #items_zapped | - | | $[0, 2.3]$ |
| | Time per item | t_item | s | | $[0.0, 1.8 \times 10^{2}]$ |
| | Abandons | #abandon | - | | $[0, 12]$ |
| | Time of abandons | t_abandon | s | | $[0, 1.2 \times 10^{2}]$ |
| | Accumulated time | t_accum | s | | $[0.6, 1.8 \times 10^{2}]$ |
| | Hovering text | #hover_text | - | | $[0.1, 0.3]$ |
| | Hovered answers | #ans_hovered | - | | $[0.2, 1]$ |
| | Selected answer ratio | answer_ratio | - | | $[9.9 \times 10^{-3}, 1]$ |
| | Time before click | t_bef_click | s | | $[0, 51]$ |
| | Pause before click | pause_bef_click | s | | $[0, 17]$ |
| | Click Time | t_click | s | | $[0, 5]$ |
| | Distance from answer | distance_answer | px | | $[2.2, 6.5 \times 10^{2}]$ |
| | <-turns | < −turns | - | | $[0.1, 6.9]$ |
| | Correction Within Item | #correc_within_ item | - | | $[0, 0.5]$ |
| | Inter-item Interval | inter − item_interval | s | | $[2.4 \times 10^{-2}, 1.8 \times 10^{2}]$ |
| | Correction Between Item | #correc_between_item | - | | $[0, 0.9]$ |
| | Revisits | #revisits | - | | $[0, 4.8]$ |
| | Correction time | t_correc | s | | $[0, 179]$ |

### 4.1.2. Features selection

Some machine learning classifiers are not sensitive enough to detect the influence of relevant features in the presence of many variables (Sperandei, 2014). It is therefore helpful 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 used for this and, if two features had an absolute coefficient higher than 0.9, one of them was left out.

### 4.1.3. Model training and testing

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 uncertainty in this proof of concept study. To solve this, mouse movement videos of 6 individuals answering a 60 item survey (360 questions in total) were observed and rated by three independent raters in terms of low or high uncertainty. The final examples of items for training and testing were selected only if rated as uncertain or certain by at least 2 of the raters. In the end, 51 items were rated as uncertain and 124 as certainty. The remaining 185 items were not rated by one rater and were excluded from further analysis.

*10-fold cross validation* was applied for model training and testing. In this procedure, the data is divided into 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.

### 4.1.4. 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 the 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 (3) and (4), where $p$ represents the probability of an event, $\beta_i$ illustrates the regression coefficients and $x_i$ are the input features (Sperandei, 2014).

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n \quad (3)$$

Solving for $p$,

$$p = \frac{1}{1 + e^{-(\beta_0 + ... + \beta_n x_n)}} \quad (4)$$

When $p \geqslant 0.5$ it is predicted $Y = 1$ (uncertainty), otherwise, $Y = 0$, where $Y$ is the outcome variable (Rohilla Shalizi, 2019). From equation (4), it is possible to verify that a positive $\beta_i$ increases (and a negative $\beta_i$ decreases) the probability of $Y = 1$.

### 4.2. Model evaluation

In binary classification, data is constituted by two opposite classes, positives and negatives. Accordingly, the possible outcomes comprise TP, TN, FP and 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

**Table 5.** The ten most relevant features.

| Feature | Regression coefficient |
| --- | --- |
| <-Turn | 1.47 |
| Length normalized | 1.23 |
| Length | 1.19 |
| Distance from answer normalized | −0.93 |
| Interactions | 0.65 |
| Accumulated time | 0.61 |
| Straightness | −0.49 |
| Pause before click | 0.31 |
| Corrections between item | −0.31 |
| Distance from answer | −0.29 |

in equation (5). *Specificity*, on the other hand, provides the probability of a question associated with certainty being correctly classified and it is illustrated by equation (6).

$$Sensitivity = \frac{TP}{TP + FN} \quad (5)$$

$$Specificity = \frac{TN}{TN + FP} \quad (6)$$

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 (7).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

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 (8) as the harmonic mean between *precision* and *recall*. *Recall* is a synonym of *sensitivity*, as it is possible to verify in equation (9). *Precision*, on its turn, represents the probability of a certainty event being classified as an uncertainty event, as shown in equation (10) (Sun et al., 2007).

$$f1\ score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (8)$$

Where

$$Recall = Sensitivity \quad (9)$$

And

$$Precision = \frac{FP}{FP + TN} \quad (10)$$

### 4.3. Results and discussion

Feature selection resulted in the elimination of the features *time before click*, *hover selected answer*, *straightness normalized*, *revisits*, *revisits normalized* and *hovered answers normalized*. These features were highly inter-correlated with other features, containing therefore a high degree of redundant information. Their elimination is likely to enhance the machine learning performance (Langley and Sage, 1997). The final number of features was 24.

Table 5 shows the ten most relevant features for classification of uncertainty, based on the reported strength of the regression coefficients (in the order of the highest to the lowest absolute value). On this basis, the number of *<-turns* is the most relevant feature for classification. Use of this feature increases the probability of detecting an uncertainty event. This feature means that subjects tend to move the pointer across the alternative response options more frequently when uncertain. This is in line with Zushi et al. (2012).

The feature *distance* suggests that subjects tend to move the pointer more frequently from one response option to another while deciding which to select, consistent with Zushi et al. (2012). The feature *distance from answer* indicated that individuals tend to keep the pointer closer to the finally selected alternative when uncertain, even though the pointer moves longer distances when uncertain.

**Table 6**. Model performance evaluation measures.

| Sensitivity | Specificity | Accuracy | F1 Score |
|---|---|---|---|
| 0.78 | 0.94 | 0.89 | 0.81 |
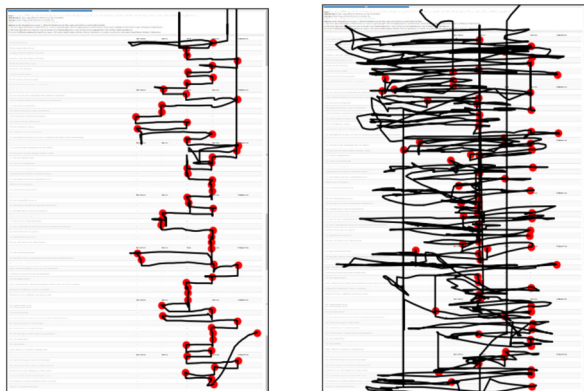


**Fig. 11.** Pointer movements in an online survey. The left panel shows the subject with the least uncertainty and the right panel shows the subject with the greatest uncertainty.

The analysis of the regression coefficient of *interactions* revealed that subjects tend to visit items more often that, appear to be associated with greater uncertainty. In these items, individuals take longer to give a response (*accumulated time* and deviate more from the straight-line trajectory between successive answers (*straightness*).

Interestingly, the results show that the number of corrected responses is associated with a decrease in the probability of identifying an uncertainty event. This suggests that pointer movements reveal less uncertainty as subsequent corrections reflect great certainty by the subject in their corrections.

### 4.3.1. *Model evaluation*

The model evaluation measures - *sensitivity*, *specificity* and *accuracy* - are presented in Table 6.

The evaluation of the uncertainty model in terms of its classification performance showed a *sensitivity* of 0.78 and *specificity* of 0.94. The former indicates that uncertainty events were correctly classified in 78% of the times. The latter indicates that 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., 2017). This improvement might relate to the choice of features used to indicate uncertainty. Using *F1 score* as another way of indicating classification accuracy using a single value, the estimated performance of the model was 0.81. This value suggests that the performance of the model is very good.

Following the application of the model to all participants' questions, the percentage of questions associated with uncertainty was computed. Fig. 11 shows the difference in pointer movements between the two individuals with the greatest difference in uncertainty, based on our features. Visual inspection of this example reveals a sense of how the two individuals can be differentiated in terms of features such as turns (within a question), distance (between response options), distance from answer (when making a decision), accumulated time (to give a response) and the straightness (or deviation from the straight-line trajectory between successive answers).

## 5. General discussion

This study presents an exploratory investigation of pointer activity collected during user interaction with an online survey. The results shows that a broad range of spatial and temporal pointer features and micro-behaviours (distinct combinations of pointer movements) in interaction with the survey.

The seven spatial features described in this study were previously reported for biometric analysis (Gamboa, 2008). Angles and curvatures were used to analyse web browsing behaviour (Chudá and Krátky, 2014). Curvatures were also shown to relate to emotions (Yamauchi and Xiao, 2018) and, together with length, to predict user engagement (Tzafilkou and Protogeros, 2018; Arapakis and Leiva, 2016). None of these features was used in the context of surveys.

Of the temporal features, the velocity of pointer movement is common to all such studies. Velocity was used in connection with biometry (e.g. Gamboa, 2008), emotion (e.g. Hibbeln et al., 2017), user engagement (e.g. Arapakis and Leiva, 2016), web browsing (e.g. Ahmed and Traore, 2007) and usability studies (Arroyo et al., 2006). The acceleration and angular velocity were used in contexts of biometry (Gamboa, 2008) and web browsing (Chudá and Krátky, 2014). Acceleration is also used to predict user engagement (Arapakis and Leiva, 2016). The horizontal and vertical velocity and the jerk are only reported in a study of biometry (Gamboa, 2008). Total time was just used to predict user engagement (Tzafilkou and Protogeros, 2018). However, the features were never used before in the context of online surveys.

The delineation of micro-behaviours in the pointer data is a main contribution of this work, as most of these have not been reported previously. The number and duration of pauses have been used to assess usability, biometry, mood or cognitive abilities (Arroyo et al., 2006; Gamboa, 2008; Seelye et al., 2015; Zimmermann et al., 2003). The analysis of hovering events was also performed in several studies (Huang et al., 2011; Tzafilkou and Protogeros, 2018; Seelye et al., 2015; Arapakis and Leiva, 2016). For example, Arapakis and Leiva (2016) analyzed the hover distributions and clicks to verify the number of search results hovered before the user clicks. The click duration and the direction change (i.e. <-turn) were used to predict user engagement (Arapakis and Leiva, 2016), emotions (Yamauchi and Xiao, 2018) and mood (Zimmermann et al., 2003). Pauses before clicks were also used to analyse usability (Arroyo et al., 2006) and user engagement (Arapakis and Leiva, 2016). Two patterns only considered in an usability study were the hovering text and scrolling events (Arroyo et al., 2006). To the best of our knowledge, none of the remaining reported micro-behaviour in this work has been considered elsewhere.

The three studies in this paper represent a proof of concept. The aim of the first and the second studies was to develop a data processing procedure and to identify a set of pointer features for further use. The aim of the third study was to explore the feasibility of using these features to detect subjects' uncertainty when making responses. We used averaged accuracy measures in order to evaluate the overall uncertainty model's accuracy (based on a set of training data) as a means to demonstrating potentially feasibility. These first results suggest that this approach is promising and, on this basis, suggest that this approach warrants further investigation.

The use of prediction reliability estimates is important as these indicate the degree to which the uncertainty classifications are individually predictive of uncertainty. These estimates help to validate predictions derived from our classification and regression models. This is lacking in Study 3. To this end, a further study is needed that includes subjective report of response certainty (i.e. *stated choice certainty*) after answering each item. This would enable us to validate reliability estimations of uncertainty based on pointer movement data against the criterion of subjects' subjective reports of their choice certainty. These reports must therefore be reliable. Experimental control of factors that might modulate responses to and stated choice certainty for each item should, therefore, be considered in the experimental design (Mattmann et al., 2019). This would allow examination of, for example, whether knowledge of having to state choice certainty influences item responses, how stable item responses are with and without this requirement across testing (e.g. test-retest) and different surveys, depending on what information the subject is asked, and how this requirement influences

consistency of responses across the items of a survey (Mattmann et al., 2019).

While this approach would help to validate the use of pointer location, movements and clicks as a measure of choice uncertainty, there is a general dearth of literature on the relationship between pointer activity and underlying perceptual, cognitive and affective psychological mechanisms. Most research on dynamics of pointer movements relates to improving design of interfaces (e.g. Chan et al., 2001). Generally, pointer movements are considered to indicate the user's attention (Arapakis and Leiva, 2016) in that the pattern of pointer location, movement to a location (i.e. pointing) and click behaviour relate closely to what the person is processing (cf. eye-mind hypothesis, Just and Carpenter, 1980). For practical purposes, such as improving design of interfaces, these may be considered self-explanatory without the need for interpretation (cf. this view in eye tracking, Holmqvist et al., 2011).

However, these patterns may indicate specific underlying cognitive and emotional states. In applied research, for example, pointer movements such as hover and scrolling are reported to accurately reveal a web users search intent, interest and satisfaction with the search results (Guo and Agichtein, 2010; Huang et al., 2011). The use of these same features also revealed effects in this paper. However, in the absence of further studies on stated choice certainty and associated pointer activity, as suggested in the preceding, it is difficult to characterize latent processes underlying choice making. We cannot assume a one-to-one correspondence when interpreting, for example, hover and scrolling behaviour in web-based searches and hover and scrolling behaviour when providing choice responses to items under uncertainty, in part because the specific task is different and may invite different interpretation (Schindler and Lilienthal, 2019).

It is possible in the context of our specific task that these features (using hover and scroll to illustrate) indicate increased use of attentional resources while resolving conflicting information within an item (i.e. hover) (cf. Glaholt et al., 2009) in order to provide a response among several possible alternative options. Scrolling might facilitate choosing a response that is coherent with other responses in the survey (cf. Choi and Pak, 2005). Alternatively, hover might reflect increased cognitive effort while retrieving relevant information (cf. Rayner, 1998; Findlay and Gilchrist, 2008) from memory in order to respond to the specific question. Accurate inference of specific cognitive process from patterns of pointer activity requires careful experimental design that allows task-specific interpretation. For example, the pointer's movement trajectory has been shown to be modulated by underlying neural processes (Dhawale et al., 2017) that relate, for example, to processing of sensory information (e.g. Faisal et al., 2008), motor planning (Churchland et al., 2006) and motor execution (Jones et al., 2002). Similarly, studies of eye movement behaviour show, for example, that dwell time (which is comparable to our measure of hover) is a measure of visual attention, relates to specific underlying neural processes during decision making, and can influence response selection when subject dwell longer on information (Glaholt et al., 2009; Lim et al., 2011; Armel et al., 2008).

The selection via mouse click of a graphical element with a pointing device is a primary task in web-based interfaces. The analysis of patterns of mouse clicks, using for example clickthrough data (Farris et al., 2010) provides insight into user behaviour for the evaluation of webpages (Farney, 2011). However, mouse clicks alone do not indicate why the user chooses and clicks on a particular element, which information was considered before making that choice, whether there was conflict between possible options in the decision making process before making a choice, how certain the user was of the choice, or, for example, how the user subjectively experienced the web page content where that choice was made (Joachims et al., 2017; Agichtein et al., 2006; Tzafilkou et al., 2014). By using its high temporal sensitivity to capture a real-time trace of the dynamics of pointer-based HCI and the interaction of the pointer with specific web content, pointer-based measures of choice uncertainty might give insight into these aspects of web-based

information search, choice evaluation and decision-making. Assuming that patterns of pointer behaviour can be shown to relate to internal cognitive processes (Freeman et al., 2011), this could contribute to developing a better understanding of latent (i.e. only indirectly observable) aspects of users' subjective experience, behaviour and responding during HCI (Stillman et al., 2018).

## 6. Conclusion

This exploratory investigation delineated spatio-temporal pointer features to capture a real-time trace of users' pointer interactions with a graphical user interface. Results show that specific features and hitherto unreported micro-behaviours (patterns of pointer interactions with the content of the user interface) can be used to detect response uncertainty. These results suggest that this approach is promising and warrants further investigation in order to develop a better understanding of the relationship between pointer activity and underlying perceptual, cognitive and affective psychological mechanisms.

## References

Agichtein, E., Brill, E., Dumais, S., Ragno, R., 2006. Learning user interaction models for predicting web search result preferences. In: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'06 (P. Association for Computing Machinery, New York, NY, USA, pp. 3–10.

Ahmed, A.A.E., Traore, I., 2007. A new biometric technology based on mouse dynamics. IEEE Trans. Dependable Secure Comput. 4, 165–179. http://ieeexplore.ieee.org/document/4288179/.

Annett, M., 1970. A classification of hand preference by association analysis. Br. J. Psychol. 61, 303–321.

Arapakis, I., Leiva, L.A., 2016. Predicting user engagement with direct displays using mouse cursor information. In: SIGIR 2016 - Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM Press, New York, New York, USA, pp. 599–608.

Armel, K.C., Beaumel, A., Rangel, A., 2008. Biasing simple choices by manipulating relative visual attention. Judgm. Decis. Mak. 3, 396–403.

Arroyo, E., Selker, T., Wei, W., 2006. Usability tool for analysis of web designs using mouse tracks. In: Conference on Human Factors in Computing Systems - Proceedings. ACM Press, New York, New York, USA, pp. 484–489.

Atterer, R., Wnuk, M., Schmidt, A., 2006. Knowing the user's every move: user activity tracking for website usability evaluation and implicit interaction. In: Proceedings of the 15th International Conference on World Wide Web. ACM Press, New York, New York, USA, pp. 203–212.

Cepeda, C., Rodrigues, J., Dias, M.C., Oliveira, D., Rindlisbacher, D., Cheetham, M., Gamboa, H., 2018. Mouse tracking measures and movement patterns with application for online surveys. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). In: LNCS, vol. 11015, pp. 28–42.

Chan, A., Lau, R.W.H., Si, A., 2001. A motion prediction method for mouse-based navigation. In: Proceedings. Computer Graphics International 2001, pp. 139–146.

Chen, M.C., Anderson, J.R., Sohn, M.H., 2001. What Can a Mouse Cursor Tell Us More? CHI'01 Extended Abstracts on Human Factors in Computing Systems - CHI'01, vol. p. 281. ACM Press, New York, New York, USA.

Choi, B., Pak, A., 2005. A catalog of biases in questionnaires. Prev. Chronic Dis. 2.

Chudá, D., Krátky, P., 2014. Usage of Computer Mouse Characteristics for Identification in Web Browsing. ACM International Conference Proceeding Series, vol. 883. ACM Press, New York, New York, USA, pp. 218–225.

Churchland, M.M., Afshar, A., Shenoy, K.V., 2006. A central source of movement variability. Neuron 52, 1085–1096. http://pmc/articles/PMC1941679/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC1941679/.

Conrad, F.G., Schober, M.F., Coiner, T., 2007. Bringing features of human dialogue to web surveys. Appl. Cogn. Psychol. 21, 165–187.

CrazyEgg, 2018 (n.d.). Retrieved on 28 November 2018, from https://www.crazyegg.com/.

Dhawale, A.K., Smith, M.A., Ölveczky, B.P., 2017. The role of variability in motor learning. Annu. Rev. Neurosci. 40, 479–498. https://pubmed.ncbi.nlm.nih.gov/28489490/.

Dias, M.C., Cepeda, C., Rindlisbacher, D., Battegay, E., Cheetham, M., Gamboa, H., 2019. Predicting response uncertainty in online surveys: a proof of concept. In: BIOSIGNALS 2019 - 12th International Conference on Bio-Inspired Systems and Signal Processing, Proceedings; Part of 12th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2019, Systems and Signal Processing, Proceedings. SCITEPRESS - Science and Technology Publications, pp. 155–162.

Digital Experience Analytics, 2018 (n.d.). Retrieved on 28 November 2018, from https://www.clicktale.com/.

Dillon, A., Watson, C., 1996. User analysis in HCI - the historical lessons from individual differences research. Int. J. Hum.-Comput. Stud. 45, 619–637. https://www.sciencedirect.com/science/article/pii/S1071581996900713.

Dix, A., 2004. Human-Computer Interaction. Pearson/Prentice-Hall.

Faisal, A., Selen, L.P.J., Wolpert, D., 2008. Noise in the nervous system. Nat. Rev. Neurosci. 9, 292–303.

Farney, T.A., 2011. Click analytics: visualizing website use data. Inf. Technol. Libr. 30. https://ejournals.bc.edu/index.php/ital/article/view/1771.

Farris, P.W., Bendle, N.T., Pfeifer, P.E., Reibstein, D.J., 2010. Marketing Metrics: The Definitive Guide to Measuring Marketing Performance, 2nd ed. Wharton School Publishing.

Findlay, J.M., Gilchrist, I.D., 2008. Active Vision: The Psychology of Looking and Seeing. Oxford University Press.

Freeman, J.B., Dale, R., Farmer, T.A., 2011. Hand in motion reveals mind in motion. http://journal.frontiersin.org/article/10.3389/fpsyg.2011.00059/abstract.

Gamboa, H., Fred, A., 2004. A behavioral biometric system based on human-computer interaction. In: Jain, A.K., Ratha, N.K. (Eds.), Biometric Technology for Human Identification. In: International Society for Optics and Photonics, vol. 5404, pp. 381–392. http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=843484.

Gamboa, H.F.S., 2008. Multi-modal behavioral biometrics based on HCI and electrophysiology. Ph.D. thesis. Instituto Superior Tecnico. http://www.lx.it.pt/~afred/pub/thesisHugoGamboa.pdf.

Glaholt, M.G., Wu, M.-C., Reingold, E.M., 2009. Predicting preference from fixations. PsychNology J. 7, 141–158. http://dblp.uni-trier.de/db/journals/psychology/psychnology7.html%23GlaholtWR09.

Goecks, J., Shavlik, J., 2000. Learning users' interests by unobtrusively observing their normal behavior. In: Proceedings of the 5th International Conference on Intelligent User Interfaces IUI'00. Association for Computing Machinery, New York, NY, USA, pp. 129–132.

Guo, Q., Agichtein, E., 2010. Ready to buy or just browsing? detecting web searcher goals from interaction data. In: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval SIGIR'10. Association for Computing Machinery, New York, NY, USA, pp. 130–137.

Hibbeln, M., Jenkins, J.L., Schneider, C., Valacich, J.S., Weinmann, M., 2017. How is your user feeling? Inferring emotion through human-computer interaction devices. MIS Q. 41, 1–21.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J., 2011. Eye Tracking: A Comprehensive Guide to Methods and Measures. Oxford University Press.

Horwitz, R., Kreuter, F., Conrad, F., 2017. Using mouse movements to predict web survey response difficulty. Soc. Sci. Comput. Rev. 35, 388–405.

Hotjar, 2018 (n.d.). Retrieved on 28 November 2018, from https://www.hotjar.com/.

Huang, J., White, R.W., Dumais, S., 2011. No clicks, no problem: using cursor movements to understand and improve search. In: Conference on Human Factors in Computing Systems - Proceedings, pp. 1225–1234.

Inspectlet, 2018 (n.d.). Retrieved on 28 November 2018, from https://www.inspectlet.com/.

Joachims, T., Granka, L., Pan, B., Hembrooke, H., Gay, G., 2017. Accurately interpreting clickthrough data as implicit feedback. SIGIR Forum 51, 4–11.

Jones, K.E., Hamilton, A.F., Wolpert, D.M., 2002. Sources of signal-dependent noise during isometric force production. J. Neurophysiol. 88, 1533–1544. PMID: 12205173.

Just, M., Carpenter, P., 1980. A theory of reading: from eye fixations to comprehension. Psychol. Rev. 87 (4), 329–354.

Langley, P., Sage, S., 1997. Scaling to Domains with Many Irrelevant Features. Computational Learning Theory and Natural Learning Systems, vol. 4. Cambridge MA MIT Press: R. Greiner.

Lim, S.-L., O'Doherty, J.P., Rangel, A., 2011. The decision value computations in the vmpfc and striatum use a relative value code that is guided by visual attention. J. Neurosci. 31, 13214–13223. https://www.jneurosci.org/content/31/37/13214. https://www.jneurosci.org/content/31/37/13214.full.pdf.

LimeSurvey, 2018 (n.d.). Retrieved on 28 November 2018, from https://www.limesurvey.org/.

Mattmann, M., Logar, I., Brouwer, R., 2019. Choice certainty, consistency, and monotonicity in discrete choice experiments. J. Environ. Econ. Policy 8, 109–127.

Park, H.A., 2013. An introduction to logistic regression: from basic concepts to interpretation with particular attention to nursing domain. J. Korean Acad. Nursing 43, 154–164. http://www.ncbi.nlm.nih.gov/pubmed/23703593.

Pimenta, A., Carneiro, D., Novais, P., Neves, J., 2013. Monitoring mental fatigue through the analysis of keyboard and mouse interaction patterns. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). In: LNAI, vol. 8073, pp. 222–231.

Pocius, K.E., 1991. Personality factors in human-computer interaction: a review of the literature. Comput. Hum. Behav. 7, 103–135. https://www.sciencedirect.com/science/article/pii/074756329190002I.

Pusara, M., Brodley, C.E., 2004. User re-authentication via mouse movements. In: VizSEC/DMSEC'04: Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security. ACM Press, New York, New York, USA, pp. 1–8.

Rayner, K., 1998. Eye movements in reading and information processing: 20 years of research. Psychol. Bull. 124, 372–422. http://www.ncbi.nlm.nih.gov/pubmed/9849112.

Revett, K., Jahankhani, H., de Magalhães, S.T., Santos, H.M.D., 2008. A survey of user authentication based on mouse dynamics. In: Global e-Security: 4th International Conference, ICGeS 2008, London, UK, June 23-25, 2008, Proceedings. Springer, Berlin, Heidelberg, pp. 210–219.

Rodden, K., Fu, X., 2007. Exploring how mouse movements relate to eye movements on web search results pages. In: Special Interest Group on Information Retrieval Workshop on Web Information Seeking and Interaction, pp. 29–32. http://www.rodden.org/kerry/publications/wisi07/. http://www.australianscience.com.au/research/google/32735.pdf%23page=33.

Rohilla Shalizi, C., 2019. Advanced Data Analysis from an Elementary Point of View. Technical Report Cambridge University Press. https://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/ADAfaEPoV.pdf.

Schindler, M., Lilienthal, A.J., 2019. Domain-specific interpretation of eye tracking data: towards a refined use of the eye-mind hypothesis for the field of geometry. Educ. Stud. Math. 101, 123–139.

Schneider, I.K., van Harreveld, F., Rotteveel, M., Topolinski, S., van der Pligt, J., Schwarz, N., Koole, S.L., 2015. The path of ambivalence: tracing the pull of opposing evaluations using mouse trajectories. Front. Psychol. 6, 996. http://www.ncbi.nlm.nih.gov/pubmed/26236267. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4505081.

Seelye, A., Hagler, S., Mattek, N., Howieson, D.B., Wild, K., Dodge, H.H., Kaye, J.A., 2015. Computer mouse movement patterns: a potential marker of mild cognitive impairment. Alzheimers Dement. 1, 472–480. https://www.sciencedirect.com/science/article/pii/S2352872915000792. http://linkinghub.elsevier.com/retrieve/pii/S2352872915000792.

Shalabi, L.A., Shaaban, Z., Kasasbeh, B., 2006. Data mining: a preprocessing engine. J. Comput. Sci. 2, 735–739. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.121.2072%26rep=rep1%26type=pdf.

Sperandei, S., 2014. Understanding logistic regression analysis. Biochem. Med. 24, 12–18. http://www.ncbi.nlm.nih.gov/pubmed/24627710. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3936971.

Stillman, P.E., Shen, X., Ferguson, M.J., 2018. How mouse-tracking can advance social cognitive theory. Trends Cogn. Sci. 22, 531–543. https://pubmed.ncbi.nlm.nih.gov/29731415/.

Sun, Y., Kamel, M.S., Wong, A.K., Wang, Y., 2007. Cost-sensitive boosting for classification of imbalanced data. Pattern Recognit. 40, 3358–3378. https://www.sciencedirect.com/science/article/abs/pii/S0031320307001835.

Tan, P.K., Downey, T.J., Spitznagel, E.L., Xu, P., Fu, D., Dimitrov, D.S., Lempicki, R.A., Raaka, B.M., Cam, M.C., 2003. Evaluation of gene expression measurements from commercial microarray platforms. Nucleic Acids Res. 31, 5676–5684. http://www.ncbi.nlm.nih.gov/pubmed/14500831. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC206463. https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkg763.

Tzafilkou, K., Protogeros, N., 2018. Mouse behavioral patterns and keystroke dynamics in end-user development: what can they tell us about users' behavioral attributes?

Comput. Hum. Behav. 83, 288–305. https://www.sciencedirect.com/science/article/pii/S0747563218300700.

Tzafilkou, K., Protogeros, N., Yakinthos, C., 2014. Mouse tracking for web marketing: enhancing user experience in web application software by measuring self-efficacy and hesitation levels. Int. J. Strategic Innovat. Marketing.

Witten, I.H.I.H., Frank, E., 2005. Data Mining: Practical Machine Learning Tools and Techniques. Morgan, Kaufman.

Yamauchi, T., Xiao, K., 2018. Reading emotion from mouse cursor motions: affective computing approach. Cogn. Sci. 42, 771–819.

Zimmermann, P., Guttormsen, S., Danuser, B., Gomez, P., 2003. Affective computing—a rationale for measuring mood with mouse and keyboard. Int. J. Occup. Saf. Ergon. 9, 539–551.

Zushi, M., Miyazaki, Y., Norizuki, K., 2012. Web application for recording learners' mouse trajectories and retrieving their study logs for data analysis. Knowledge Managm. E-Learn. 4, 37–50. https://www.researchgate.net/publication/268064257_Web_application_for_recording_learners'_mouse_trajectories_and_retrieving_their_study_logs_for_data_analysis/fulltext/546601540cf2052b50a14b59/268064257_Web_application_for_recording_learners'_mouse_traje.