Is mouse dynamics information credible for user behavior research? An empirical investigation

Article *in* Computer Standards & Interfaces · March 2024

DOI: 10.1016/j.csi.2024.103849

Eduard Kuric
Slovak University of Technology in Bratislava

Computer Standards & Interfaces 90 (2024) 103849

Contents lists available at ScienceDirect

Computer Standards & Interfaces

journal homepage: www.elsevier.com/locate/csi

Is mouse dynamics information credible for user behavior research? An empirical investigation

Eduard Kuric [a,b,*], Peter Demcak [b], Matus Krajcovic [b], Peter Nemcek [a,b]

[a] Faculty of Informatics and Information Technologies, Slovak University of Technology, Ilkovicova 2, Bratislava, 84216, Slovakia
[b] UXtweak Research, Vysoka 26, Bratislava, 81106, Slovakia

ARTICLE INFO

Dataset link: github.com/micemicsresearch/mouse-dynamics-data-credibility

Keywords:
Mouse configuration
Measurement
Data validity
User behavior modeling
Biological sex classification
Machine learning

ABSTRACT

Mouse dynamics, information on user's interaction with a computer mouse, are in vogue in machine learning for purposes such as recommendations, personalization, prediction of user characteristics and behavioral biometrics. We point out a blind spot in current works involving mouse dynamics that originates in underestimating the gravity of the characteristics of the mouse device and configuration on the data that mouse dynamics are inferred from. In a controlled study with $N = 32$ participants, across three kinds of mouse interaction activities, we collect data for mouse dynamics utilizing a variety of mouse parameter configurations. We show that mouse dynamics commonly used in studies can be significantly altered by differences in mouse parameters. Out of 108 evaluated mouse dynamics metrics, 95 and 84 are affected between two conducted studies. A machine learning model's performance can be warped by the mouse parameters being used. We demonstrate on a prediction task that mouse parameters cannot be approached uniformly and without consideration. We discuss methodological implications — how mouse dynamics studies should account for the diversity of mouse-related conditions.

1. Introduction

A significant portion of human lives in the 21st century takes place online, with productive-age people spending 6 h 37 min of their day online.[1] Among other activities, users engage with the Web to read news, study, enjoy entertainment or pay their bills. Each user comes into this digital environment with a different set of knowledge, background, skills, experiences, literacy (computer or domain-specific), needs, interests and intents (long-term and short-term). Today's apps and websites are often intelligent systems — capable of personalizing user interfaces, making recommendations or otherwise adapting to the specifics of a particular user.

When inferring information about users on the Web, there are a multitude of sources of data that current approaches make use of. Among them is the data that traces movements of a cursor controlled by a user with a mouse device. A type of information that can be extracted from these are the user's mouse dynamics. Though the usage of mobile touch devices (where there can be no mouse dynamics to speak of) is booming worldwide, desktop devices continue to comprise a sizeable portion of the web traffic market share.[2] The persistence of desktop computers can be attributed to the higher degree of control that they grant the end user. Such fine control may be unnecessary for casual browsing, but remains preferable for tasks of greater weight such as office work, shopping, banking or trading. The pursuit of studying mouse dynamics can be expected to maintain its relevance into the future.

Mouse dynamics are features of the user. They characterize how the user operates a computer mouse — what shape are the trajectories of their mouse movements, how they click, how accurate they are, etc. Since each user of a computer mouse has their unique upper limb movements, the mouse dynamics differ from user to user. Currently, mouse dynamics are widely inferred for user modeling with a variety of purposes, from prediction of user intent [1–3], user experience [4,5], the user's characteristics such as gender or age [6–9], for authentication [10–13] or for personalized recommendations [14,15].

One of the known problems of mouse tracking, according to Schoemann et al. [16], is that a variety of study design features can affect mouse tracking data. This includes the mouse sampling rate, cursor speed and training (warm-up) activities. According to Kieslich et al. [17], there are no guidelines available for how mouse devices should be configured for experiments that involve mouse tracking. Only 5% of

* Corresponding author at: Faculty of Informatics and Information Technologies, Slovak University of Technology, Ilkovicova 2, Bratislava, 84216, Slovakia.
E-mail address: eduard.kuric@stuba.sk (E. Kuric).

[1] DATAREPORTAL 2022 October Global Statshot Report: https://datareportal.com/reports/digital-2022-october-global-statshot.
[2] Statcounter desktop vs. mobile vs. tablet market share statistics: https://gs.statcounter.com/platform-market-share/desktop-mobile-tablet.

https://doi.org/10.1016/j.csi.2024.103849
Received 9 November 2023; Received in revised form 10 February 2024; Accepted 27 February 2024
Available online 4 March 2024

current mouse tracking studies state all the design features described by Schoemann et al. [16], whether explicitly or indirectly. Cursor speed and mouse sampling rates are among the least reported study design features.

Failure to transparently report information about mouse configuration may present an obstacle for the experiment's reproducibility, and in some cases, raise concerns about validity. Certain approaches in research methodology (within-subject design, data normalization) may alleviate concerns by mitigating the impact of mouse configuration. It is not the intent of this paper to directly invalidate any previous research. Rather, better understanding of the impacts of mouse configuration may lead to better understanding of existing knowledge, improvements in methodology and pursuit of ecological validity. Works that pay less attention addressing variability in mouse configurations can be reasonably expected to be impacted more severely.

To the best of our knowledge, there is currently no research dedicated specifically to investigating the effects that the selection of a mouse device and its settings has on the properties of mouse tracking data and mouse dynamics features. The lack of a firm standard presents an issue for designing mouse tracking experiments, where the results are inherently linked to the parameters of the mouse configuration used, no matter whether research design is controlled or uncontrolled, in what domain and for what purpose. The question of whether results are actually dependent or independent on the mouse parameters is itself unknown. Some degree of relation has been observed, with Freeman and Ambady [18] finding that high cursor speed (sensitivity) can alter mouse trajectory data, while Grage et al. [19] tie increased cursor speed to an adaptation of the hand/cursor movement ratio, accompanied by increased frequency of sudden high-speed "changes of mind". Works that perform normalization of mouse dynamics to account for mouse configuration variability are scarce, mostly focused on personal user differences or noise reduction within the same user. Investigation of whether normalization is also needed to account for mouse configuration differences is lacking.

This leads us to the question of whether – and to what extent – mouse dynamics are dependent on the properties of the mouse setup on a computer (the mouse device itself, along with its configuration). Does diversity of mouse configurations plausible in real-world conditions cause differences in raw mouse interaction data? What does a change on mouse parameters mean for the validity of the results from a study investigating mouse dynamics, or for the reported performance of a method (i.e., machine learning) that has mouse dynamics as its inputs? If users are given a mouse configuration that they are unfamiliar with, do they adapt to it in a way that significantly affects their behavior? For more methodically sound research of users' mouse behavior, we see demand for answering these questions.

The primary contributions of this paper are:

- Statistical evaluation of the effect of mouse devices and their configurations on mouse dynamics metrics. In a comprehensive list of metrics utilized in research, most are shown to be significantly altered by the mouse parameters (95/108 in one study, 84/108 in a second study).
- Using classification of the user's biological sex as an example of a method involving mouse dynamics, the comparison of performances between samples collected with different mouse configurations demonstrates that performance is worse when the user's mouse configuration is not included in the training set.
- Discussion of methodological considerations accounting for mouse parameter configuration in research involving mouse dynamics, given its potential impact on representations of user mouse behavior.
- A three-stage experiment design for obtaining a dataset that allows for comparing mouse dynamics between different mouse configurations.

2. Background and literature review

2.1. Computer mouse parameters

A number of mouse configuration parameters affect the user's experience with a mouse and the properties of mouse inputs harnessed from user–computer interaction. Some of these parameters are the physical attributes of the mouse itself (e.g., the resolution of the sensor), while others are software settings (e.g., the mouse sensitivity setting in the operating system). Below is a (nonexhaustive) list of mouse parameters that can be conceived as impactful on the reliability of mouse tracking data[3]:

Resolution. Measured in – and commonly only referred to as – Counts Per Inch (CPI), resolution of a mouse is the number of distinct steps that the sensor can divide the distance of one inch (2.54 cm) into, so that it can distinguish between individual steps. Some mouse manufacturers use Dots Per Inch (DPI) as a synonym for CPI. However, DPI as a term has origins in printers and is inaccurate when describing mouse devices. DPI is primarily used as a marketing term, for customers who are familiar with it, commonly gamers. In gaming, lower CPI is preferable for tasks requiring higher precision, while higher CPI is recommended when interacting with distant targets to reduce the size of necessary motions, even if at the cost of precision [20].

Sensitivity. Operating systems provide sensitivity as a setting that acts as a multiplier of the CPI, software-wise adjusting the ratio of mouse movement to movement on the screen. Microsoft Windows provides 20 sensitivity options from 1 to 20.[4] Games and other software applications may offer their own independent sensitivity setting.

Effective CPI. eCPI (commonly referred to as eDPI, see Resolution) is calculated as CPI multiplied by sensitivity, acting as a measure of the real sensitivity collectively for a mouse device and software configuration setup.

Tracking speed. There is a limit to the movement speed that a mouse may reliably detect. The unit used to measure tracking speed of a mouse is Inch Per Second (IPS). The IPS is determined by the frame rate of the sensor — the frequency at which the sensor captures a frame (image) of the surface. To accurately determine mouse movements of a specific speed, the device needs to be able to calculate distance traveled by calculating the position of each frame in reference to the previous one.

Polling rate. The rate (measured in Hz) that signifies how many times per second a USB mouse reports its position to the computer (commonly 125–1000 Hz, but can go as high as 8000 Hz). This is separate from the frame rate of the sensor, which is commonly higher than the polling rate, meaning reported position may be calculated from multiple frames. Polling rate is responsible for the perceived responsiveness of the mouse.

Acceleration. Operating under the premise that faster mouse movements indicate when the user is aiming to travel a longer distance with the cursor, mouse acceleration is a function that alters mouse sensitivity depending on the speed of movements. Effectively, faster movements of the mouse cause the cursor to travel a longer distance than slower more careful movements of the mouse device traveling the same distance. When found as a software feature in an operating system, it may be enabled/disabled. The primary use case for mouse acceleration is standard desktop computer use, where it allows for moving the cursor across the full screen without lifting the mouse. Due to acceleration's interference with developing muscle memory for precision mouse aiming, some users prefer to disable it while gaming.[5]

[3] Mouse buyer's guide of mouse physical attributes: https://sensor.fyi/info/.

[4] Liquipedia mouse settings guide: https://liquipedia.net/counterstrike/Mouse_Settings#Windows_Sensitivity.

[5] Guide to mouse settings: https://www.techspot.com/article/2556-master-mouse-settings/.

The screen resolution, while not a literal parameter of the mouse but rather the screen, can also directly impact the mouse dynamics due to the size of the overall space available for a cursor to move in Freeman and Ambady [18]—how many pixels a mouse movement may cover in its path at maximum and what relative percentual distance on the screen the same movement will cover at the same eCPI. Other examples of circumstances within the environment that may impact mouse behavior in real-world conditions include the surface beneath the mouse (friction and optical properties), the ergonomy and the weight of the mouse.

2.2. Mouse dynamics features

To model mouse movement behavior of users during specific tasks, metrics or other representations are calculated from raw mouse log data obtained by mouse tracking software, such as UXtweak[6] and MouseTracker.[7] Typically, metrics that serve as input into machine learning algorithms are inferred from coordinates along the x and y axis, logged as discrete points on a timeline demarcated by their corresponding timestamps [2–4,8,9,11,13,15,18,21–27]. Aside from calculating metrics, mouse coordinate data can be clustered [28] or passed as an input to neural networks directly [29–32].

According to Balen et al. [9], metrics that represent mouse dynamics can be grouped into three categories: temporal, spatial and accuracy metrics. Temporal metrics are bound to cursor velocity, ballistic component (phase of cursor acceleration and deceleration during the initial cursor movement), reaction time, movement correction and button press times. Spatial metrics include mouse trajectory lengths, direction changes and movement variability — all related to the distances between consecutive points. Movement accuracy metrics include metrics related to errors while clicking on the action stimulus.

Across literature, most common features used to describe user behavior on the web are cursor velocity, acceleration, duration, distance, straightness of the cursor path and deviation from the ideal path. A compilation of mouse dynamics metrics described in literature can be found in Table 1. This includes elementary metrics, as well as aggregation variations – e.g., velocity mean, maximum, minimum – and logical variants — e.g., 2-dimensional or separately along the x and y axis. See Table 2 for a summary of works where the individual metrics were deployed.

2.3. User modeling and behavioral research

User modeling is a technique for capturing information about the user [36]. It can support personalization of user interfaces and content (e.g., rearrangement of navigation, display of messaging respective to knowledge) or recommendations, making it fruitful in many domains, including education apps, ecommerce, online libraries and encyclopedias, virtual assistants, e-government sites and others. The information used to build (and update) a user model may be gathered either from explicit feedback (information submitted directly by users, commonly via a form) or implicit feedback (observation of users). Implicit and explicit feedback can also be combined (e.g., recommendations based on the user's favorite product categories and search history).

Data sources employed by user modeling can be limited in the information that they can provide. Server logs are one such common source of data. Some aspects of the user's path on the web (e.g., scanning over the page with a mouse cursor) cannot be traced in high-level server logs. Some research refers to pageviews data as "click events" or "click streams" [37–39], even though they do not provide data on actual clicks (e.g., which element was clicked). Higher granularity of data can capture the full interaction (e.g., cursor movements, scrolling, inputs). Mouse dynamics modeled from mouse tracking data have been shown to boost performance in a behavioral prediction task [1].

A variety of published works conduct research utilizing mouse interaction data. Their topics commonly pertain to machine learning classification and extraction of features for the formation of prediction models. These works commonly do not pay attention to mouse parameter configurations, e.g., not listing utilized CPI or mouse sensitivity in experiment specification, or conducting a controlled experiment using a singular mouse parameter configuration while ignoring the diversity of mouse configurations present in data from real users. There exists a valid concern about the methods and techniques covered.

If mouse dynamics are dependent on mouse parameter configuration (as investigated in this paper), the implications on research would be multi-fold, whether experiments are conducted in controlled or uncontrolled environments. Result reproducibility and transparency can be cited as primary concerns. Validity of research outcomes documented in works may potentially be compromised or be interpreted incorrectly due to the lack of full context (e.g., whether a new method is an actual improvement over the state of the art can become unclear). The role of mouse parameter configuration may become a relevant question (e.g., the robustness of prediction models to changes in mouse configuration). In case of between-subjects research design (where comparisons are performed between different participants, comprising the majority of reviewed research employing mouse dynamics), hidden variables of mouse configuration may render comparison of independent groups invalid.

Within-subject research can be considered as resistant to concerns about its validity or reproducibility due to a neglect of mouse configuration. Here, the typical premise is that conditions where participant behavior is observed and compared are consistent as they are obtained from the same participant. As such, if mouse configurations exert a meaningful effect on mouse dynamics, within-subject research can benefit from more holistic addressing of mouse configuration to further enhance the robustness and rigorousness of its methodology. For illustration, Cai et al. [13] refer to the use of multiple workstations over multiple sessions, during an experiment involving what is described succinctly as an HP optical mouse. An explicit assurance can be preferred over an implicit assumption, that all these mouse devices are identical over all sessions, as is the mouse sensitivity setting in the operating system (i.e., that participants were not allowed to change it to their preference).

Applicability is another potential concern. Let us consider an ecommerce website, where a prediction model classifies anonymous visitors into age categories according to their mouse dynamics. A recommendation algorithm then takes the age category as a parameter for recommending products to the user. When trained on data gathered in laboratory settings where users all use the same mouse, mouse movement speed may be found as an effective feature for predicting age. However, it is questionable whether this would still be true in real conditions, where parameters such as mouse CPI, tracking speed, material properties of the surface beneath the mouse or the size of the screen are unknown variables.

The lack of proper examination of mouse parameter configuration may be impactful in a diversity of research domains. Despite their contributions, which we do not aim to discount, works listed below illustrate various degrees of cases where caring to explicitly incorporate the aspect of mouse configuration into methodology, or accounting for it further, may be beneficial. See also Table 8 for a structured overview of research designs employed and approaches to mouse configuration taken.

Recommendations and personalization are two very common aspects of intelligent information systems, intended for serving users information relevant to them and in a personally-tailored fashion. Social

[6] UXtweak platform for user experience research, containing tools that enable remote mouse tracking via installed snippet or web browser extension: https://www.uxtweak.com/.

[7] MouseTracker software package for mouse movement research, capable of recording mouse data on a computer locally: http://www.mousetracker.org/.

Table 1
Overview of mouse-movement-inferred features (mouse dynamics) across literature. Conceptually related metrics that appear with varied mutations are grouped together, aggregates and variants listed.

Feature	Description	Aggregates	Variants
Temporal features			
Velocity	Change in distance over time	Min, max, mean, sd, min–max difference	x-axis, y-axis, smoothed
Acceleration	Change in velocity over time	Min, max, mean, sd, min–max difference	x-axis, y-axis, positive, negative
Jerk	Change in acceleration over time	Mean, sd	Positive, negative
Snap	Change in jerk over time	Mean, sd	Positive, negative
Angular velocity	Change of angle over time	Mean, sd	–
Movement duration	Time between clicks	–	Total
Reaction time	Time until response	–	–
Initiation time	Time until mouse movement	–	–
Pause	Idle cursor time	Count, sum	–
Spatial features			
Movement distance	Path length	–	Smoothed, total
Straightness	Path length ratio to ideal path	Mean, sd	–
Path deviation	Perpendicular distance to ideal path	Max	AUC
Jitter	Smoothed to real path length ratio	–	–
Angle	Current movement direction	–	–
Flips	Movement directional changes	Count	x-axis, y-axis
Path crossings	Intersections of path and ideal path	Count	–
Curvature	Change of angle over distance	Mean, sd	–
Inflection points	Curvature changes, flex points	Count	–
Stimuli interaction features			
Clicks	Number of stimuli presses	Count	–
Hold time	Stimuli press duration	–	–
Click error	Stimuli center and press distance	–	x-axis, y-axis, absolute
Time to click	Stimuli hover to press time	–	–
Scroll	Page scrolling	–	Horizontal, vertical

Table 2
Overview of mouse-movement-inferred features (mouse dynamics) with recent sources. A multitude of works utilizes varied sets of mouse dynamics.

Feature	References
Temporal features	
Velocity	Monaro et al. [2], Guo and Agichtein [3], Fernández-Fontelo et al. [4], Pentel [7], Balen et al. [9], Antal and Egyed-Zsigmond [11], Cai et al. [13], Hucko et al. [15], Freeman and Ambady [18], Kratky and Chuda [21], Freihaut and Göritz [22], Pepa et al. [23], Maldonado et al. [25], Zheng et al. [27], Shen et al. [33]
Acceleration	Guo and Agichtein [3], Fernández-Fontelo et al. [4], Antal and Egyed-Zsigmond [11], Hucko et al. [15], Freeman and Ambady [18], Kratky and Chuda [21], Freihaut and Göritz [22], Pepa et al. [23], Maldonado et al. [25], Shen et al. [33]
Jerk	Antal and Egyed-Zsigmond [11], Hucko et al. [15], Freihaut and Göritz [22]
Snap	Freihaut and Göritz [22]
Angular velocity	Kratky and Chuda [21]
Movement duration	Balen et al. [9], Antal and Egyed-Zsigmond [11], Hucko et al. [15], Kratky and Chuda [21], Freihaut and Göritz [22], Shen et al. [26], Seelye et al. [34]
Reaction time	Monaro et al. [2], Fernández-Fontelo et al. [4], Balen et al. [9], Freeman and Ambady [18]
Initiation time	Guo and Agichtein [3], Fernández-Fontelo et al. [4], Antal and Egyed-Zsigmond [11], Freeman and Ambady [18]
Pause	Cai et al. [13], Kratky and Chuda [21], Pepa et al. [23], Seelye et al. [34]
Spatial features	
Movement distance	Fernández-Fontelo et al. [4], Gardey et al. [5], Balen et al. [9], Antal and Egyed-Zsigmond [11], Cai et al. [13], Sulikowski et al. [14], Hucko et al. [15], Kratky and Chuda [21], Freihaut and Göritz [22], Seelye et al. [34]
Straightness	Pentel [7], Balen et al. [9], Antal and Egyed-Zsigmond [11], Freeman and Ambady [18], Kratky and Chuda [21], Freihaut and Göritz [22]
Path deviation	Yamauchi et al. [8], Antal and Egyed-Zsigmond [11], Freeman and Ambady [18], Kratky and Chuda [21], Freihaut and Göritz [22], Shen et al. [26]
Jitter	Kratky and Chuda [21]
Angle	Pentel [7], Yamauchi et al. [8], Antal and Egyed-Zsigmond [11], Freihaut and Göritz [22], Zheng et al. [27]
Flips	Pentel [7], Balen et al. [9], Kratky and Chuda [21], Freihaut and Göritz [22]
Path crossings	Balen et al. [9]
Curvature	Antal and Egyed-Zsigmond [11], Kratky and Chuda [21], Seelye et al. [34]
Inflection points	Kratky and Chuda [21]
Stimuli interaction features	
Clicks	Kuric et al. [1], Cai et al. [13], Kratky and Chuda [21], Pepa et al. [23], Khan et al. [35]
Hold time	Balen et al. [9], Cai et al. [13], Pepa et al. [23], Shen et al. [26]
Click error	Balen et al. [9]
Time to click	Balen et al. [9], Zheng et al. [27]
Scroll	Sulikowski et al. [14]

media and video sharing platforms utilize recommendation algorithms to suggest posts and new content creators in the feed. Online stores recommend products of potential interest. Personalization is intelligent adaptation of a system to its user, e.g., search results ordered depending on the user's profile. Accommodating complex search tasks attempted by users is a multidimensional challenge, where mouse movements serve as common predictors of task dimensionality [40]. Sulikowski et al. [14] use fuzzy modeling on purchase intent regarding recommendations. A dataset containing mouse tracking data within the Document Object Model (DOM) of 5 ecommerce sites was recorded via a browser extension. No reference to mouse configuration is made. Hucko et al. [15] present a personalized onboarding tool where assistive messages are displayed when confusion is detected from mouse dynamics. The precision of achieved prediction is 63%. Whether the mouse parameter configurations are identical across all 20 working stations in the lab, or what are the exact configurations, is unknown.

User experience is the overall cognitive and emotional effect that a product, service or company has on the user who interacts with it [41]. Gardey et al. [5] evaluate UX of particular elements in web forms, namely by predicting their interaction effort. De Santana et al. [42] predict interaction with explainable AI elements using high-granularity interaction data (including mouse interaction data) in a remote uncontrolled study. A LSTM proposed by Chen et al. [43] predicts search engine satisfaction based on relationships of movements between regions. In either work, mouse configuration is not referenced as a subject of consideration.

Affective computing is the interdisciplinary field of study dedicated to detection, interpretation and simulation of affective states that are tied to human emotions. Pepa et al. [23] detect the stress of users via classification, utilizing 22 mouse dynamics and 15 keyboard features. The experiment involves 4 tasks (2 difficulty variants each), chosen for their ability to invoke stress. Participants used their own computer and mouse during the experiment to simulate real conditions. However, no attention is given to mouse configuration on the participants' end. Khan et al. [35] point to the possibility of predicting personality traits of users based on mouse and keyboard events. They utilize data from two studies — one collecting logs of real user activity over a period of several days, the second involving participants engaged in programming tasks while listening to mood-evoking music. The authors make no mention of what mouse devices were used or how they were set up.

Prediction of user characteristics such as sex or age can provide profiling/statistical data about anonymous website visitors to help product, research or marketing teams better understand their existing audiences. Predicting characteristics also has direct applications for making systems react appropriately to actions made by users. User behavior may give away information about the user's innate characteristics, without the need of any explicit action on the user's part. An exemplary use case is in preventing minor children from accessing age-restricted content. Pentel [7] extracts features from standard keyboard and mouse use to classify the user's age and sex. Six datasets are used — 2 from real system use, 4 harnessed in experiments for other mouse tracking studies. No mentions account for mouse selection or configuration.

Fernandez-Lanvin et al. [6] demonstrate that men and women, as well as different age groups, can be differentiated by mouse movements while performing simple mouse actions such as pointing and clicking, dragging and dropping, and selecting items. Yamauchi et al. [8] discovered significant differences between mouse dynamics of men and women in an experiment with a simple decision task. The six work stations used for the experiment utilize the same optical mouse model and have mouse sensitivity set to medium. Authors admit the lack of validation on different input devices as a limitation. Kratky and Chuda [21] utilize mouse movement coordinate data from a real website to calculate metrics from individual distances, times and angles. These features are used as inputs to classification algorithms predicting the age and gender of participants, the resulting F-score being 60%. Used mouse configurations are unknown and presumably varied since data was obtained from real webpage users. Balen et al. [9] also classified the gender of 94 participants solely from mouse dynamics features with the accuracy of 76%. The sole mention of mouse configuration is that the used mouse's approximate polling rate is 100 Hz.

Accessibility is the domain of human–computer interaction concerned with making user experiences more easily approachable by more people, particularly persons faced with challenges in everyday life due to a physical or a mental disability. Seeking to make a contribution in the creation of user interfaces for users with cerebral palsy, Almanji et al. [44] assess how mouse dynamics in the categories of rapidity and accuracy reflect the user's arm dexterity. The sole reference to mouse parameters used is that all participants interacted with a "standard" mouse and that the polling frequency of mouse movement positions was 100 Hz.

Behavioral biometrics, the pursuit of identifying individuals unobtrusively based on indicators describing their behavior, has thoroughly explored mouse dynamics as a means of differentiating between humans. Several attempts [31,32] were made towards authenticating users in pre-existing datasets with different use contexts where no information on mouse configuration has been provided. Shen et al. [26] authenticate users by recording mouse movements in a short predetermined task while Cai et al. [13] tackle the issue of variability of mouse dynamics during regular usage. In both cases, identical but further unspecified mouse configuration parameters are used, leaving the authentication performance under realistic conditions as an open question. Siddiqui et al. [10] employ mouse dynamics from people playing the game Minecraft as biometrics, with the hardware and software configuration identical for all participants, documented except the characteristics of the mousepad (although default game settings are mentioned, parameters such as mouse sensitivity are not further specified).

Zheng et al. [27] admit potential unreliability of mouse dynamics for authentication across different computers due to a variety of factors, including mouse brand, sensitivity and mousepad. Mouse dynamics have been used to detect impostors [11] or distinguish humans from bots [12]. Antal et al. [45] provide a mouse dynamics dataset collected on participant's own devices with unknown mouse configurations, evaluated using convolutional neural networks. This dataset can be used to train systems for authentication, detection of bots or training human-like mouse trajectory generative models [46]. Shen et al. [33] show that in the domain of mouse dynamics authentication, the variability of mouse dynamics (caused by factors such as usage of a new mouse, GUI settings, usage scenario, emotional and physical state) found in data collected over longer periods of time leads to lower performance of the prediction of identity, but reduction of dimensionality leads to significantly improved results. The authors also found that motor-skill features (e.g., the time elapsed during a single click) are less liable to being affected by variability than schematic features (e.g., the mouse action histogram). Mouse parameters may be considered as precursors of variability of mouse dynamics, which are owed further research.

User intent prediction is a collective term for approaches that seek to automatically recognize what the user aims to do, whether to capitalize on those intents and provide better customer service, or to spot malicious intent and take appropriate preventative measures. User intent has been predicted from mouse dynamics and without attention being paid to mouse parameter configuration [47] in cases such as likelihood of returning as a customer [1] or readiness to make a purchase [3] in ecommerce, or the intent to lie [2] in personality questionnaires.

User attention prediction evaluates how users visually scan user interfaces or which elements capture their focus, a potentially valuable source of feedback [48]. While not a replacement for true gaze tracking, it can serve as a viable substitute in conditions where eye-tracking is unavailable, since eye-tracking technology requires dedicated hardware and as such is challenging to scale. Arapakis and Leiva [30] predict interest in ads in search engine results with mouse dynamics features non-specific to any particular web page design, but no mentions of accounting for heterogeneous mouse configurations are made.

Cognitive assessment is a domain that finds promise in mouse dynamics's ability to reflect human cognitive processes and the mental capabilities of human subjects. Seelye et al. [34] predict mild cognitive impairments in older people, which correspond with less efficient mouse movements with more frequent pauses in day-to-day computer use. Maldonado et al. [25] detect changes in decisions, which manifest as changes in mouse trajectory. Neither comment on the mouse configurations used during their experiments, nor the concern that a different configuration may affect interpretation of mouse data.

Notably, it is precisely for the above flexibility in predicting all kinds of facts about the user that *privacy and ethical concerns* have already been raised about the use of mouse tracking technology. Leiva et al. [29] present a method for decreasing the effectiveness of machine learning profiling based on mouse dynamics by introducing noise into mouse tracking data, thus granting users recourse against potentially invasive and undesirable tracking.

2.4. Normalization of mouse dynamics features

Among works utilizing mouse dynamics, in the context of potential implications of inconsistent mouse configuration, special attention is owed to methods that involve feature normalization. If mouse dynamics are meaningfully impacted by mouse configuration (as this paper aims to validate), normalization may be a methodologically essential step for eliminating mouse configuration variables and utilizing mouse dynamics to achieve more reliable research results. A survey of related works at the time of writing reveals that normalization of mouse dynamics features is rare — see Table 8.

Two normalization approaches are found in research that may mitigate the effects of mouse configuration. Their distinction — whether the normalization baseline originates in statistical norm (average) behavior, or reference observations of specific behavior that is considered normal.

Statistical norm baseline. Wilson et al. [49] predict whether an essay is genuine or plagiarism generated by a large language model tool, based on mouse dynamics collected as the user answers follow-up questions about the essay's contents. Mouse parameter configurations are not controlled during the experiment. Z-score normalization is used within-participant to account for individual differences between participants, calculated from both experimental conditions (essay written by the participant and generated by AI). Cai et al. [13] perform normalization and feature reduction for a user authenticity task, to eliminate noise present in mouse dynamics present even when the same users perform the same actions.

Reference observation baseline. Fernández-Fontelo et al. [4] research the possibility of employing mouse dynamics for prediction of the difficulty of individual multiple-choice questions in a survey, which may enable reactive modifications of the survey for respondents who are struggling. Participants use their own unknown equipment (including a computer mouse) to complete the survey. The authors propose two personalization methods to correct mouse dynamics to the compound of personal differences between participants (e.g., habits) and the hardware they are using. Behavior in survey questions where questions are not manipulated to be more difficult serves as the baseline.

Normalization may have the potential to control mouse parameter configuration, with additional aspects to consider:

- In problems investigated by within-subject research, the requirement of a suitable baseline may pose a challenge for ecological validity of findings outside the lab. For illustration, detection of stress in controlled conditions may be improved by normalization, using mouse dynamics from controlled non-stressful conditions as a baseline. In real-world environments (e.g., e-commerce, banking, e-health, education), interrupting the user's natural activity to artificially induce a peaceful state may not be an option or could induce further stress if perceived as intrusive. Depending on the nature (e.g., stressfulness) of the uncontrolled environment and individual user differences (e.g., anxiousness), a statistical norm is also not guaranteed to present a suitable baseline.
- In between-subject research design, creating a baseline that eliminates the effects of unknown mouse configuration, yet preserves other unknown distinctions between participants may not be viable. This presents a challenge for research problems that do not lend themselves to within-subject research design, such as prediction of user characteristics that are fixed (e.g., sex, permanent health impairment) or impractical to control within-participant (e.g., age). Alternative normalization approaches, such as machine learning prediction of mouse configuration serving as the input of a correction method may be more applicable.
- Current mouse dynamics normalization approaches have been designed and evaluated primarily to correct personal differences between users [4,49] or variability between sessions of the same user [13]. In real-world conditions, mouse configuration should not be conflated with individual differences, since users may swap devices or change mouse configurations. The specifics of the interaction between normalization and variable mouse configurations remain an open question. Further validation can be seen as warranted regarding its effectiveness and potential risks (e.g., feature homogenization).

Normalization of mouse dynamics to account for variability in mouse configuration parameters is a complex yet under-explored issue. This could be attributed to a lack of research investigating the effects of mouse configuration, potentially resulting in underestimation of their significance for mouse dynamics. Aiming to address this, it is not in the intended scope of this work to propose a mouse configuration normalization method for any specific mouse-dynamics-related research problem, but rather to validate pursuit of designing normalization methods that would tackle the challenges illustrated above.

3. Aim of the study

The parameters of a computer mouse directly impact how the mouse behaves when a user interacts with it. Thus, configurations of mouse parameters can impact the resulting mouse dynamics. In this work, we aim to investigate what effects the mouse parameter configurations can have on properties of mouse dynamics and their usefulness in machine learning prediction tasks. Among the parameters of mouse configuration, we choose to focus on CPI and sensitivity (and by extension their product, eCPI), to mirror heterogeneous mouse devices and sensitivity settings on real-world computers belonging to the end users. Relevant research reports these parameters as potentially having the greatest impact [18,19]. The following research questions are postulated:

RQ1*: What effect do different mouse eCPI configurations have on metrics calculated from raw mouse interaction data?* Our goal is to determine whether eCPI – as a defining characteristic of the mouse – affects mouse dynamics metrics and if so, which metrics are affected the most by the differences in mouse configurations. We hypothesize the majority of features will be affected in some manner. Notably, mutations in both temporal and spatial metrics were expected, namely to metrics that model velocity and shape of the cursor trajectory, since higher eCPI causes the cursor to move faster and become more unwieldy, hence difficult to aim accurately.

RQ2*: What amount of mouse activity do users need to adapt to a new mouse, so that differences in mouse dynamics metrics can be safely compared between adapted users?* When participants are given a mouse

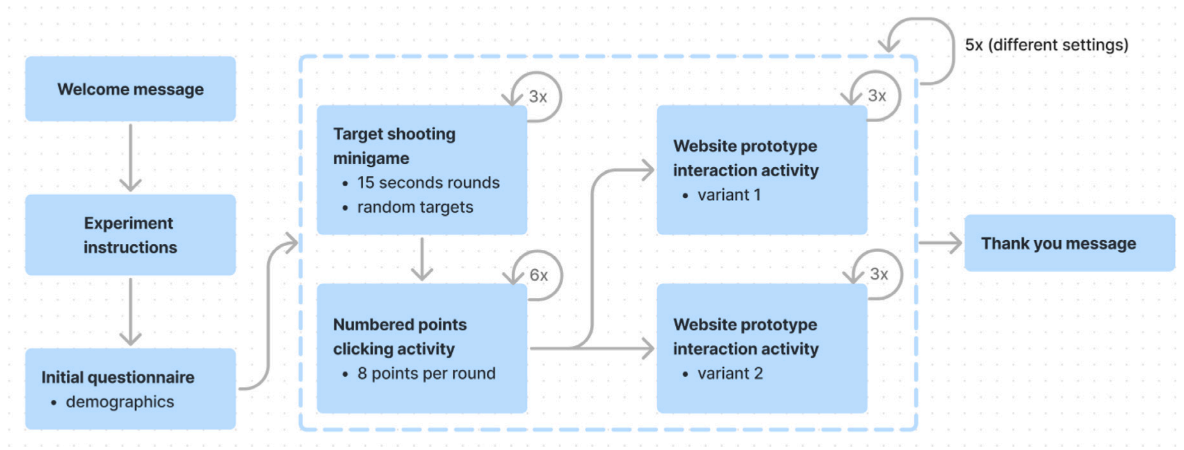


Fig. 1. Diagram of the experiment for acquiring data for mouse dynamics in 3 types of tasks under 5 different mouse parameter configurations from every participant.

they are not used to (with different eCPI making the cursor move slower or faster than a reference device), participants can be expected to take some time to adapt to the new mouse. Adaptation can be reasonably thought to lead to the emergence of changes in mouse dynamics when represented on a learning curve. Differences in mouse dynamics metrics between individual mouse parameter configurations can be caused not only by the congenital traits of the configuration, but also by the user being unaccustomed to the mouse and other settings. It is therefore desirable to understand how users accustom themselves to a new mouse and from what point in the interaction can their mouse dynamics be relevantly compared.

***RQ3**: What is the effect of different mouse eCPI configurations on the performance of prediction using mouse dynamics features?* If there are significant differences in mouse dynamics depending on the mouse parameter configuration, it can be questioned how this affects performance in prediction tasks that adopt mouse dynamics as features. The task of classifying the user's biological sex is chosen to examine whether prediction performance varies when different mouse parameter configuration samples are used for training and testing.

4. Methods

To investigate how mouse configuration can translate into changes in mouse dynamics information, an experiment was conducted to compare varied mouse configurations (see Fig. 1). Obtained information is analyzed statistically and indirectly through performance of machine learning algorithms utilizing them as input.

4.1. Procedure

The experiment consists of three mouse-focused activities (see 4.2 Materials) where the user's mouse movements are tracked. The three activities are completed five times, each iteration being structurally identical, but utilizing a different eCPI configuration. The activity is altered between iterations to assure that participants are moving the cursor naturally, relying on visual inputs instead of their memory to solve tasks. A web application was created for the purpose of guiding participants through the experiment and collecting task and mouse tracking data, based on the UXtweak[8] mouse event data collector software. The changing of mouse parameter configurations (which includes swapping of hardware) is handled by the moderator. Participants complete tasks in either ascending or descending eCPI order, with either of these options selected randomly for every participant, while assuring an even ratio of the two orders.

[8] UXtweak Session Recording tool capable of capturing mouse tracking data on the web: https://www.uxtweak.com/session-recording-tool.

4.2. Materials

The experimental activity repeated by participants with multiple mouse parameter configurations is internally divided into three stages: the target shooting minigame activity, the numbered points clicking activity and the website prototype interaction activity. The purpose of this division of activities is to collect mouse interaction data under different contextual conditions, which are further elaborated on in the specification of individual activities.

4.2.1. Target shooting minigame activity

Each iteration starts with a gamified mouse-intensive activity to let participants get accustomed to a new mouse configuration and to decrease the impact of the previously used mouse configuration on their modus of operating the mouse. Post-game, the participant's mouse dynamics should represent normal mouse behavior, rather than mouse behavior influenced by switching between mouse configurations. The participants' capacity of adapting to the mouse (see RQ2) is investigated during this activity. The secondary purpose of the game is to render the lengthy experiment more enjoyable for the participants.

Participants are asked to shoot targets within a time limit by clicking them to earn a high score (see paper's data repository for full instructions). As an added motivation to perform well in the game, participants can see the top three highest scores achieved by previous players (see paper's data repository for the points calculation algorithm).

In each iteration, the minigame is played in three 15-second rounds. The game is split into multiple rounds to measure improvements in game performance between rounds. The game is also split into rounds for consideration of the participants' ability to achieve a higher degree of focus if partaking in shorter game intervals interspersed with breaks, rather than in a single long and cohesive time interval. A new round is triggered via the explicit click of a button.

During a round, targets are displayed within a 1000x600 pixel (px) area (see Fig. 2). The position of targets is random, with the minimum distance from the previous target being one eighth of the full width on the x axis (125 px) and one eighth of the full height on the y axis (75 px). The random positioning of targets makes the game unlike the other two mouse tracking activities where the intended click positions are predetermined to allow for direct comparison between mouse dynamics of mouse parameter configuration samples. This allows the game to serve as a varied warm-up where participants cannot be construed as conditioned to any specific mouse movement patterns. Due to the differences in the expected mouse movement lengths and directions of individual participants, the data from this minigame is not used for evaluation of mouse dynamics metrics in relation to mouse configurations. As such, the rest of the experiment evaluates mouse

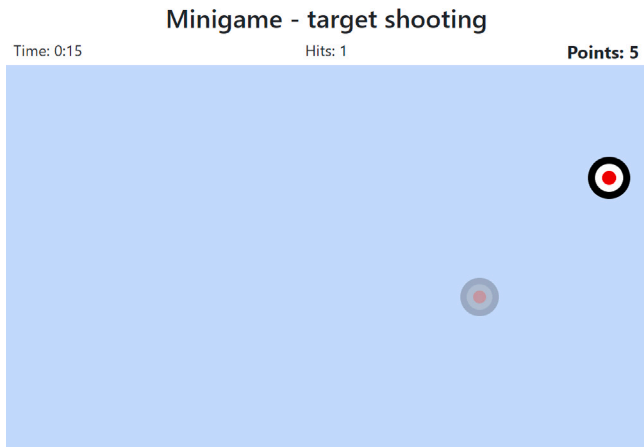


Fig. 2. First of three activities in the experiment is a target shooting minigame task, to help participants get accustomed to new mouse configurations. Targets appear in random locations, earning participants points depending how fast or close to the center they click.

dynamics from users who have had the opportunity to get used to their current mouse parameter configuration.

A new target is displayed during the minigame immediately after the previous target has been clicked. If the target is missed, nothing happens — the participant may retry shooting the same target. The target circle's radius is 30 px, with two smaller concentric circles inside (radiuses 20 px and 10 px), dividing the target into zones. Over the period of 2 s after appearing, the target gradually expands from its center until its radius reaches 1.5 times its original length.

4.2.2. Numbered points clicking activity

The second activity in the experiment has two goals. It is to generate mouse interaction data where:

- There is equal representation of diverse classes of mouse movements from all participants and in all observed mouse parameter configurations.
- The ideal paths are the same for all mouse parameter configurations (or as similar as possible) to support comparisons of mouse dynamics metrics between configurations (see RQ1) and of prediction performance depending on the configurations included in training (see RQ3).

The activity designed for this purpose is inspired by a similar game described by Pentel [24]. Participants are asked to click squares in a grid in the same order as they are numbered in (see paper's data repository for full instructions).

The area where the participants complete the clicking activity (see Fig. 3) has the dimensions of 1000 × 600 px. This area is covered by a 10 × 6 grid (60 tiles in total) that is invisible to the participants. The numbered squares that the participants are asked to click in a specified order are sized 50 × 50 px. Each clickable square occupies one tile in the grid by itself. The square may be located in the center of a tile, or be anchored in any of the tile's corners. This method of positioning is adopted so that the clickable squares themselves do not have the appearance of a grid, having multiple possible positions even within the same column or row.

The activity is split into multiple rounds, with each round starting when the participant clicks the Next button in a popup window. The popup is used to assure that the cursor of all participants starts in approximately the same position. The activity consists of 6 rounds in total, each involving a sequence of 8 numbered rectangles. These numbers can be determined as the smallest number meeting the requirement to cover all types of mouse movements (lengths, directions) intended

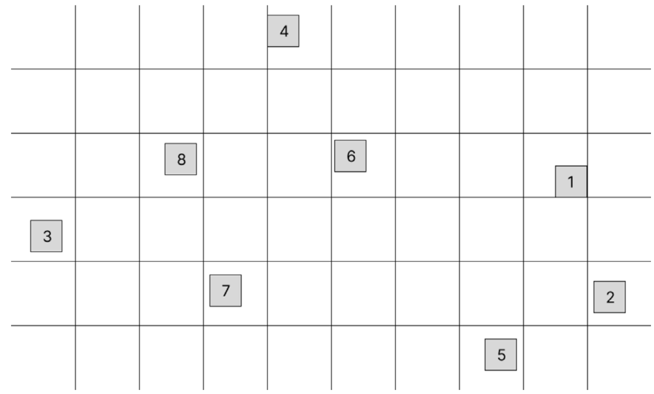


Fig. 3. Grid for generating tile patterns in numbered points clicking activity, portraying a sample tile sequence. The grid is not shown to the participants. The tile that a square is placed in within a sequence is consistent for all mouse parameter configurations.

to track, while obfuscating the repetition of patterns between mouse configurations via a method designed for the experiment.

To classify the types of mouse movements, two key aspects are defined that characterize a mouse movement — its direction and length. The direction of each movement is classified into one of 8 directions — cardinal (north, south, west, east) and intercardinal (northwest, northeast, southwest, southeast). By their length, movements belong into one of three categories — short, medium and long. Combining each direction and length once, there have to be 24 mouse movements. This can be structured as 3 rounds, each involving a sequence of 8 numbered clickable rectangles.

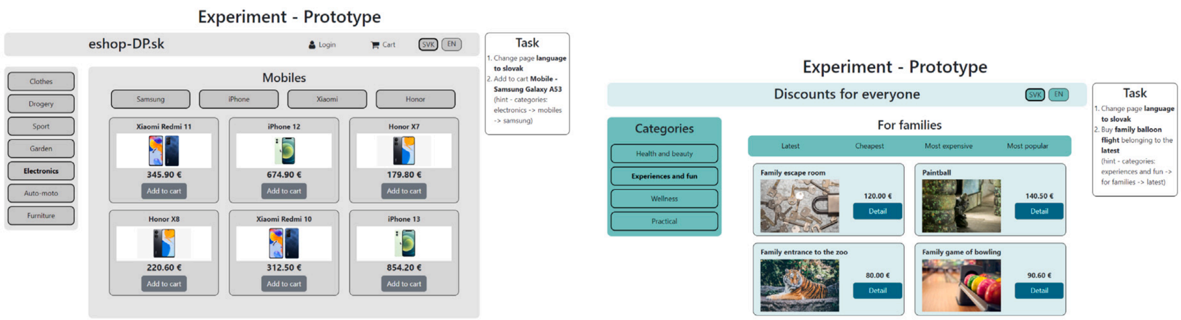
Maintaining consistent sequences of tiles across different mouse parameter configurations can allow for more exact comparison of the effects that the mouse configurations have on mouse dynamics. However, it also creates the risk of participants training themselves on repeating sequence patterns, becoming more efficient in later iterations of the activity. To avoid this, the number of rounds is doubled to the final total of six. Three sequences of squares have the exact same appearance for every mouse configuration while the other three are semi-randomized — the tile patterns remain fixed but the position of squares within the tiles is randomized, giving the sequences a different appearance. By also presenting the six sequences to participants in random order, the fact that the sequences repeat themselves is obscured.

4.2.3. Website prototype interaction activity

The third observed activity has the same goals as the second, but with the added aspect that the sequence is clicked by participants to solve a realistic task in the interactive prototype of a website. Effectively, the task is designed for the clicking behavior to be similar to a usability test. This is to allow for the analysis of the effects of mouse parameter configurations in a more practical scenario on top of the abstract numbered point clicking task.

As a requirement for comparing mouse parameter configurations, the ideal mouse movements in prototype interactions needed to be essentially the same for each configuration. To minimize the chance that participants would train themselves on the same sequence by completing the same task multiple times, similarity of mouse movements is obscured by presenting tasks with different meanings. At least one task variant was needed for each of the five mouse parameter configurations.

Two website prototypes were designed — one of an ecommerce website, the other of a coupon website. Three task variants were created for both sites. In each variant, the context of the task provided within the instructions is different. Labels within the prototypes are adjusted during each task to adapt to its context. Between the six task variants



(a) Variant 1 - Ecommerce website prototype (b) Variant 2 - Coupon website prototype

Fig. 4. Last of three activities in the experiment is an interaction with a prototype. Variant prototypes have different appearance and tasks, but identical positions of clickable elements.

in total, the ideal mouse paths (layout of elements and the positions of buttons to be clicked) are either completely identical (within the same prototype) or with negligible differences (between the prototypes). The ideal mouse movements have the same length and direction between the two prototypes, only being distinguished by a small number of pixels for the sake of granting prototypes distinct appearances). The study moderator assigns tasks in a deterministic order so that each task is completed by participants an even number of times with each configuration. See Fig. 4 for illustration of prototypes and paper's data repository for full instructions.

To prevent participants from getting lost within the prototype, at any moment during the task, only a single button in it is clickable.

4.3. Data collection and measures

In a controlled experiment, mouse tracking data was obtained for a selection of mouse configurations. Each participant completed the procedure with every mouse parameter configuration. CPI (Count Per Inch) and sensitivity settings in the operating system were chosen as the primary independent variables (collectively making up the eCPI variable by multiplication). CPI was chosen as an inherent property of the physical device owned by the user that is potentially but not always adjustable, while the sensitivity setting in the operating system is a commonly used option for modifying the speed of cursor movements. To ensure that the eCPI is the sole variable responsible for potential differences between samples, the aim was to minimize the impact of outside factors of the environment by conducting the experiment presentially in laboratory conditions.

4.3.1. Mouse parameter configurations

To render comparisons between mouse parameter configurations feasible, specific mouse parameter configurations are granted focus. Web browsers throttle the frequency of mouse events due to performance reasons. A full list of mouse events used to calculate aggregate events may be accessed,[9] but for the purposes of this work, mouse events are observed at a consistent polling frequency of 60 Hz, as returned by the V8 JavaScript engine in Google Chrome Browser.

In total, three computer mouse devices are used for the experiment, chosen to represent a scale of common mouse devices,[10] including average, above-average and gaming-mouse grade CPI (1000, 2400 and 4800 respectively). The typical CPI range for a non-gaming mouse is between 800 and 1600, while a gaming mouse commonly provides multiple CPI settings. The default mouse sensitivity of the Windows operating system (10) was used for all devices. With the 2400 CPI mouse, the experiment was also conducted on system sensitivity levels 5 and 18, to achieve the effect where the mouse cursor moves with the perceived speed closer to the low and high CPI mouse devices. See Table 3 for summary of configurations with corresponding eCPI[11] values.

4.3.2. Mouse dynamics metrics

Mouse movement metrics are calculated from pairs of consecutive mouse tracking data points — their coordinates along horizontal/vertical axis and corresponding timestamps. Among others, base metrics include distance, time and angle differences. One hundred twenty-one (121) mouse movement metrics are resolved in total by aggregating the base metrics per round, calculating their means, standard deviations, etc., (see Table 4). The list of all formulas corresponding with all the metrics is found in Appendix B.

Specific mouse dynamics of temporal and spatial nature were picked from related works according to plausible expectations for how they may be affected by mouse configuration. Multiple aggregates are deployed at once, since each of them represents a different aspect of mouse dynamics, some of which may be affected by changes in mouse configuration, while others may not. For example, the 5th percentile may be more informative for some metrics than the minimum value.

4.3.3. Questionnaire

To collect additional information about the participants and the research activity, a questionnaire appears before every activity iteration. This questionnaire is divided into two blocks — the profiling questionnaire and the meta information questionnaire.

The profiling block of the questionnaire is filled in by the participants only before the first experiment iteration, enabling analysis of the sample's demographics and computer literacy. Questions in the questionnaire include the participant's age, sex (male, female and intersex), computer expertise, computer use frequency and employment status. The moderator handles noting down the currently set up mouse parameter configuration.

4.4. Participants

In total, 32 participants ($M_{age} = 31.3, SD_{age} = 13$) have completed participation in the controlled experiment. Average length of the experiment is 22 min ($min = 17, max = 28$). There is an equal 16-to-16 ratio of the binary sexes (no intersex answers were received), the age distribution between the sex groups being similar ($M_f = 33.44, SD_f =$

[9] Raw mouse events are accessible in a web browser as coalesced events: https://www.w3.org/TR/pointerevents3/#dfn-coalesced-events.

[10] Patriot Group (computer hardware manufacturer) blog article — "Gaming Mouse DPI: Is it Important?": https://store.patriotmemory.com/blogs/news/gaming-mouse-dpi-is-it-important.

[11] eCPI calculation verified via eDPI calculator by Omni Calculator: https://www.omnicalculator.com/other/edpi.

Table 3
Five mouse configurations utilized by participants during the experiment.

Config. no.	Mouse model	CPI	Sensitivity (OS Windows)	eCPI
#1	Microsoft Wireless Mobile Mouse 1850	1000	10	10,000
#2	Dell wm116p 1850	2400	5	12,000
#3	Dell wm116p 1850	2400	10	24,000
#4	Dell wm116p 1850	2400	18	43,200
#5	Trust GXT 101 GAV	4800	10	48,000

Table 4
Mouse dynamics metrics evaluated for impact of mouse configuration.

Feature name	Aggregates	Feature description
distance	*Mean, sd, median, min, max, Q5, Q95, sum*	Distance between consecutive points
flip_x	*Count*	x-axis movement direction change
flip_y	*Count*	y-axis movement direction change
duration	*Sum*	Movement time between consecutive points
velocity	*Mean, weight. mean, sd, median, min, max, Q5, Q95*	Cursor distance over time
velocity_x	*Mean, weight. mean, sd, median, min, max, Q5, Q95*	Cursor distance over time (x-axis)
velocity_y	*Mean, weight. mean, sd, median, min, max, Q5, Q95*	Cursor distance over time (y-axis)
velocity_smooth	*Mean, weight. mean, sd, median, min, max, Q5, Q95*	Smoothed cursor distance over time
pace	*Mean, weight. mean, sd, median, min, max, Q5, Q95*	Time over cursor distance
acceleration	*Mean, weight. mean, sd, median, min, max, Q5, Q95*	Cursor velocity over time
acceleration_x	*Mean, weight. mean, sd, median, min, max, Q5, Q95*	Cursor velocity over time (x-axis)
acceleration_y	*Mean, weight. mean, sd, median, min, max, Q5, Q95*	Cursor velocity over time (y-axis)
acceleration_positive	*Mean, weight. mean, sd, median, min, max, Q5, Q95*	Cursor velocity over time (positive)
acceleration_negative	*Mean, weight. mean, sd, median, min, max, Q5, Q95*	Cursor velocity over time (negative)
angle	*Mean, sd, median, min, max, Q5, Q95*	Current cursor movement direction
corner	*Count*	Orthogonal angle change
velocity_angular	*Mean, weight. mean, sd, median, min, max, Q5, Q95*	Movement direction difference over time
curvature	*Mean, sd, median, min, max, Q5, Q95*	Angle change over distance
deviation	*Mean, sd, median, max, Q5, Q95*	Perpendicular distance to ideal path
straightness	–	Ratio of ideal path to the movement one

$14.30, M_m = 29.19, SD_m = 11.70$). Majority of participants (84%) use the computer at least once a day. Men in the sample report using the internet more often than women, and express higher confidence with using a computer ($M_f = 6.94, SD_f = 2.62, M_m = 8.44, SD_m = 1.60$). This is in general alignment with known statistics, such as that internet usage is more widespread among men.[12] Groups of each sex contain representatives of varied (primarily white-collar) professions; participants who are students pursue a variety of fields.

A pilot experiment was conducted beforehand with a single participant and conditions identical to the genuine participation phase. No significant findings impacting the experiment design were made during the pilot. Inspection of mouse interaction data revealed mouse movements corresponding with relevant mouse targets, qualifying the experiment design as collecting the required data.

4.5. Data processing and analysis

Mouse tracking data is aggregated into mouse dynamics metrics (see Table 4) within the scope of each participant's completion of a task using a given mouse parameter configuration.

The Shapiro test is used to assess the normality of calculated metrics. For assessing differences between mouse configurations and between the sexes, paired and unpaired tests are used depending on the sample: paired non-parametric Friedman tests with Kendall's W effect size, non-parametric Kruskal–Wallis tests with η^2 effect size or Mann–Whitney U nonparametric tests with the r effect size. Out of the total 123 features (121 mouse dynamics metrics plus mouse CPI and sensitivity), 108 were selected as informative and used in further analysis. Remaining 15 features with zero variance were discarded.

[12] Statista - Internet usage penetration ratios among genders as of October 2023: https://www.statista.com/statistics/1387693/penetration-rate-of-the-internet-by-gender/.

4.6. Dataset

The processed dataset (available in the paper's data repository) consists of the full sample obtained from the three experiments, each instance containing a vector of 123 features. For clarity, the dataset is divided by the experiment activity where its data originates from:

Target shooting minigame dataset. While the mouse dynamics from warm-up activity are not utilized directly for analysis of results, feature sets are calculated for the sake of completeness and transparency. The sample size is 480 — the product of the number of participants (32), the number of mouse configurations (5) and the number of minigame rounds (3).

Numbered points clicking dataset. Records of participants clicking on sequences of points with varied mouse configuration. The number of instances in the dataset (958), is the product of the number of participants (32), the number of mouse configurations (5) and the number of rounds involving clicked point sequences (6). Two instances are subtracted due to missing data caused by technical malfunction during the experiment involving two participants — #7 (female) and #15 (male). These participants are excluded from statistical analysis which entails paired tests, but are kept as valid inputs for the classification task.

Website prototype interaction dataset. 160 instances that capture 32 participants' interactions with website prototypes using 5 different mouse configurations. Utilized in statistical analysis.

4.7. Prediction method

The prediction task of classifying biological sex is understood as a binary classification, simplified compared to real-world biological sex by containing no intersex option. For machine learning application, metrics are standardized and transformed using a Yeo–Johnson transformer. Six machine learning models are used — Logistic regression, Support vector machine (SVM), Random forest, XGBoost, Catboost, and LightGBM.

To assess whether the mouse configuration differences have actual impact on sex classification performance, classification models are trained on data from a conjunction of a subset of the five tested mouse configurations. The conjunction of the training set from multiple configurations is done to assure sufficient size of the training sample. The model was trained either on the mouse configurations with the lowest eCPI (#1, #2, #3), or with the highest eCPI (#3, #4, #5). To assess differences, prediction was tested on unseen users using the same configuration as the training sample, or those using the remaining configurations (#4 and #5 for the first training group and #1 and #2 for the second one respectively).

Identifying a concern that the training and testing datasets would not contain data from the same participants (even though they are using different mouse configurations) the sample is split for the prediction task. The training sample contains data from 24 participants (12 women, 12 men) and a testing sample from 8 participants (4 women, 4 men). All machine learning performance metrics are calculated as the mean of 10 trials, with the participants split into different testing and training sets randomly for each trial (the same 10 splits utilized for every machine learning method).

Feature selection is performed by choosing statistically significant features determined by the Mann Whitney U-test, followed by scikit-learn's[13] SelectFromModel feature selection. Optimization of prediction is achieved via randomized grid search for hyperparameter tuning, with tweaking of the cutoff threshold value using AUC scores (the best true positive to false positive ratio). Accuracy (ACC), recall (R), precision (P), F1 score ($F1$) and AUC score (AUC) are observed as performance metrics.

5. Results

5.1. Mouse parameter impact on mouse dynamics (RQ1)

RQ1: What effect do different mouse eCPI configurations have on metrics calculated from raw mouse interaction data?

To assess the impact of individual mouse configurations, employing the Friedman test, differences between five mouse configurations are gauged by calculating mouse dynamics metrics using data from the numbered points clicking and prototype interaction activities. The full list with all affected metrics can be found in Table 9.

During the numbered points clicking activity, mouse sensitivity significantly affects 83 out of 108 calculated metrics. Among the five most affected are the median of positive acceleration, $\chi^2(2, N = 480) = 150.81, p < .001, W = .42$, mean of positive acceleration, $\chi^2(2, N = 480) = 133.30, p < .001, W = .37$, maximum velocity, $\chi^2(2, N = 480) = 115.88, p < .001, W = .32$, median of negative acceleration, $\chi^2(2, N = 480) = 115.14, p < .001, W = .32$ and mean of regular acceleration, $\chi^2(2, N = 480) = 112.30, p < .001, W = .31$.

During the prototype activity, 73 out of 108 metrics are significantly affected by mouse sensitivity, most among them the fifth percentile of acceleration, $\chi^2(2, N = 480) = 37.56, p < .001, W = .59$, 95th percentile of acceleration, $\chi^2(2, N = 480) = 36.19, p < .001, W = .57$, fifth percentile of acceleration along the y axis, $\chi^2(2, N = 480) = 36.19, p < .001, W = .57$, 95th percentile of positive acceleration, $\chi^2(2, N = 480) = 35.69, p < .001, W = .56$ or mean of negative acceleration, $\chi^2(2, N = 480) = 34.56, p < .001, W = .54$.

CPI has a significant impact on 89 out of 108 mouse dynamics metrics during the numbered points activity. The highest impact was discovered to be on the median of positive acceleration $\chi^2(2, N = 480) = 208.21, p < .001, W = .58$, median of negative acceleration $\chi^2(2, N = 480) = 207.74, p < .001, W = .58$, mean of positive acceleration $\chi^2(2, N = 480) = 178.01, p < .001, W = .49$, 5th percentile of acceleration, $\chi^2(2, N = 480) = 174.98, p < .001, W = .49$ or total duration, $\chi^2(2, N = 480) = 169.88, p < .001, W = .47$.

During the prototype activity, CPI affected 76 out of 108 features. Metrics such as mean of acceleration, $\chi^2(2, N = 480) = 34.31, p < .001, W = .54$, 95th percentile of positive acceleration, $\chi^2(2, N = 480) = 33.25, p < .001, W = .52$, mean of positive acceleration, $\chi^2(2, N = 480) = 32.25, p < .001, W = .50$, mean of acceleration along the y axis, $\chi^2(2, N = 480) = 31.94, p < .001, W = .50$ or 95th percentile of regular acceleration, $\chi^2(2, N = 480) = 31.94, p < .001, W = .50$ were significantly affected.

In all cases, among the most impacted metrics are the aggregates of acceleration, velocity and distance, although each feature group (e.g., acceleration related metrics, velocity related metrics) contains significantly affected aggregates. When participants are divided into five groups by eCPI (product of CPI and mouse sensitivity), eCPI affects 95 out of 108 features in the numbered points clicking activity and 84 out of 108 in the prototype activity.

5.2. Adjustment to new mouse configuration (RQ2)

RQ2: What amount of mouse activity do users need to adapt to a new mouse, so that differences in mouse dynamics between adapted users can be safely compared?

During each iteration's target shooting minigame, when participants first start using a new mouse configuration, according to the Friedman test, there are significant differences between the three consecutive rounds in terms of game performance, evaluated as target hits for the whole sample, $\chi^2(2, N = 495) = 77.46, p < .001, W = .24$.

When evaluating differences in game performance for individual mouse configurations in a specific order of iterations (ascending or descending eCPI), a significant difference between consecutive rounds of the minigame is present only for the highest eCPI configuration when it is presented as first, $H(2, n = 48) = 10.21, p = .006, \eta^2 = .18$ (Kruskal–Wallis test). When the order of iterations is discounted, significant differences between rounds are present solely for the highest eCPI configuration. Following post hoc tests in both cases show that differences are pronounced only between the first and the third (last) round ($p = .002$).

Conducting comparison between different placements of the same mouse configuration depending on the iteration order, significant differences are found for the high eCPI configurations #4 and #5. For the second-highest eCPI configuration #4, depending on whether it was presented during the second or the fourth iteration, significant differences in game performance exist during the first round, $z = -1.98, p = .045, \eta^2 = -.35$ (Mann Whitney test). For the highest eCPI configuration #5, depending on whether it was presented during the first or the last iteration, significant differences manifest between both the first round, $z = -3.13, p = .002, r = -.55$, and the second round, $z = -2.02, p = .042, r = -.36$. These differences can be further highlighted by observing that when the activity starts with highest eCPI, target hits improve between rounds the most significantly ($M_{r1} = 10.13, M_{r2} = 11.81, M_{r3} = 13.06$), while during the last iteration the target hits are more consistent ($M_{r1} = 13.38, M_{r2} = 13.63, M_{r3} = 13.94$).

5.3. Mouse parameter impact on classification of sex (RQ3)

RQ3: What is the effect of different mouse eCPI configurations on the performance of prediction using mouse dynamics features?

Prediction of biological sex is developed to illustrate the prospect of eCPI influencing a machine learning model's accuracy. Since innovation or optimization to achieve the best prediction of biological sex are not part of the focus of this work, standard methods and models are employed in a straightforward manner. To apply machine learning to mouse dynamics features for the task of predicting the user's biological sex, relevant mouse dynamics features are selected according to their statistical significance, as calculated by the Mann–Whitney U test. Features in the prediction are resolved from the data obtained as part of the numbered points clicking activity.

[13] Scikit-Learn, python library with machine learning tools: https://scikit-learn.org/stable/.

Table 5
Before the tuning, prediction of biological sex using mouse dynamics data from all mouse configurations yields the best results for Logistic regression, Random Forest and CatBoost. *After tuning, the most accurate model is CatBoost, achieving the F1-score of .625.

Model	Sensitivity & CPI info	Features	Accuracy	Precision	Recall	F1 score	AUC score
CatBoost	–	15	.587	.588	.587	.587	.635
	Included	17	.590	.591	.590	.589	.635
LightGBM	–	15	.584	.584	.584	.583	.614
	Included	17	.571	.571	.571	.570	.605
Logistic Regression	–	15	.608	.608	.608	.608	.643
	Included	17	.606	.606	.606	.606	.641
Random Forest	–	15	.591	.592	.591	.591	.627
	Included	17	.587	.587	.587	.586	.628
SVM	–	15	.534	.536	.535	.533	.554
	Included	17	.529	.531	.529	.526	.546
XGBoost	–	15	.567	.567	.566	.566	.601
	Included	17	.585	.586	.585	.584	.607

Table 6
Best performing models after hyperparameter tuning, feature selection and ROC threshold optimization.

Tuned model	Accuracy	Precision	Recall	F1 score	AUC score
CatBoost	.612	.639	.616	**.592**	.621
Logistic Regression	.6	.619	.6	.575	.585
Random Forest	.610	.626	.610	.585	.6

Sex differences significantly affect 61 out of 108 mouse dynamics features according to Mann Whitney tests. The most impacted features include standard deviation of curvature, $z = 6.01, p < .001, r = .20$, negative weighted mean of acceleration, $z = -5.46, p < .001, r = -.18$, the 95th percentile of curvature, $z = 5.4, p < .001, r = .18$, mean of curvature, $z = 5.39, p < .001, r = .18$ and the 5th percentile of negative acceleration, $z = -5.15, p < .001, r = -.17$. Overall, the most affected features are related to curvature, acceleration and velocity, even though there is at least one aggregate in each feature group that is affected. Corresponding to the size of the sample, the top 15 most affected features were selected as the input for machine learning. The full list with all affected features is found in Table 9.

To validate classification of biological sex based on mouse dynamics features as a suitable representation of a prediction task where the impact of mouse configuration can be explored, machine learning models were trained using data from all mouse configurations in the numbered points clicking activity. A version of each model is also trained with additional features explicitly containing information on mouse sensitivity and CPI to assess their potential role in prediction. See Table 5 to see the performance metrics of all models before the tuning.

Before performance metrics of the prediction methods are reported, note that the objective of this work is not present a novel model with remarkable performance. Rather, it is to show how performance metrics can be changed by difference in mouse dynamics. Among the baseline models, the best is Logistic Regression, ($F1 = .608, AUC = .627$) without CPI and sensitivity features and ($F1 = .606, AUC = .628$) and with the same features. Direct addition of CPI and sensitivity features does not significantly impact the classification performance metrics. Furthermore, feature selection did not select CPI and mouse sensitivity as relevant features. Three best-performing algorithms – CatBoost, Random Forest and Logistic Regression – were tuned for final classification of biological sex. After hyperparameter tuning, feature selection and choosing an optimized classification probability threshold in accordance with the ROC curve (see Fig. 5) the final best performing model is CatBoost that classifies unseen participants' sex with $ACC = .612, F1 = .592, AUC = .621$ (see Table 6).

With classification of biological sex based on mouse dynamics proven as viable through a selection of multiple machine learning methods, the prerequisite condition for assessing the impact of mouse configuration on a realistic case of classification has been satisfied. To investigate the answer to the research question, prediction performance is assessed separately for seen and unseen mouse configurations (see 4.7 Prediction method). Resulting performance metrics (see Table 7) show that when the models are tested with mouse dynamics obtained with the same mouse configuration, overall, the performance metrics achieve better results than with other unseen configurations (even if sometimes the size of the difference is small). Similar result differences are also visible in AUC scores. Answer to research question RQ3 follows: evaluation with different mouse eCPI configurations can result in non-negligible differences in prediction performance.

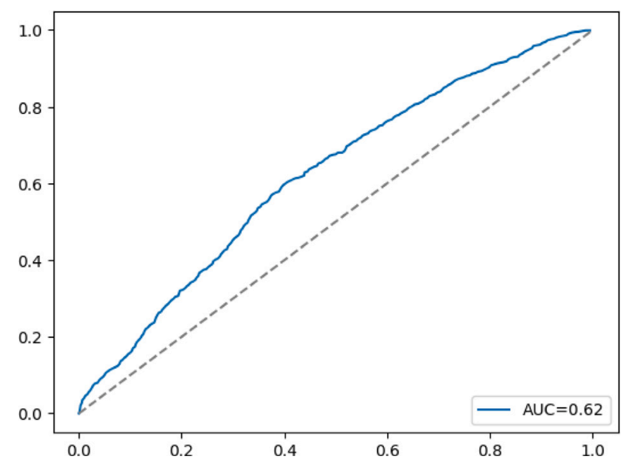


Fig. 5. ROC curve (average) of final CatBoost predictive models with the F1 score of .592, used for the model optimization.

6. Discussion

Learnability of mouse control (see RQ2) presents itself as a valid concern for experiment design, particularly in experiments with users asked to work with high-eCPI mouse devices. Findings presented here can be attributed to the effect of mouse configurations themselves, rather than the temporary effects of the participants adapting to the mouse configuration. This is courtesy of the target shooting minigame activity, which precedes the other activities involved with evaluation of the effects of mouse configurations on mouse dynamics and machine learning prediction featuring them (RQ1 and RQ3 respectively). The

Table 7
Comparison of performances of classification models according to the source of mouse dynamics features used for training and testing. Datasets #1 to #5 group mouse dynamics from five mouse parameter configurations in ascending order of eCPI (1 - lowest, 5 - highest). When trained on low eCPI and tested on high eCPI, or the exact opposite, prediction performance metrics drop.

Model	Trained on	Tested on	Accuracy	Precision	Recall	F1 score	AUC score
CatBoost	#1, #2, #3	#1, #2, #3	.553	.553	.553	.550	.596
		#4, #5	.576	.578	.576	.574	.605
	#3, #4, #5	#3, #4, #5	.587	.588	.587	.586	.636
		#1, #2	.574	.575	.574	.572	.608
LightGBM	#1, #2, #3	#1, #2, #3	.554	.553	.552	.551	.576
		#4, #5	.526	.526	.526	.526	.557
	#3, #4, #5	#3, #4, #5	.568	.570	.570	.567	.605
		#1, #2	.563	.565	.563	.561	.572
Logistic Regression	#1, #2, #3	#1, #2, #3	.601	.601	.601	.600	.638
		#4, #5	.575	.575	.575	.574	.609
	#3, #4, #5	#3, #4, #5	.620	.620	.620	.619	.646
		#1, #2	.583	.583	.583	.583	.606
Random Forest	#1, #2, #3	#1, #2, #3	.562	.562	.563	.560	.595
		#4, #5	.541	.541	.541	.540	.576
	#3, #4, #5	#3, #4, #5	.606	.609	.607	.605	.646
		#1, #2	.572	.573	.572	.569	.608
SVM	#1, #2, #3	#1, #2, #3	.614	.615	.615	.613	.639
		#4, #5	.576	.576	.576	.575	.608
	#3, #4, #5	#3, #4, #5	.612	.611	.611	.610	.643
		#1, #2	.577	.577	.577	.577	.599
XGBoost	#1, #2, #3	#1, #2, #3	.562	.563	.562	.560	.580
		#4, #5	.547	.548	.547	.547	.575
	#3, #4, #5	#3, #4, #5	.586	.586	.586	.585	.627
		#1, #2	.571	.572	.570	.568	.567

difference in game performance between rounds 1 and 3 being no longer present between rounds 2 and 3 attests to the participants growing accustomed to the mouse configuration in the course of the game. Achieving more consistent game performance exclusively in instances of users wielding a mouse with high eCPI, confirms that the user has adapted to mouse configuration, not only the task (game) itself. Because participants may be used to different mouse configurations, in experiments where participants are not using their own devices, a warm-up activity of similar intensity to the target-shooting minigame should be employed, letting participants adapt to the mouse configuration to a sufficient degree. In within-subject research methods, this may be seen as needed, to prevent the acclimatization to a mouse configuration from having an effect on user behavior during the initial condition.

According to our analysis of the answer to RQ1, the majority of mouse dynamics metrics evaluated are significantly affected by mouse configuration. Among the top affected features are primarily the ones related to acceleration and velocity, but there are significantly affected metrics in all assessed metric groups. The evaluated metrics were all adopted in reference to existing research on mouse dynamics, which to various degrees neglects to address mouse parameter configuration (see 2.2 Mouse dynamics features). Thus, the sets of mouse configuration parameters utilized in research design (whether specified implicitly or explicitly) should be considered a key design variable. This claim does not seek to invalidate any previous research that employs mouse dynamics without fully addressing mouse configuration, but rather to contextualize it. More transparent and thorough reporting of mouse configuration can prevent potential methodological concerns about the presence of unaccounted-for third variables. Broader discussion about ecological validity can be opened about methods and techniques to be used in uncontrolled environments, where mouse configuration parameters may vary between different users in real-world conditions, but also between interactions by the same user.

Commonly used mouse dynamics depend on both the counts per inch (CPI) of the mouse device and the sensitivity setting in the operating system. CPI and sensitivity act as variables in the function mapping the speed of the physical mouse device to the effective speed of the virtual mouse cursor. Reports of controlled experiments linked to mouse dynamics should state the details of the mouse configuration being used to ensure the experiment's reproducibility, in agreement with Schoemann et al. [16], so that future research is comparable to the state of the art. Without information on mouse configuration, results achieved by research have ambiguous meaning (eCPI could be the hidden variable that caused any differences). The information that research should report includes the mouse CPI and sensitivity set in the operating system or application. The state of the acceleration setting being turned on or off should be reported as well, since it also affects the mutual relation of hand speed and cursor speed.

Aforementioned differences in mouse dynamics caused by the mouse configuration are observed not only in the more abstract numbered points clicking activity, but also in the website prototype interaction activity. Mouse dynamics' adjustment to mouse parameters may therefore be reasonably expected in realistic mouse interaction scenarios, such as internet browsing. Discrepancies in which of the specific mouse dynamics are affected between the two activities can be explained by differences in the activity context (task, user interface layout). The numbered points clicking task is cognitively linear, engaging only the pre-school level ability to count to 8. By contrast, in the task of visually scanning the user interface of a website and determining the strategy to solve a problem (the meaning of which is also more open to personal interpretation than straightforward counting), the participant's reasoning skills enter the equation. Mouse dynamics may manifest differently due to variable type of cognitive activity.

When high eCPI is presented to participants as the first mouse configuration, learning (witnessed via the improvement in game performance between rounds) is significant. By comparison, if participants are exposed to high eCPI during the latter half of the experiment, there is a comparatively notable lack of similar performance growth. This suggests that the act of learning to use a specific mouse configuration can warp mouse behavior further, particularly during tasks that users are unfamiliar with. When a participant is learning in multiple aspects at the same time — mastering an unfamiliar mouse configuration while also completing a research task that they have just been introduced to, this can be expected to multiplicatively increase the steepness of the learning curve.

Notably, the target shooting minigame designed for our experiment is a mouse-interaction-heavy activity, with the 15-second rounds having a low level of granularity for the marking of changes in performance. It is possible that a certain degree of learning is needed to become accustomed to eCPI values other than the highest eCPI, but this was not captured due to the granularity level. The target shooting minigame seen in the study is intensive enough to arguably enable fast learning, within the scope of a single round. If interactions that are more sparse on mouse movements were presented without even a single round of a similar warm-up, mouse dynamics may continue being affected by continuous learning. Further research on this topic is warranted.

The machine learning task of predicting the user's sex based on mouse dynamics features shows a significant impact of mouse configuration on prediction performance (see RQ3). The final tuned model is able to detect the user's sex with the F1 score of .592. However, investigation shows that prediction performance does tend to drop when the classified input uses a mouse configuration with an eCPI that was not part of the training sample. Because of the significant statistical effect of mouse configuration on most mouse dynamics, we posit that the differences in performance caused by mouse configuration could potentially even be higher in prediction models that are more effective than the model used in the case study. A model that is seemingly well-adjusted to solving its particular prediction task could behave in unexpected ways when exposed to data retrieved from unknown mouse configurations. This bears significant implications on methodological credibility of any past and future works that use mouse dynamics as the basis of making predictions. Without accounting for varied mouse configurations, there are limitations to understanding the meaning of results achieved via mouse-dynamics-based methods, their replication and comparison with other methods.

The fact that mouse parameters used as explicit features are found irrelevant for prediction may appear counterintuitive at first, given the above findings about mouse configuration impacting mouse dynamics and prediction performance. However, this can be interpreted as the information about eCPI already being encoded in the mouse dynamics that it affects, thus making a further explicit declaration or the parameters redundant information for classification purposes.

Differences between the mouse dynamics of the biological sexes (a prerequisite for successful classification) could be caused by a variety of factors, including physiological, psychological, social or behavioral. Because of people's individual differences, while some factors may be inherently linked with biological sex, others may merely have some degree of correlation without being definitive indicators (e.g., male/female stereotypical habits such as gaming with preferences for specific genres or platforms could impact typical mouse behavior). Identifying, understanding and controlling these variables may lead to more effective classification of the biological sex. Since the reason for the use of classification in this research is to explore how classification results can be affected by alterations to mouse configurations (the primary controlled conditions), rather than the development of classification itself, consideration of these aspects is out of scope of this paper.

In the more website prototype interaction activity, reduced differences between the sexes may be explained by higher cognitive complexity of the task. Explorative analysis shows that mouse trajectories are considerably more different in this task, as participants actively look for the solution (instead of simply visually scanning for a number in the numbered points clicking activity). This is true for both observed sexes. Different methods of aggregating mouse dynamics may be suitable to identify sex differences in website interaction.

7. Implications

The following implications can be drawn for practical application in mouse dynamics research:

- Methodologies for conducting controlled experiments to obtain data for mouse dynamics should report the parameters of the mouse configuration being used (CPI, sensitivity at minimum). In within-subject research, explicit attention should be paid to mouse configuration being consistent between condition environments. For examples of impacted research, see: Monaro et al. [2], Yamauchi et al. [8], Balen et al. [9], Siddiqui et al. [10], Hucko et al. [15], Almanji et al. [44].
- To achieve more flexible machine learning models, multiple varied mouse configurations should be deployed in parallel as part of data collection methods.
- If the source of mouse dynamics is an uncontrolled experiment where reporting the parameters of the observed mouse configurations is not feasible, the methods utilizing thus obtained datasets should be validated on a testing sample representing varied mouse configurations. For examples of impacted research, see: Kuric et al. [1], Fernández-Fontelo et al. [4], Gardey et al. [5], Sulikowski et al. [14], Pepa et al. [23], Arapakis and Leiva [30], De Santana et al. [42].
- For mouse dynamics to mirror natural mouse behavior, they should be harnessed from eCPI that the participants are used to, either by verifying the participant's native mouse configuration, or exposing them to the new configuration via sufficient warm-up.
- Mouse dynamics normalization approaches (as covered in 2.4 Normalization of mouse dynamics features) may have value (or could be seen as imperative) for methodologically correct use of mouse dynamics in real-world conditions where consistent mouse configuration is not guaranteed.

8. Limitations and future work

Conducting a controlled experiment involving multiple different mouse configurations from the same users is time-taxing. Although the obtained dataset is of suitable size to answer our research questions, a more rich dataset could help answer followup questions, such as how mouse dynamics change when the user is adjusting to a new mouse configuration. It may be desirable to conduct an experiment with known information about mouse parameter configuration over longer periods of time, with more participants and involving real usage of a computer (e.g. internet browsing). A bigger dataset could provide additional variability that may contribute to more highly accurate predictive models and statistical results.

All participants in the experiment use mouse configurations in one of two predetermined orders (ascending and descending eCPI) rather than a different (e.g. fully randomized order). This means that the dataset does not contain data on swaps between mouse configurations that might be valuable for further analysis of user adaptation to a different mouse configuration. Future research dedicated to mouse learnability may be inclined to accentuate its focus on the order in which mouse devices are presented.

Of all the mouse configuration parameters, two were assessed in this study — the CPI of the mouse hardware and the mouse sensitivity setting via software. The acceleration setting in the operating system was not assessed by this work. This setting (as explained in 2.1 Computer mouse parameters) causes the cursor to move longer distances the faster the mouse movements are and vice versa. This could potentially have an impact on the mouse dynamics features as well, as well as other factors such as mouse polling frequency, weight, ergonomics, screen size and physical properties of the mousepad/surface beneath the mouse.

Normalization of participants' game performance between rounds of the initial target shooting minigame when the mouse eCPI is high shows that users learn to use the new mouse configuration. Due to the low granularity of the rounds, the experiment does not prove or disprove whether more nuanced deviation in eCPI also leads to temporary abnormalities in mouse behavior of a smaller scope. It is possible that the user may quickly adapt to small eCPI changes during a fast-paced game, where there is ample opportunity to practice their aim. Meanwhile, during less aim-focused interactions, mouse behavior abnormalities may be diffused into mouse dynamics over the length of a longer interaction. Assessment of differences in initial mouse movements after changing mouse eCPI in realistic usage conditions could justify or repudiate the need for a warm-up task when tracing the usage of a mouse that the user is unfamiliar with.

Given that the inclusion of a mouse configuration in the training set increases the likelihood that samples using this configuration will be classified correctly, this raises the question about the relationship between the number of mouse parameter configurations used for training and prediction performance. More representative training set could simply mean more accurate prediction, or there could be a tradeoff between the number of mouse configurations and the accuracy of the prediction model. Future research could design methods for optimization of threshold on the number of mouse configurations, or for selection of the most effective mouse configurations to train prediction models on.

The role that mouse configuration plays for mouse dynamics may be more crucial in use cases like biometric authentication. In comparison to predictions of more general phenomena, such as user characteristics, intent, attention or emotion, identification of a specific user is at severe risk of becoming unreliable if mouse dynamics are highly dependent on what mouse parameter configuration the individual user is currently on.

9. Conclusion

Are mouse dynamics a credible source of information that can drive unobtrusive user modeling, recommendations, or authentication? A conclusion can be drawn that they still probably can be. But a methodology supporting the usage of mouse dynamics also needs to acknowledge that the mouse at the user's disposal is not always the same device. Mouse dynamics obtained with each individual mouse are ruled by a unique set of technological and environmental circumstances, introducing an additional layer of variability to account for. Contemporary classification methods demonstrably perform worse on unseen mouse configurations. At the same time, it is common for works focusing on mouse dynamics to lack proper reporting on the mouse parameters used and/or to not evaluate how different mouse parameter configuration may affect its results. With this work, we hope to open up a discussion that will lead to improvement of future methodologies involving mouse dynamics and to advancement in ecological validity of related methods.

CRediT authorship contribution statement

Eduard Kuric: Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Peter Demcak:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Matus Krajcovic:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Peter Nemcek:** Writing – original draft, Visualization, Validation, Software, Investigation, Formal analysis, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data, analysis scripts written using Python as well as the experiment's source code are openly available in a public paper repository at github.com/micemicsresearch/mouse-dynamics-data-credibility.

Acknowledgments

This work was supported by the Operational Program Integrated Infrastructure for the project: Support of Research Activities of Excellence Laboratories STU in Bratislava, project No. 313021BXZ1, co-funded by the European Regional Development Fund (ERDF).

Appendix A. Research utilizing mouse dynamics

Mouse dynamics as utilized for varied purposes in literature, structured by type of research and how mouse configurations is addressed, available in Table 8.

Appendix B. Mouse dynamics metrics formulas

Assuming $\Delta x = x_{i+1} - x_i$, $\Delta y = y_{i+1} - y_i$ and $\Delta t = t_{i+1} - t_i$ (coordinate and time differences between consecutive data points), list of derived metrics is available below. These metrics are further aggregated as counts, means, etc.

The *distance* is an euclidean distance between two consecutive points:

$$distance = \sqrt{(\Delta x)^2 + (\Delta y)^2} \tag{1}$$

The *flip* is a change in direction. Direction is a binary variable of value 1 if the distance change of the coordinate is positive (movement to the right for *flip_x* or down for *flip_y*), otherwise it is −1:

$$direction\_x = \begin{cases} +1 & \Delta x > 0 \\ -1 & \Delta x < 0 \\ 0 & otherwise \end{cases} \tag{2}$$

$$flip\_x = \begin{cases} 1 & direction\_x_i = 1 \land direction\_x_{i-1} = -1 \\ 1 & direction\_x_i = -1 \land direction\_x_{i-1} = 1 \\ 0 & otherwise \end{cases} \tag{3}$$

$$direction\_y = \begin{cases} +1 & \Delta y > 0 \\ -1 & \Delta y < 0 \\ 0 & otherwise \end{cases} \tag{4}$$

$$flip\_y = \begin{cases} 1 & direction\_y_i = 1 \land direction\_y_{i-1} = -1 \\ 1 & direction\_y_i = -1 \land direction\_y_{i-1} = 1 \\ 0 & otherwise \end{cases} \tag{5}$$

The *duration* is simply the time difference between consecutive points:

$$duration = \Delta t \tag{6}$$

The *velocity* is a change of *distance* over *duration*. When considering individual axes, only the distance in that particular axis comes into the equation for *velocity_x* and *velocity_y*:

$$velocity = \frac{distance}{duration} \tag{7}$$

$$velocity\_x = \frac{\Delta x}{duration} \tag{8}$$

$$velocity\_y = \frac{\Delta y}{duration} \tag{9}$$

The *velocity_smooth* is a change in *distance* calculated from smoothed coordinates with Savitzky–Golay filter over *duration*:

$$velocity\_smooth = \frac{\sqrt{(\Delta x_{savgol})^2 + (\Delta y_{savgol})^2}}{duration} \tag{10}$$

The *pace* is a change of *duration* over the *distance*:

$$pace = \frac{duration}{distance} \tag{11}$$

Table 8
Overview of research utilizing mouse dynamics. The majority of experiment design is between-subject, which is more vulnerable to differences in mouse configuration. Normalization approaches that may mitigate the effects of mouse configuration have been tackled, but are rare.

Source	Area of research	Experiment design	Environment	Mouse configuration	User differences normalization
Sulikowski et al. [14]	E-commerce recommendations	Between-subject	Remote uncontrolled	Unspecified	None
Hucko et al. [15]	Confusion detection	Between-subject	Laboratory controlled	Unspecified	None
De Santana et al. [42]	XAI interaction prediction	Between-subject	Remote uncontrolled	Unspecified	None
Chen et al. [43]	Search engine satisfaction prediction	Between-subject	Laboratory controlled	Unspecified	None
Khan et al. [35]	Personality traits measurement	Between-subject	Remote uncontrolled	Unspecified	None
Pentel [7]	Age and gender prediction	Between-subject	Remote uncontrolled	Unspecified	None
Fernandez-Lanvin et al. [6]	Age and gender personalization	Between-subject	Remote uncontrolled	Unspecified	None
Yamauchi et al. [8]	Gender differences	Between-subject	Laboratory controlled	Dell 0C8639 USB 2 Button Scrollwheel Optical Mouse, pointer speed set to medium	None
Kratky and Chuda [21]	Gender classification	Between-subject	Remote uncontrolled	Unspecified	None
Balen et al. [9]	Gender classification	Between-subject	Laboratory controlled	Unspecified	Not of individual differences, based on target parameters (e.g,, distance, size)
Almanji et al. [44]	Effects of cerebral palsy impairment	Between-subject	Laboratory controlled	Lenovo M/N: LXH-MOAFUO USB - PN 25007694	Not of individual differences, spatially normalized measures
Chong et al. [31]	User authentification	Between-subject	Remote uncontrolled	Unspecified	Not of individual differences, standardization to 0-1 interval
Hu et al. [32]	User authentification	Between-subject	Remote uncontrolled	Unspecified	None
Siddiqui et al. [10]	User authentification	Between-subject	Laboratory controlled	Dell MS116 USB Optical Mouse	None
Zheng et al. [27]	User authentification	Between-subject	Laboratory controlled and remote uncontrolled	Unspecified	None
Antal and Egyed-Zsigmond [11]	Intrusion detection	Between-subject	Remote uncontrolled	Unspecified	None
Chu et al. [12]	Bot detection	Between-subject	Remote uncontrolled	Unspecified	None
Antal et al. [45]	Dataset for user authentification	Between-subject	Remote uncontrolled	Unspecified	None
Antal et al. [46]	Creating human like trajectories	Between-subject	Remote uncontrolled	Unspecified	None
Shen et al. [33]	User authentification	Between-subject	Remote uncontrolled	Unspecified	None
Martín-Albo et al. [47]	User intent detection	Between-subject	Remote uncontrolled	Unspecified	None
Kuric et al. [1]	Repeat purchase prediction	Between-subject	Remote uncontrolled	Unspecified	None
Monaro et al. [2]	Fake response detection	Between-subject	Laboratory controlled	Unspecified	Not of individual differences, linear interpolation of trajectory points (spatial normalization)
Leiva and Huang [48]	Mouse-dynamics data compression	Between-subject	Laboratory controlled and remote uncontrolled	Unspecified	None
Arapakis and Leiva [30]	User attention prediction	Between-subject	Remote uncontrolled	Unspecified	Not of individual differences, interpolation by the size of viewport
Seelye et al. [34]	Cognitive impairment prediction	Between-subject	Remote uncontrolled	Unspecified	None
Maldonado et al. [25]	Cognitive processes prediction	Between-subject	Remote uncontrolled	Unspecified	None
Grage et al. [19]	Mouse-tracking design	Between-subject	Laboratory controlled	Logitech RX1500, sampling rate 92 Hz, resolution 1000 dpi	None
Fernández-Fontelo et al. [4]	Survey question difficulty prediction	Between-subject	Remote uncontrolled	Unspecified	Personalization (normalization) method that adjusts for respondents' baseline mouse behavior observed while answering questions without manipulated higher difficulty.
Leiva et al. [29]	Web browsing privacy	Between-subject and within-subject	Remote uncontrolled	Unspecified	None
Shen et al. [33]	User authentification	Between-subject and within-subject	Remote uncontrolled	Unspecified	None
Cai et al. [13]	Mouse feature variability assessment	Between-subject and within-subject	Laboratory controlled	USB HP optical mouse, further unspecified	Normalization and feature reduction to mitigate noise, improving performance in authentication task. Unknown whether the approach may effectively normalize mouse configuration differences in addition to between-session noise — lab experiment utilizes further unspecified HP optical mouse.
Gardey et al. [5]	Interaction effort prediction	Within-subject	Remote uncontrolled	Unspecified	None
Pepa et al. [23]	Stress detection	Within-subject	Remote uncontrolled	Unspecified	None
Guo and Agichtein [3]	Web searcher goals prediction	Within-subject	Remote uncontrolled	Unspecified	None
Freihaut and Göritz [22]	Stress measurement	Within-subject	Laboratory controlled	Logitech B100, optical USB mouse with 800 dpi	Not of individual differences, linear interpolation of trajectory points (spatial normalization)
Wilson et al. [49]	Plagiarism detection	Within-subject	Remote controlled	Unspecified	Z-score normalization method between samples where participant answers follow-up questions about an essay they authored and an AI-generated essay.

Table 9
Statistical tests results for differences between mouse configurations (Sensitivity and CPI - Friedman test) and sexes (Mann Whitney test) during individual activities.

Feature group	Feature name	Points clicking activity			Website prototype activity		
		Sensitivity $\chi^2(2, N = 480)$	CPI $\chi^2(2, N = 480)$	Sex $z(N = 900)$	Sensitivity $\chi^2(2, N = 480)$	CPI $\chi^2(2, N = 480)$	Sex $z(N = 900)$
Acceleration	*acceleration_max*	53.58***	33.08***	1.37	18.81***	11.44**	−0.6
	acceleration_mean	107.03***	67.41***	1.19	16.19***	34.31***	−0.67
	acceleration_mean_weighted	1.96	0.01	0.54	1.31	0.56	−0.7
	acceleration_median	9.06*	55.31***	0.7	7.6*	29.11***	2.02*
	acceleration_min	27.23***	7.54*	−2.59*	21.0***	12.25**	−0.53
	acceleration_q5	35.73***	174.98***	−5.02***	37.56***	25.0***	0.86
	acceleration_q95	88.04***	128.81***	3.79***	36.19***	31.94***	−0.84
	acceleration_std	83.81***	63.74***	3.14**	31.69***	17.06***	−0.34
Negative acceleration	*acceleration_negative_max*	4.05	5.46	−0.86	1.69	9.81**	0.13
	acceleration_negative_mean	85.28***	138.01***	−4.6***	34.56***	16.19***	0.34
	acceleration_negative_mean_weighted	4.74	149.21***	−5.46***	13.19**	7.75*	−0.13
	acceleration_negative_median	115.14***	207.74***	−1.38	28.94***	31.69***	0.8
	acceleration_negative_min	27.23***	7.54*	−2.59*	21.0***	12.25**	−0.53
	acceleration_negative_q5	23.41***	86.93***	−5.15***	19.0***	12.44**	0.27
	acceleration_negative_q95	5.68	34.88***	0.17	5.69	8.06*	0.38
	acceleration_negative_std	31.51***	24.58***	4.24***	25.0***	13.19**	0.06
Positive acceleration	*acceleration_positive_max*	53.58***	33.08***	1.37	18.81***	11.44**	−0.6
	acceleration_positive_mean	133.3***	178.01***	3.3***	33.06***	32.25***	−0.84
	acceleration_positive_mean_weighted	97.34***	107.88***	3.62***	25.19***	10.75**	0.08
	acceleration_positive_median	150.81***	208.21***	3.01**	30.44***	28.31***	−0.99
	acceleration_positive_min	0.11	4.9	0.28	6.44*	6.72*	1.25
	acceleration_positive_q5	10.68**	26.21***	−1.25	13.0**	4.56	−0.47
	acceleration_positive_q95	79.54***	118.18***	3.77***	35.69***	33.25***	−1.45
	acceleration_positive_std	72.74***	75.03***	2.47*	20.44***	15.25***	−0.5
Acceleration X	*acceleration_x_max*	71.71***	27.41***	1.72	17.31***	4	0.43
	acceleration_x_mean	112.3***	80.48***	1.11	13.19**	22.94***	−0.26
	acceleration_x_mean_weighted	5.13	1.82	−0.73	0.5	1.76	1.67
	acceleration_x_min	28.54***	2.47	−3.66***	16.75***	9.19*	−1.37
	acceleration_x_q5	19.14***	161.81***	−5.09***	24.19***	23.31***	0.46
	acceleration_x_q95	89.21***	136.9***	4.31***	30.44***	23.69***	−0.45
	acceleration_x_std	81.23***	74.01***	4.16***	26.31***	10.56**	0.22
Acceleration along Y axis	*acceleration_y_max*	29.01***	38.81***	0.35	10.69**	14.31***	−1.19
	acceleration_y_mean	88.08***	51.41***	0.62	19.19***	31.94***	−0.64
	acceleration_y_mean_weighted	4.87	4.65	−0.89	2.25	0.45	2.07*
	acceleration_y_min	11.57**	25.83***	−1.12	11.44**	8.69*	1.89
	acceleration_y_q5	57.1***	150.83***	−3.41***	36.19***	18.06***	0.87
	acceleration_y_q95	93.23***	131.21***	2.42*	33.81***	21.81***	−0.92
	acceleration_y_std	67.51***	80.01***	1.31	21.44***	21.44***	−0.94
Angle	*angle_mean*	11.51**	7.08*	2.03*	1.31	0.81	1.61
	angle_median	3.61	6.22*	0.64	1.52	0.78	1.22
	angle_min	0.22	3.56	2.39*	2.87	3.83	0.18
	angle_q5	3.94	1.94	0.36	4.87	2.32	0.19
	angle_q95	6.23*	24.72***	2.08**	2.11	2.35	0.75
	angle_std	10.54**	3.48	3.05**	9.81**	5.69	0.86
Corner	*corner_count*	39.63***	42.33***	3.16**	3.65	2.05	1.42
Curvature	*curvature_max*	2.01	0.3	2.39*	1.86	1.08	2.51*
	curvature_mean	11.91**	5.54	5.39***	2.69	1.75	4.54***
	curvature_min	1.54	32.73***	−1.5	2	1.73	1.5
	curvature_q5	1.98	13.75**	−3.1**	2.44	3.3	−2.03*
	curvature_q95	3.55	36.24***	5.44***	7.63*	3.42	3.1**
	curvature_std	0.71	38.21***	6.01***	3.56	3.94	1.55
Deviation	*deviation_max*	24.84***	45.48***	0.07	5.81	13.0**	−1.05
	deviation_mean	14.21***	12.74**	−0.42	7.94*	2.44	−1.88
	deviation_median	2.41	4.41	0.15	1.69	0.56	−1.96
	deviation_q5	10.83**	1.23	−2.78**	0.06	0.81	−2.08*
	deviation_q95	16.14***	20.54***	0.2	3.81	7.75*	−1.45
	deviation_std	11.88**	42.43***	0.31	5.69	13.19**	−0.85
Distance	*distance_max*	89.64***	48.34***	1.16	26.06***	25.75***	−0.29
	distance_mean	67.9***	60.81***	2.16*	23.69***	17.31***	−1.48
	distance_median	76.45***	109.0***	−2.25*	19.9***	7.82*	−2.2*
	distance_q5	6.8*	39.69***	−2.22*	4.67	4	−0.13
	distance_q95	10.84**	67.54***	3.47***	10.56**	18.81***	−1.35
	distance_ratio	38.43***	96.4***	−1.1	10.69**	7.94*	1.25
	distance_std	70.93***	74.34***	3.29***	25.75***	30.06***	−0.64
	distance_sum	37.74***	84.63***	1	16.94***	7.0*	−1.49

(*continued on next page*)

The *acceleration* is a change of *velocity* over *duration*. Similarly to *velocity*, *acceleration* is also separately calculated for x and y axes as *acceleration_x* and *acceleration_y*. Additionally, positive *acceleration_pos* and negative *acceleration_neg* acceleration are reported separately:

$$acceleration = \frac{\Delta velocity}{duration} \tag{12}$$

$$acceleration\_x = \frac{\Delta velocity\_x}{duration} \tag{13}$$

$$acceleration\_y = \frac{\Delta velocity\_y}{duration} \tag{14}$$

$$acceleration\_pos = acceleration;\ \{acceleration > 0\} \tag{15}$$

$$acceleration\_neg = acceleration;\ \{acceleration < 0\} \tag{16}$$

The *angle* is the current angle of mouse movement represented as inverse of tangent of coordinate change proportion:

$$angle = tan^{-1}\frac{\Delta y}{\Delta x} \tag{17}$$

The *velocity_angular* is the change of *angle* over *duration*:

$$velocity\_angular = \frac{\Delta angle}{duration} \tag{18}$$

Table 9 (*continued*).

Feature group	Feature name	Points clicking activity			Website prototype activity		
		Sensitivity $\chi^2(2, N = 480)$	CPI $\chi^2(2, N = 480)$	Sex $z(N = 900)$	Sensitivity $\chi^2(2, N = 480)$	CPI $\chi^2(2, N = 480)$	Sex $z(N = 900)$
Duration	*duration_sum*	77.68***	169.88***	−3.44***	1.31	3.81	−1.05
Flip	*flip_x_count*	42.02***	10.06**	2.36*	0.34	1.46	−0.32
	flip_y_count	26.0***	49.26***	0.86	1.78	3.32	0.35
Pace	*pace_max*	15.22***	38.49***	−1.43	3	22.75***	−1.15
	pace_mean	7.54*	75.63***	0.76	7.75*	22.94***	−0.69
	pace_mean_weighted	22.8***	28.31***	−0.82	8.06*	15.25***	−1.31
	pace_median	103.08***	112.73***	0.82	17.31***	13.65**	2.28*
	pace_q5	18.4***	40.62***	−2.98**	6.92*	20.69***	0.21
	pace_q95	49.11***	20.43***	2.04*	3.42	0.19	1.28
	pace_std	40.13***	43.01***	−1.17	4.75	19.0***	−1.23
Angular velocity	*velocity_angular_max*	5.24	9.64**	−0.12	1.19	1.73	0.32
	velocity_angular_mean	4.48	0.9	−0.68	0.44	0.25	−2.4*
	velocity_angular_mean_weighted	1.2	5.45	1.05	3.25	0.56	−0.75
	velocity_angular_min	6.88*	6.79*	−1.04	1.94	2.44	−0.7
	velocity_angular_q5	59.12***	8.33*	−4.5***	6.75*	12.43**	−2.37*
	velocity_angular_q95	44.39***	13.35**	4.8***	20.46***	12.11**	2.17*
	velocity_angular_std	16.31***	12.21**	3.04**	1.19	13.56**	2.09*
Velocity	*velocity_max*	115.88***	33.41***	2.55*	30.06***	21.81***	0.7
	velocity_mean	59.34***	57.34***	2.26*	23.25***	16.19***	−1.39
	velocity_mean_weighted	2.8	0.23	4.3***	13.56**	16.19***	−0.48
	velocity_median	76.14***	125.35***	−2.2*	18.0***	8.6*	−2.31*
	velocity_q5	20.18***	45.69***	−2.32*	0.44	3.06	−0.1
	velocity_q95	3.74	56.14***	3.41***	12.56**	19.75***	−1.44
	velocity_std	60.3***	64.63***	3.93***	21.81***	27.44***	−0.29
Smooth velocity	*velocity_smooth_max*	23.01***	30.71***	1.11	33.25***	23.69***	0.34
	velocity_smooth_mean	40.68***	46.71***	2.45*	20.44***	24.25***	−1.55
	velocity_smooth_mean_weighted	12.14**	2.34	4.57***	14.25***	15.25***	−0.4
	velocity_smooth_median	110.53***	110.81***	−0.46	19.0***	12.25**	−2.05*
	velocity_smooth_q5	3.34	75.21***	1.53	9.0*	8.69*	−1.83
	velocity_smooth_q95	11.08**	57.41***	3.31***	9.44**	20.31***	−1.38
	velocity_smooth_std	24.7***	55.21***	3.31***	25.0***	28.56***	−0.79
Velocity along X axis	*velocity_x_max*	94.02***	27.88***	2.42*	20.44***	13.56**	1.33
	velocity_x_mean	30.53***	30.18***	2.81**	15.75***	14.81***	−0.98
	velocity_x_mean_weighted	2.23	0.83	4.4***	11.31**	12.25**	−0.26
	velocity_x_median	36.17***	35.3***	−2.81**	5.52	3.18	−1.77
	velocity_x_q95	2.41	24.23***	4.09***	12.25**	21.87***	−0.73
	velocity_x_std	43.68***	38.81***	4.58***	20.25***	28.0***	0.2
Velocity along Y axis	*velocity_y_max*	32.41***	48.87***	−0.01	12.25**	15.44***	−1.44
	velocity_y_mean	46.43***	44.88***	1.37	19.75***	21.44***	−1.88
	velocity_y_mean_weighted	1.9	3.6	3.01**	11.44**	17.06***	−1.03
	velocity_y_median	34.46***	63.67***	−0.91	15.71***	12.35**	−2.25*
	velocity_y_q95	17.82***	38.53***	1.8	14.6***	14.33***	−1.41
	velocity_y_std	39.41***	64.81***	1.28	20.31***	18.25***	−1.78

*** $p < .001$.
** $p < .01$.
* $p < .05$.

The *curvature* is the change of *angle* over *distance*:

$$curvature = \frac{\Delta angle}{distance} \tag{19}$$

The *deviation* is the current perpendicular distance to the ideal path, assuming (x_s, y_s) and (x_e, y_e) are the ideal path starting and ending points:

$$deviation = \left| \frac{(x_e - x_s, y_e - y_s) \times (x - x_s, y - x_s)}{\sqrt{(x_e - x_s)^2 + (y_e - y_s)^2}} \right| \tag{20}$$

The *straightness* is a ratio of ideal path length to the total distance, assuming (x_s, y_s) and (x_e, y_e) are the ideal path starting and ending points. It is calculated for the whole movement at once as:

$$straightness = \frac{\sqrt{(x_e - x_s)^2 + (y_e - y_s)^2}}{\sum_{x,y} distance} \tag{21}$$

The corner is a binary variable representing whether the current *angle* of a movement is more or less than 90 degrees.

$$corner = \begin{cases} 1 & \Delta angle > \frac{\pi}{2} \\ 0 & \Delta angle \leq \frac{\pi}{2} \end{cases} \tag{22}$$

Appendix C. Statistical tests

Statistical tests results for differences between mouse configurations and sexes can be found in Table 9.

References

[1] E. Kuric, A. Puskas, P. Demcak, D. Mensatorisova, Effect of low-level interaction data in repeat purchase prediction task, Int. J. Hum.-Comput. Interact. (2023) 1–19, http://dx.doi.org/10.1080/10447318.2023.2175973.

[2] M. Monaro, C. Mazza, M. Colasanti, S. Ferracuti, G. Orrù, A. di Domenico, G. Sartori, P. Roma, Detecting faking-good response style in personality questionnaires with four choice alternatives, Psychol. Res. 85 (8) (2021) 3094–3107, http://dx.doi.org/10.1007/s00426-020-01473-3.

[3] Q. Guo, E. Agichtein, 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, 2010, pp. 130–137, http://dx.doi.org/10.1145/1835449.1835473.

[4] A. Fernández-Fontelo, P.J. Kieslich, F. Henninger, F. Kreuter, S. Greven, Predicting question difficulty in web surveys: A machine learning approach based on mouse movement features, Soc. Sci. Comput. Rev. 41 (1) (2023) 141–162, http://dx.doi.org/10.1177/08944393211032950.

[5] J.C. Gardey, J. Grigera, A. Rodríguez, G. Rossi, A. Garrido, Predicting interaction effort in web interface widgets, Int. J. Hum.-Comput. Stud. 168 (2022) 102919, http://dx.doi.org/10.1016/j.ijhcs.2022.102919.

[6] D. Fernandez-Lanvin, J. de Andres-Suarez, M. Gonzalez-Rodriguez, B. Pariente-Martinez, The dimension of age and gender as user model demographic factors for automatic personalization in e-commerce sites, Comput. Stand. Interfaces 59 (2018) 1–9, http://dx.doi.org/10.1016/j.csi.2018.02.001.

[7] A. Pentel, Predicting age and gender by keystroke dynamics and mouse patterns, in: Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization, UMAP '17, Association for Computing Machinery, New York, NY, USA, 2017, pp. 381–385, http://dx.doi.org/10.1145/3099023.3099105.

[8] T. Yamauchi, J.H. Seo, N. Jett, G. Parks, C. Bowman, Gender differences in mouse and cursor movements, Int. J. Hum.-Comput. Interact. 31 (12) (2015) 911–921, http://dx.doi.org/10.1080/10447318.2015.1072787.

[9] N.V. Balen, C. Ball, H. Wang, Analysis of targeted mouse movements for gender classification, EAI Endorsed Trans. Secur. Saf. 4 (11) (2017) http://dx.doi.org/10.4108/eai.7-12-2017.153395.

[10] N. Siddiqui, R. Dave, N. Seliya, Continuous user authentication using mouse dynamics, machine learning, and minecraft, in: 2021 International Conference on Electrical, Computer and Energy Technologies, ICECET, 2021, pp. 1–6, http://dx.doi.org/10.1109/ICECET52533.2021.9698532.

[11] M. Antal, E. Egyed-Zsigmond, Intrusion detection using mouse dynamics, IET Biometrics 8 (5) (2019) 285–294, http://dx.doi.org/10.1049/iet-bmt.2018.5126.

[12] Z. Chu, S. Gianvecchio, H. Wang, Bot or human? A behavior-based online bot detection system, in: From Database To Cyber Security: Essays Dedicated To Sushil Jajodia on the Occasion of His 70th Birthday, Springer International Publishing, Cham, 2018, pp. 432–449, http://dx.doi.org/10.1007/978-3-030-04834-1_21.

[13] Z. Cai, C. Shen, X. Guan, Mitigating behavioral variability for mouse dynamics: A dimensionality-reduction-based approach, IEEE Trans. Hum.-Mach. Syst. 44 (2) (2014) 244–255, http://dx.doi.org/10.1109/THMS.2014.2302371.

[14] P. Sulikowski, T. Zdziebko, O. Hussain, A. Wilbik, Fuzzy approach to purchase intent modeling based on user tracking for E-commerce recommenders, in: 2021 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE, 2021, pp. 1–8, http://dx.doi.org/10.1109/FUZZ45933.2021.9494585.

[15] M. Hucko, L. Gazo, P. Simun, M. Valky, R. Moro, J. Simko, M. Bielikova, YesElf: Personalized onboarding for web applications, in: Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization, in: UMAP'19 Adjunct, Association for Computing Machinery, New York, NY, USA, 2019, pp. 39–44, http://dx.doi.org/10.1145/3314183.3324978.

[16] M. Schoemann, D. O'Hora, R. Dale, S. Scherbaum, Using mouse cursor tracking to investigate online cognition: Preserving methodological ingenuity while moving toward reproducible science, Psychon. Bull. Rev. 28 (3) (2021) 766–787, http://dx.doi.org/10.3758/s13423-020-01851-3.

[17] P.J. Kieslich, M. Schoemann, T. Grage, J. Hepp, S. Scherbaum, Design factors in mouse-tracking: What makes a difference? Behav. Res. Methods 52 (1) (2020) 317–341, http://dx.doi.org/10.3758/s13428-019-01228-y.

[18] J.B. Freeman, N. Ambady, MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method, Behav. Res. Methods 42 (1) (2010) 226–241, http://dx.doi.org/10.3758/BRM.42.1.226.

[19] T. Grage, M. Schoemann, P.J. Kieslich, S. Scherbaum, Lost to translation: How design factors of the mouse-tracking procedure impact the inference from action to cognition, Attent. Percept. Psychophys. 81 (7) (2019) 2538–2557, http://dx.doi.org/10.3758/s13414-019-01889-z.

[20] B. Boudaoud, J. Spjut, J. Kim, Mouse sensitivity in first-person targeting tasks, in: 2022 IEEE Conference on Games, CoG, 2022, pp. 183–190, http://dx.doi.org/10.1109/CoG51982.2022.9893626.

[21] P. Kratky, D. Chuda, Estimating gender and age of web page visitors from the way they use their mouse, in: Proceedings of the 25th International Conference Companion on World Wide Web, in: WWW '16 Companion, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 2016, pp. 61–62, http://dx.doi.org/10.1145/2872518.2889384.

[22] P. Freihaut, A.S. Göritz, Using the computer mouse for stress measurement – An empirical investigation and critical review, Int. J. Hum.-Comput. Stud. 145 (2021) 102520, http://dx.doi.org/10.1016/j.ijhcs.2020.102520.

[23] L. Pepa, A. Sabatelli, L. Ciabattoni, A. Monteriù, F. Lamberti, L. Morra, Stress detection in computer users from keyboard and mouse dynamics, IEEE Trans. Consum. Electron. 67 (1) (2021) 12–19, http://dx.doi.org/10.1109/TCE.2020.3045228.

[24] A. Pentel, Patterns of confusion: Using mouse logs to predict user's emotional state, in: UMAP Workshops, 2015, pp. 1–6.

[25] M. Maldonado, E. Dunbar, E. Chemla, Mouse tracking as a window into decision making, Behav. Res. Methods 51 (3) (2019) 1085–1101, http://dx.doi.org/10.3758/s13428-018-01194-x.

[26] C. Shen, Z. Cai, X. Guan, Y. Du, R.A. Maxion, User authentication through mouse dynamics, IEEE Trans. Inf. Forensics Secur. 8 (1) (2013) 16–30, http://dx.doi.org/10.1109/TIFS.2012.2223677.

[27] N. Zheng, A. Paloski, H. Wang, An efficient user verification system via mouse movements, in: Proceedings of the 18th ACM Conference on Computer and Communications Security, CCS '11, Association for Computing Machinery, New York, NY, USA, 2011, pp. 139–150, http://dx.doi.org/10.1145/2046707.2046725.

[28] D. Wulff, J. Haslbeck, P. Kieslich, F. Henninger, M. Schulte-Mecklenbeck, Mouse-tracking: Detecting types in movement trajectories, 2019, pp. 131–145, http://dx.doi.org/10.31234/osf.io/6edca.

[29] L.A. Leiva, I. Arapakis, C. Iordanou, My mouse, my rules: Privacy issues of behavioral user profiling via mouse tracking, in: Proceedings of the 2021 Conference on Human Information Interaction and Retrieval, CHIIR '21, Association for Computing Machinery, New York, NY, USA, 2021, pp. 51–61, http://dx.doi.org/10.1145/3406522.3446011.

[30] I. Arapakis, L.A. Leiva, Learning efficient representations of mouse movements to predict user attention, in: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 1309–1318, http://dx.doi.org/10.1145/3397271.3401031.

[31] P. Chong, Y. Elovici, A. Binder, User authentication based on mouse dynamics using deep neural networks: A comprehensive study, IEEE Trans. Inf. Forensics Secur. 15 (2020) 1086–1101, http://dx.doi.org/10.1109/TIFS.2019.2930429.

[32] T. Hu, W. Niu, X. Zhang, X. Liu, J. Lu, Y. Liu, An insider threat detection approach based on mouse dynamics and deep learning, Secur. Commun. Netw. 2019 (2019) 3898951, http://dx.doi.org/10.1155/2019/3898951.

[33] C. Shen, Z. Cai, X. Guan, H. Sha, J. Du, Feature analysis of mouse dynamics in identity authentication and monitoring, in: 2009 IEEE International Conference on Communications, 2009, pp. 1–5, http://dx.doi.org/10.1109/ICC.2009.5199032.

[34] A. Seelye, S. Hagler, N. Mattek, D.B. Howieson, K. Wild, H.H. Dodge, J.A. Kaye, Computer mouse movement patterns: A potential marker of mild cognitive impairment, Alzheimer's Dementia Diagn. Assess. Dis. Monit. 1 (4) (2015) 472–480, http://dx.doi.org/10.1016/j.dadm.2015.09.006.

[35] I.A. Khan, W.-P. Brinkman, N. Fine, R.M. Hierons, Measuring personality from keyboard and mouse use, in: Proceedings of the 15th European Conference on Cognitive Ergonomics: The Ergonomics of Cool Interaction, ECCE '08, Association for Computing Machinery, New York, NY, USA, 2008, pp. 1–8, http://dx.doi.org/10.1145/1473018.1473066.

[36] S. Dara, C.R. Chowdary, C. Kumar, A survey on group recommender systems, J. Intell. Inf. Syst. 54 (2) (2020) 271–295, http://dx.doi.org/10.1007/s10844-018-0542-3.

[37] W.W. Moe, P.S. Fader, Dynamic conversion behavior at E-commerce sites, Manage. Sci. 50 (3) (2004) 326–335, http://dx.doi.org/10.1287/mnsc.1040.0153.

[38] K.B. Subramanya, A. Somani, Enhanced feature mining and classifier models to predict customer churn for an E-retailer, in: 2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence, 2017, pp. 531–536, http://dx.doi.org/10.1109/CONFLUENCE.2017.7943208.

[39] B. Zhao, A. Takasu, R. Yahyapour, X. Fu, Loyal consumers or one-time deal hunters: Repeat buyer prediction for E-commerce, in: 2019 International Conference on Data Mining Workshops, ICDMW, 2019, pp. 1080–1087, http://dx.doi.org/10.1109/ICDMW.2019.00158.

[40] J. Liu, Deconstructing search tasks in interactive information retrieval: A systematic review of task dimensions and predictors, Inf. Process. Manage. 58 (3) (2021) 102522, http://dx.doi.org/10.1016/j.ipm.2021.102522.

[41] E. Kuric, P. Demcak, M. Krajcovic, G. Nguyen, Cognitive abilities and visual complexity impact first impressions in five-second testing, Behav. Inf. Technol. (2023) http://dx.doi.org/10.1080/0144929X.2023.2272747.

[42] V.F. De Santana, A. Fucs, V. Segura, D.B. de Moraes, R. Cerqueira, Predicting the need for XAI from high-granularity interaction data, Int. J. Hum.-Comput. Stud. 175 (2023) 103029, http://dx.doi.org/10.1016/j.ijhcs.2023.103029.

[43] L. Chen, A. Fan, H. Shi, G. Chen, Region-action LSTM for mouse interaction sequence based search satisfaction evaluation, Inf. Process. Manag. 57 (6) (2020) 102349, http://dx.doi.org/10.1016/j.ipm.2020.102349.

[44] A. Almanji, T. Claire Davies, N. Susan Stott, Using cursor measures to investigate the effects of impairment severity on cursor control for youths with cerebral palsy, Int. J. Hum.-Comput. Stud. 72 (3) (2014) 349–357, http://dx.doi.org/10.1016/j.ijhcs.2013.12.003.

[45] M. Antal, N. Fejér, K. Buza, SapiMouse: Mouse dynamics-based user authentication using deep feature learning, in: 2021 IEEE 15th International Symposium on Applied Computational Intelligence and Informatics, SACI, 2021, pp. 61–66, http://dx.doi.org/10.1109/SACI51354.2021.9465583.

[46] M. Antal, K. Buza, N. Fejer, SapiAgent: A bot based on deep learning to generate human-like mouse trajectories, IEEE Access 9 (2021) 124396–124408, http://dx.doi.org/10.1109/ACCESS.2021.3111098.

[47] D. Martín-Albo, L.A. Leiva, J. Huang, R. Plamondon, Strokes of insight: User intent detection and kinematic compression of mouse cursor trails, Inf. Process. Manag. 52 (6) (2016) 989–1003, http://dx.doi.org/10.1016/j.ipm.2016.04.005.

[48] L.A. Leiva, J. Huang, Building a better mousetrap: Compressing mouse cursor activity for web analytics, Inf. Process. Manag. 51 (2) (2015) 114–129, http://dx.doi.org/10.1016/j.ipm.2014.10.005.

[49] D. Wilson, P. Burnett, J.S. Valacich, J. Jenkins, Human or ai? Using digital behavior to verify essay authorship, in: ICIS 2023 Proceedings, 2023, p. 6.