ORIGINAL ARTICLE

WILEY

Mouse movements reflect personality traits and task attentiveness in online experiments

Kimberly L. Meidenbauer[1] | Tianyue Niu[1] | Kyoung Whan Choe[1,2] | Andrew J. Stier[1] | Marc G. Berman[1,2,3]

[1]Environmental Neuroscience Lab, Department of Psychology, The University of Chicago, Chicago, Illinois, USA

[2]Mansueto Institute for Urban Innovation, The University of Chicago, Chicago, Illinois, USA

[3]The Neuroscience Institute, The University of Chicago, Chicago, Illinois, USA

Correspondence
Kimberly L. Meidenbauer and Marc G. Berman, Environmental Neuroscience Lab, Department of Psychology, The University of Chicago, 5848 S University Avenue, Chicago, IL 60637, USA.
Email: meidenbauer@uchicago.edu and bermanm@uchicago.edu

Abstract

Objective: In this rapidly digitizing world, it is becoming ever more important to understand people's online behaviors in both scientific and consumer research settings. The current work tests the feasibility of inferring personality traits from mouse movement patterns as a cost-effective means of measuring individual characteristics.

Method: Mouse movement features (i.e., pauses, fixations, speed, and clicks) were collected while participants ($N = 791$) completed an online image choice task. We compare the results of standard univariate and three forms of multivariate partial least squares (PLS) analyses predicting Big Five traits from mouse movements. We also examine whether mouse movements can predict a proposed measure of task attentiveness (atypical responding), and how these might be related to personality traits.

Results: Each of the PLS analyses showed significant associations between a linear combination of personality traits (high Conscientiousness, Agreeableness, Openness, and low Neuroticism) and several mouse movements associated with slower, more deliberate responding (less unnecessary clicks and more fixations). Additionally, several click-related mouse features were associated with atypical responding on the task.

Conclusions: As the image choice task itself is not intended to assess personality in any way, our results validate the feasibility of using mouse movements to infer internal traits across experimental contexts.

KEYWORDS
Big Five, mouse movements, multiverse, personality, PLS

Kimberly L. Meidenbauer and Tianyue Niu contributed equally to this study.

1 | INTRODUCTION

The rapid development of digital technology has led people to spend an increasing amount of time online. As people navigate through different web applications, they leave digital traces behind. One of these implicit traces is mouse movements. Regardless of the motivation, people use their mouse to guide and shape their online experiences. As humans naturally infer internal states from physical motion cues (Koppensteiner, 2013), a similar process can be used to infer an individual's internal states based on online behavioral cues. This research attempts to examine a related topic on the relation between an individual's online behaviors and personality traits. More specifically, the research question of interest is "Are mouse movement patterns exhibited in a choice-making task reflective of a person's internal states and traits?"

Mouse cursor movements are a cost-effective measurement of individuals' behaviors. Mouse trajectories have been used in previous research to track attention in computer interactions (Rodden et al., 2008) and measure website engagement (Arapakis & Leiva, 2016). For example, in a study by Arapakis et al. (2014), several speed- and distance-based mouse movement features showed significant correlations (coefficients between 0.37 and 0.4, $N = 22$) with participant ratings of how interesting they found a news article that they were reading. It has also been shown that mouse movements are significantly correlated with individuals' attitudes. In the context of evaluating implicit bias, researchers have examined the time and trajectory from one location on screen to another to evaluate hesitancy or "corrections" to initial decisions (Freeman, 2018; Hehman et al., 2015; Stolier & Freeman, 2016). Work by Tzafilkou et al. (2014) showed that mouse hover time and movement patterns can be used to predict an individual's self-efficacy and risk perception. The same researchers found that a person's attitude toward a web-based tool, such as perceived usefulness or perceived ease of use of that tool, can also be inferred from their mouse movements (Tzafilkou & Nicolaos, 2018). This research highlights the potential of mouse movement to effectively predict internal states and attributes. In this study, we specifically explore the relationship between participants' mouse movement patterns and two types of individual attributes—attentiveness to the task at hand, measured by the deviation of individual responses from random responding, and personality, as measured by the Big Five Inventory.

The Big Five Inventory (BFI) is a self-reported survey designed to measure one's personality across five dimensions: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (John & Srivastava, 1999). Each of these five dimensions consists of several subtraits.

For example, Neuroticism consists of traits such as anxiety, depression, self-consciousness, and vulnerability. Researchers have shown that personality traits are related to a wide range of behaviors. Of key relevance to the current work, it has been found that one's behaviors online are reflective of a person's offline personality (Orchard & Fullwood, 2010) outside of the computer screen.

Combining personality traits with mouse movements, studies have found that extroverts tend to exhibit higher levels of motor activity (mouse clicking) at a higher frequency in a given task (Brebner, 1983; Khan et al., 2008), and that keystrokes and mouse clicking behaviors are significantly correlated with Big Five personality traits (Khan et al., 2008). In particular, it was found by Khan et al. (2008) that the average number of mouse clicks was positively correlated with a subtrait of Conscientiousness ($r = 0.52$) and negatively correlated with a subtrait of Neuroticism ($r = -0.40$). It is worth noting that this study's small sample size ($N = 20$) was underpowered to detect a correlation of this magnitude, and therefore, the effect sizes may be overestimated (Kühberger et al., 2014; Schönbrodt & Perugini, 2013). It does provide initial evidence that personality traits can be predicted from mouse movement patterns. However, both personality traits and mouse movements can show high levels of intercorrelation, and it is therefore likely that multivariate methods which identify a combination of mouse features and personality traits would create a better understanding of this link.

Across academic- and consumer-focused research, participants' inattentiveness in online research can significantly damage the validity of a study. Different attempts have been made to filter out non-compliant responses such as setting a higher standard to select participants with good records, embedding attention check questions within a survey, or evaluating personality inventories and flagging abnormalities (Barends & de Vries, 2019). In addition to these methods, it has been found in previous research that in certain tasks involving image recognition, mouse click attention tracking can provide highly valid results that are more consistent than eye movement attention tracking (Egner et al., 2018). As a secondary aim, the current study proposes the use of atypical responding on the image rating task (deviation from the group average) as a measure of general inattentiveness and random responding as the stimuli in the task show very high inter-rater reliability ($\rho = 0.86$).

Importantly, (in)attentiveness during online research can also be reflective of a participant's personality (i.e., those higher on the BFI trait of Conscientiousness tend to show greater compliance in experiments; Berry et al., 2019; Meade & Pappalardo, 2013). Therefore, instead of assuming independence between mouse movement features,

attentiveness and personality, we conduct multivariate partial least squares (PLS) analysis to further explore the underlying relationships between these features. Additionally, we adopt a multiverse analytic approach (Steegen et al., 2016), comparing the results of PLS analysis with and without demographic factors and with and without our task attentiveness measure to evaluate the constraints of the relationship based on the variables considered.

Thus, the current work tests whether mouse movements can indeed be used to predict individuals' personalities in an online task. Specifically, we test whether mouse movements extracted from a simple image-rating task can be used to predict personality traits as assessed by the Big Five Inventory. We present results from bivariate correlations and three multivariate PLS analyses with a large sample of participants ($N = 791$) and discuss the utility of this approach in both research and consumer contexts.

2 | METHOD

2.1 | Data sources

Raw data used in this research combine data collected from two different studies at the University of Chicago. Both studies recruited participants from Amazon Mechanical Turk (AMT). All participants provided informed consent before continuing with the study procedures. In each study, participants were asked to complete an image-rating task, then fill out the Big Five Inventory (BFI) questionnaire, which includes 44 items that measure an individual's personality across five dimensions (John & Srivastava, 1999). Participants' basic demographic information, such as Gender and Age, were also collected at the end of the task.

We used the JavaScript library jQuery to record participants' mouse movements. A record is created whenever a movement occurs. A continuous cursor movement is captured at around 60 Hz, or every 17 ms. However, the exact frequency depends on the type of mouse or touchpad used. Four variables were recorded for each mouse movement entry: (1) timestamp in milliseconds, (2) the cursor's x-coordinate in pixels, (3) the cursor's y-coordinate in pixels, and (4) a dummy-coded variable for click (1 if the participant clicked in the recorded position, and 0 if the participant did not).

2.2 | Participants

Participants were recruited using CloudResearch (https://www.cloudresearch.com/; Litman et al., 2017). After

removing null values and invalid trials, the final cleaned data include 791 participants. Among them, 483 self-identified as *Male*, 303 self-identified as *Female*, and 5 self-identified as *Other*. Participants had a mean age of 38.8 years, with a standard deviation of 10.8 years.

2.3 | Image-rating task

Data for this research come from a set of image-rating tasks. Across all tasks, in each trial, participants were asked to look at 12 photos of streets taken at different angles and pick four images based on how high they were on a given attribute. A demo of the image rating task is available at: https://kywch.github.io/Image RatingStudy/multi-image-rating-demo.html. For example, in one iteration, participants were asked to choose the four images they liked the most, and in a second iteration, they chose the four images they liked the least. In other versions, participants were asked to choose the four images highest on a given perceptual feature, such as perceived walkability, orderliness, and complexity. Figure 1 shows the first four instruction pages shown to participants in one of the image-rating tasks. To ensure that the data were of sufficiently high quality, attention check questions were randomly distributed in multiple trials of the task. During these trials, participants needed to drag a corrupted image into a trash can located in the bottom-left corner of the screen. Participants' sessions were terminated if they failed to pass the attention checks twice.

2.4 | Mouse feature extraction

Code for preprocessing of mouse features can be accessed at: https://github.com/tianyueniu/mouse_movement_personality.

2.4.1 | Time-related features

Time-related features in this study refer to pauses and fixations. In this study, a long pause is defined as cursor inactivity for longer than 4 s (based on Tzafilkou & Nicolaos, 2018), whereas a fixation is defined as micro-movements within 25 pixels that lasted for more than 250 ms (based on Dalmaijer et al., 2014). The following features were extracted from raw mouse movement data.

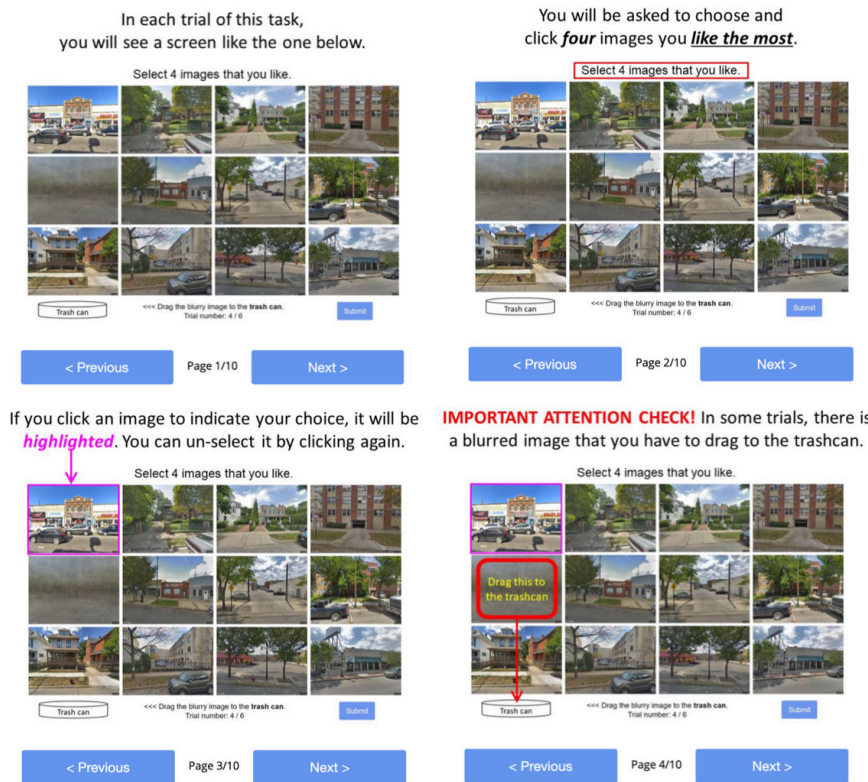> *Total_pause_cnt*: total count of long pauses across all trials.

Avg_fixation_dur: average duration per fixation across trials.

Avg_agg_fixation_dur: average total fixation time per trial.

Avg_fixation_cnt: average number of fixations detected per trial.

2.4.2 | Activity-related features

Activity-related features in this study include distance, time, and speed. The following features are extracted from raw mouse movement data.

Avg_euc_dist: average Euclidean distance traveled in pixels per trial.

Avg_euc_speed: average speed from pixel to pixel measured in milliseconds.

Avg_completion_time: average trial completion time in milliseconds per trial.

2.4.3 | Click-related features

To complete the tasks in this study, participants would have to click five times in normal trials (4 selection clicks

and 1 click on the "continue" button to move on to the next task), or six times in attention check trials (4 selection clicks, 1 drag click, and 1 "continue" button click). If a participant clicked for more times than necessary in a trial, we named the extra clicks as "reclick." The following clicks-related features are extracted from mouse movement data.

avg_click_att: average clicks made in attention trials.

reclick_percent_att: the percent of attention trials in which participants clicked more times than necessary.

avg_click_norm: average clicks made in normal trials.

reclick_percent_norm: the percent of normal trials in which participants clicked more times than necessary.

2.4.4 | Atypical responding

Lastly, in this study, atypical responding is measured as the deviation from random responding. We propose that in this context, atypical or random responding may reflect inattentiveness. This calculation of random responding was based on whether a given participant's choices can or

cannot predict the averaged choice probability across the whole group. To this end, receiver operator characteristic (ROC) curves were calculated for an individual's decision (click or not) about each image relative to the group's average decision (highly clicked or not) about each image. That is, for any given image, if a participant chooses that image (click is 1) and the group average for that image is also very high (image is often chosen and therefore, the average value is close to 1), and vice versa, that participant's choice is highly predictive of the group's choice in a leave-one-out procedure. The area under the curve (AUC) of this analysis reflects how similar the participant's responses are to the group's average responses across all images. If the clicks of an individual are highly predictive of the rest of the group, the AUC will be close to 1. In contrast, if the individual's choices are extremely different from the group's (i.e., intentional opposite responding), the AUC will be close to 0. An individual responding at random would have an AUC close to 0.5. This measure is stored as *Area_Under_Curve* in our data. We then calculated an *Abs_Area_Under_Curve*, which is equal to the absolute value of *Area_Under_Curve* - 0.5, to capture a participant's deviation from random responding, which is used as our final measure for attentiveness in our study.

2.5 | Data analysis method

Eleven features were extracted from raw mouse-tracking data as described in the previous section. The measure for attentiveness (*Abs_Area_Under_Curve*) was calculated from an individual's choice responses. Pearson correlations and OLS regression analyses were performed predicting *Abs_Area_Under_Curve* by the other 11 extracted features to explore the individual relationships between attentiveness and mouse movement features. OLS regression was used to predict each Big Five personality trait by all 11 mouse features and *Abs_Area_Under_Curve*.

Three partial least squares analyses (PLS) were performed to explore the overall relationships between extracted cursor movement features, attentiveness, and Big Five personality scores. We adopt a multiverse analytical approach here to examine the constraints of the results (Steegen et al., 2016) depending on the specific variables examined. As such, the three analyses differed on the specific input matrices included. Analysis (1) included Big Five personality measures ~ All 11 mouse movements and attentiveness (*Abs_Area_Under_Curve*), (2) included Big Five personality measures with Age and Gender ~ All 11 mouse movements and attentiveness (*Abs_Area_Under_Curve*), and (3) included Big Five personality measures ~11 mouse movements only. The primary goal of this work was to specifically identify the relationship between

personality traits and mouse movements (Analysis 3), but we reasoned that other demographic factors (Age and Gender) may influence the overall results. By comparing the results with and without these demographic variables, we can see the extent to which the mouse movement ~ personality relationship is influenced by these variables. Additionally, as our proposed attentiveness/atypicality measure was not a true "mouse movement" and may be hard to quantify in other tasks (i.e., where deviation from average responding is not possible), we wanted to test whether these results held even when removing this variable.

The current work adapted the Behavioral PLS code from https://www.rotman-baycrest.on.ca/ to be used with two matrices of behavioral data (rather than a matrix of behavioral data and a matrix of brain data, such as fMRI or EEG). Matlab code for this study is available at https://osf.io/fr74q/. This PLS analysis extracts maximally covarying latent variables from the covariance matrix. In the case of Analysis 3 (Big Five personality traits and 11 mouse movements), this is the covariance of personality measures (**X** matrix which is 791 participants × 5 personality variables) with a matrix of mouse movement features (**Y** matrix which is 791 participants × 11 mouse movement features) for all participants. Before calculating the covariance matrix, variables were first *z*-scored. Next, the covariance matrix (**S**) was calculated as $X'*Y$, yielding a 5×11 matrix. Subsequently, the singular value decomposition (SVD) was performed on the covariance matrix, **S**.

SVD takes **S** and decomposes it into three matrices $\mathbf{S} = U\Delta V^T$, where U is a singular vector (also called "salience") that represents the decomposition of **S** in mouse movement features space, V is the salience that represents the decomposition of **S** in personality space, and Δ quantifies the weighting of the singular vectors V and U. The results of the SVD are linear combinations of the two data matrices that maximize their covariance, referred to as latent variables. These extracted latent variables (LVs) are mutually orthogonal to one another, and the number of LVs is equivalent to the total number of variables on the lower rank matrix (i.e., in this analysis, there would be a total of five LVs because there are five personality trait variables in the **X** matrix or the lower number of rows or columns in the **S** covariance matrix). In other words, for each latent variable, the singular vector V reflects a linear combination of personality traits (e.g., is a vector of dimension 1 LV × 5 personality traits) and the singular vector U reflects a linear combination of mouse movement features (i.e., is a vector of dimension 1 LV × 11 mouse movements).

Additionally, each participant receives a value for where they fall on each of the LVs, referred to as *V*-scores and *U*-scores. These are calculated by multiplying the *U*

singular vector by the original data in the **Y** matrix (*z*-scored mouse movement features) to get *U*-scores, and by multiplying the *V* vector by the original data in the **X** matrix (*z*-scored personality data) to get *V*-scores.

What is ultimately reported in each LV of the PLS analyses are two sets of correlations. One is a set of correlations between the **X** matrix (*z*-scored personality data, a 791 × 5 matrix) and the *U*-scores (a 791 × 1 vector reflecting participant-level scores on where they fall on the mouse movements side of the LV). The other is a set of correlations between the **Y** matrix (*z*-scored mouse movement data, a 791 × 11 matrix) and the *V*-scores (a 791 × 1 vector reflecting participant-level scores on where they fall on the personality side of the LV). Together, these two sets of correlations identify the relationships between each matrix of data in its original form (i.e., **X** or **Y** matrix) and the other set of data that are now in latent space (i.e., *V*-scores or *U*-scores).

It is worth noting that this SVD procedure is also used in principal component analysis to identify maximally covarying sets of variables within a single set of data, but in PLS, SVD is run on the covariance matrix of two datasets, rather than the covariance of a single matrix (i.e., a single dataset). As such, the latent variables in a PLS analysis reflect a latent relationship between two sets of data (two matrices), rather than the latent structure of one set of data.

To estimate the reliability of the latent variables, permutation testing was conducted by first shuffling the order of one of the two input matrices then running SVD on the newly calculated covariance matrix with the shuffled data. This step was repeated 10,000 times to generate a null distribution and by comparing the amount of covariance explained for the original LV compared to the null distribution. From this, we could derive a *p*-value, which was estimated for each LV. Next, to test whether the pattern of effects (i.e., the linear combination of variables) was reliable, a bootstrapping procedure was employed where rows (corresponding to participants) were resampled with replacement. The procedure was conducted with 10,000 bootstrapped samples to generate 95% confidence intervals around each of the variables in the LV.

Additionally, to determine the effect sizes of the PLS analyses, we first calculated the correlation between participants' *V*-scores and *U*-scores for the first LV in each analysis. By correlating these scores across all 791 participants, we get an overall correlation coefficient for each model that reflects the average relatedness of each dataset at the participant level. However, as correlating these scores requires reducing the data to two vectors, this effect size may be an underestimate of the overall model which leverages this covariance structure across all participants.

Therefore, as an additional measure of effect size, we determined the proportion of covariance accounted for by the significant LV. To do so we first projected the two components of the LV into the original data space via $lv^J_{proj} = \frac{X^J \cdot lv}{\|X^J\|} X^J$, where $\|X\|$ is the norm of matrix X and J is the Jth subject. X^J is the personality vector for the Jth subject that is 1×5, lv is the 5×1. Next, we regressed out each component of the projected LV from the original data to get a reduced dataset X_{red}. Finally, we calculated the effect size as $e = 1 - \left(\frac{det(\Sigma_{red})}{det(\Sigma)} \right)^{2/n}$. Where Σ is the covariance matrix of the original data, Σ_{red} is the covariance matrix of the reduced data, $det(\Sigma)$ is the determinant of the covariance matrix, and n is the number of data dimensions. To understand the meaning of this effect size measure, we can first recognize that the determinant of the covariance matrix provides a measure of the volume of the point cloud defined by the data. Consequently, the square of the nth root of the determinant of the covariance matrix provides a measure of the characteristic or average spread of the data in each dimension. Together, these facts mean that e^2 can be interpreted as the average "length" of data which is explained by the LV. In this sense, e^2 is similar to a traditional R^2 effect size measure in that it roughly scales with the square of the average correlation (i.e., covariance) across different data dimensions (See Figure S1).

3 | RESULTS

3.1 | Variable and sample descriptives and data representativeness

Table 1 includes descriptive statistics for personality traits and mouse movement features.

Before performing further analysis, we compared personality distributions of the studies by Gender and Age with personality distributions found in other literature to examine the representativeness of our data. The distribution of participants' normalized Big Five personality scores by Age found in our analysis [See Supporting Information] is similar to the distribution presented in John and Srivastava's research (John & Srivastava, 1999). Differences in personality traits by Gender (Costa et al., 2001) and Age (Soto et al., 2011) found in our analysis are also supported by previous work.

3.2 | Univariate statistical analyses

3.2.1 | Exploratory analysis of mouse movements and (in)attentive responding

One preliminary aim was to test whether mouse movements are predictive of compliance in the online

TABLE 1 Descriptive statistics of normalized personality traits & mouse movement features

Personality traits	Mean	Std Deviation	Min	Max
(a) Descriptive statistics of participant's normalized Big Five personality scores (N = 791)				
Extraversion	2.82	0.96	1.00	5.00
Agreeableness	3.75	0.73	1.44	5.00
Conscientiousness	3.92	0.77	1.56	5.00
Neuroticism	2.53	0.95	1.00	5.00
Openness	3.61	0.78	1.10	5.00
Mouse movement features	**Mean**	**Std Deviation**	**Min**	**Max**
(b) Descriptive statistics of participant's mouse movement features (N = 791)				
avg_click_att	7.07	1.61	6.00	25.50
reclick_percent_att	0.33	0.21	0.00	1.00
avg_click_norm	5.62	0.72	5.00	14.24
reclick_percent_norm	0.25	0.17	0.00	1.00
avg_euc_dist	5235.33	1613.57	2246.97	15,157.24
avg_euc_speed	0.41	0.11	0.14	1.24
avg_completion_time	14,849.01	6717.49	8101.91	79,774.36
total_pause_cnt	9.46	13.66	0.00	103.00
avg_fixation_dur	868.73	386.60	419.80	3619.54
avg_agg_fixation_dur	7521.76	4169.25	1661.08	32,585.53
avg_fixation_cnt	8.75	2.54	2.60	25.96
Abs_Area_Under_Curve	0.22	0.11	0.00	0.42

Note: *avg_click_att* = average number of clicks on attention trials, *reclick_percent_att* = percent of attention trials in which participants clicked for more times than necessary, *avg_click_norm* = average number of clicks on normal trials, *reclick_percent_norm* = percent of normal trials in which participants clicked for more times than necessary, *avg_euc_dist* = average Euclidean distance traveled in pixels per trial, *avg_euc_speed* = average speed from pixel to pixel in ms, *avg_completion_time* = average trial completion time in ms, *total_pause_cnt* = total count of long pauses across all trials, *avg_fixation_dur* = average duration per fixation across trials in ms, *avg_agg_fixation_dur* = average total fixation time per trial in ms, and *avg_fixation_cnt* = average number of fixations detected per trial.

experiment. Our proposed proxy for compliance and task attentiveness is the *Abs_Area_Under_Curve* measure, in which lower values reflect more random, atypical responses in the image choice task. As measured by Pearson correlation analyses with Bonferroni correction for multiple comparisons, significant negative relationships were found between all click-related features and *Abs_Area_Under_Curve* (all $p < 0.001$): average number of clicks on attention trials ($r = -0.18$), percent of reclicks on attention trials ($r = -0.19$), average number of clicks on normal trials ($r = -0.17$), and percent of reclicks on attention trials ($r = -0.17$). Additionally, the number of fixations was positively correlated with *Abs_Area_Under_Curve* ($r = 0.11$, $p = 0.002$). No other correlations were significant (see Supporting Information for full correlation matrix). While additional work is needed to fully establish whether this measure of atypical responding does indeed reflect inattentiveness, these results provide preliminary evidence that click-based features, in particular, may be useful for capturing atypical, inattentive responding.

3.2.2 | Mouse movements and personality

Bivariate correlations between all mouse click measures and personality features are presented in Table 2.

OLS regressions were also conducted on each of the Big Five personality traits by the 11 mouse movement features and atypical responding/inattentiveness to the task (*Abs_Area_Under_Curve*) [See Supporting Information for regression tables]. Overall model R^2 values for each personality trait ranged between 0.03 (Neuroticism) and 0.08 (Conscientiousness). Only *Abs_Area_Under_Curve* was significantly predictive in the models examining Extraversion (negatively), Openness (positively), and Conscientiousness (positively) in these models. Lower *Abs_Area_Under_Curve* and higher average number of clicks during attention trials were significantly predictive of Neuroticism. Higher *Abs_Area_Under_Curve* and lower average number of clicks during attention trials were predictive of Agreeableness. No other significant relationships were found in these analyses, however, and

TABLE 2 Bivariate correlations between mouse movements and Big Five scores

	Extroversion	Agreeableness	Conscientiousness	Neuroticism	Openness
avg_click_att	**0.099**[**]	**−0.085**[*]	**−0.076**[*]	0.062	−0.028
reclick_percent_att	**0.106**[**]	−0.030	−0.049	−0.002	**−0.083**[**]
avg_click_norm	0.025	**−0.073**[*]	**−0.077**[*]	0.031	−0.041
reclick_percent_norm	0.033	**−0.071**[*]	**−0.082**[*]	0.018	−0.065
avg_euc_dist	0.016	0.002	0.014	0.007	0.036
avg_euc_speed	0.003	−0.051	**−0.081**[*]	**0.074**[*]	−0.049
avg_completion_time	0.016	0.026	0.043	−0.047	0.041
total_pause_cnt	0.028	0.014	−0.019	−0.014	−0.002
avg_fixation_dur	−0.003	−0.011	−0.013	0.002	−0.019
avg_agg_fixation_dur	−0.033	0.066	**0.107**[**]	−0.055	**0.078**[*]
avg_fixation_cnt	−0.053	**0.107**[**]	**0.156**[**]	**−0.071**[*]	**0.119**[**]
Abs_Area_Under_Curve	**−0.125**[**]	**0.119**[**]	**0.216**[**]	**−0.098**[**]	**0.117**[**]

Note: Values reflect Pearson correlation coefficients between measures; significance indicators reflect uncorrected *p*-values.

Bold values indicate significant correlations with uncorrected alpha = 0.05.

*Correlation is significant at the 0.05 level.; **Correlation is significant at the 0.01 level.

this was again likely due to multicollinearity of mouse features.

3.3 | Multivariate statistical analyses

To better examine whether mouse movements can reflect individual differences in personality traits, partial least squares (PLS) analyses were run. The results of the OLS regressions are hard to interpret due to the high intercorrelation between click-based measures, leading to multicollinearity in the multiple regression models. However, these intercorrelations may actually reflect something interesting and meaningful about the relationship between combinations of mouse movements and personality traits, so treating them separately (in the case of bivariate correlations) is also a suboptimal approach. In comparison, multicollinearity is not an issue in PLS and this task allows for all of the personality traits to be examined in the same analysis, rather than treating these personality traits as independent.

3.3.1 | PLS analysis 1: Big Five ~ mouse movements and AUC

The first latent variable (LV 1) of the PLS analysis was significant (*p* = 0.001) and explained 91% of the cross-block covariance. Figure 2 shows the results of LV 1. On the personality side, the first latent variable corresponds to higher Agreeableness, Conscientiousness, Openness, and lower Neuroticism. On the mouse movement features side, LV

1 corresponds to a lower number of clicks/reclicks, lower euclidean speed (i.e., faster mouse movements), longer aggregated fixation durations, more fixations, and less atypical, random responding (higher AUC). Euclidean distance, time, pauses, and fixation durations did not show a reliable relationship in LV 1 (as shown by 95% CI bars crossing 0). The relationship between the two sets of variables suggests that participants who are more Agreeable, Conscientious, and Open, and less Neurotic also show mouse movement patterns associated with greater attentiveness and care while doing this task (i.e., fewer unnecessary clicks, slower movements, more pauses, and less random responding).

3.3.2 | PLS analysis 2: Big Five with demographics ~ mouse movements and AUC

When Age and Gender were included in the PLS, the first latent variable (LV 1) of the PLS analysis was significant (*p* < 0.001) and explained 85% of the cross-block covariance. Figure 3 shows the results of LV 1. The overall pattern of results was similar to the first PLS analysis. With the inclusion of these demographic variables, Age, Conscientiousness, and Openness showed reliable relationships with almost all mouse features. In these results, being older and scoring higher on Conscientiousness and Openness was associated with less unnecessary clicks, slower mouse movements, more and longer fixations, and greater attentiveness/more typical responding. Relative to Analysis 1, the inclusion of age created more consistent relationships with timing-related features, although fewer personality features showed reliable loadings.

FIGURE 2 Factor loadings for LV 1 in PLS analysis 1 (with AUC). Error bars on each of the variables represent 95% confidence intervals based on bootstrapped estimates. Loadings reflect correlation coefficients for each variable in one panel (i.e., left panel: Personality) with the linear combination of other panel variables (i.e., right panel: Mouse movement features). The relationship is symmetric, so the loadings for each variable on the right panel (mouse movement features) also reflect correlation coefficients with the linear combination of left panel variables (personality).
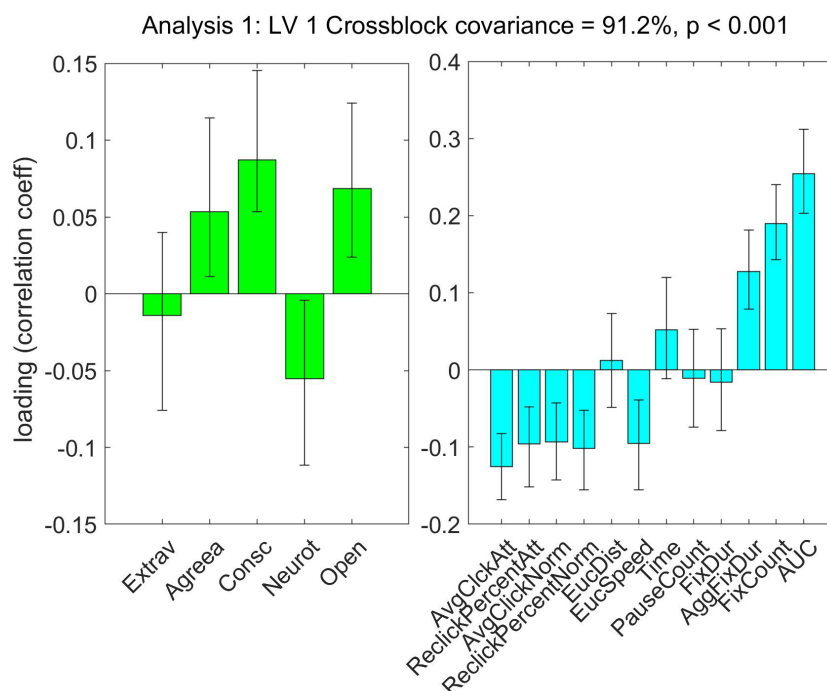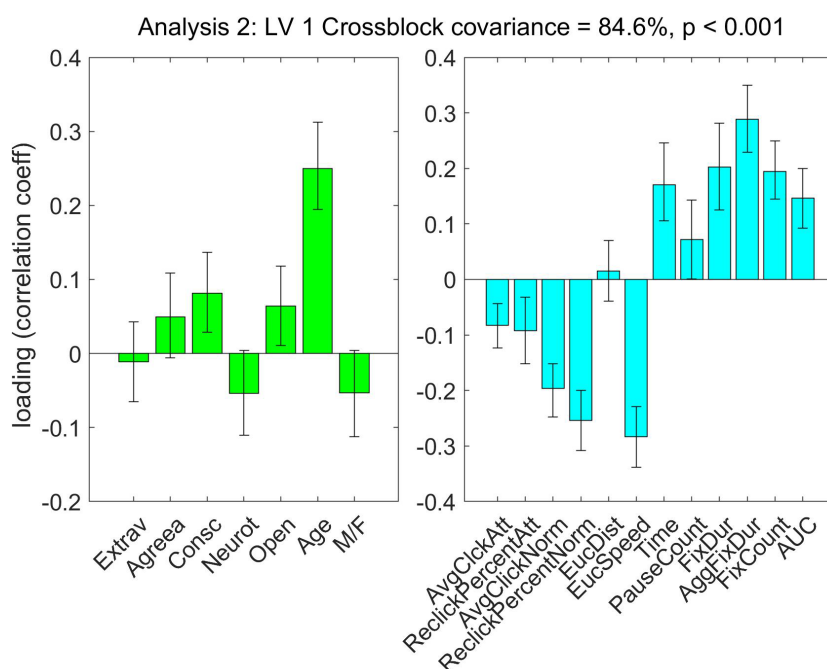

Analysis 1: LV 1 Crossblock covariance = 91.2%, p < 0.001

FIGURE 3 Factor loadings for LV 1 in PLS analysis 2 (including demographics and AUC). Error bars on each of the variables represent 95% confidence intervals based on bootstrapped estimates.

Analysis 2: LV 1 Crossblock covariance = 84.6%, p < 0.001

3.3.3 | PLS analysis 3: Big Five ~ mouse movements

When AUC was removed from the right-side matrix in the PLS regression, the first latent variable (LV 1) was still significant ($p < 0.001$) and explained 88% of the cross-block covariance. Figure 4 shows the results of LV 1. The results replicated PLS Analysis 1: even when removing the attentiveness measure (*Abs_Area_Under_Curve*), the combination of high Agreeableness, Conscientiousness, and

Openness and low Neuroticism was associated with less unnecessary clicking and more fixations.

3.3.4 | Effect sizes for PLS analyses

The first form of effect size calculated here involved correlating participants' *V*-scores and *U*-scores for each analysis, which creates an overall model correlation coefficient. These are reported for each analysis in Table 3. The effect
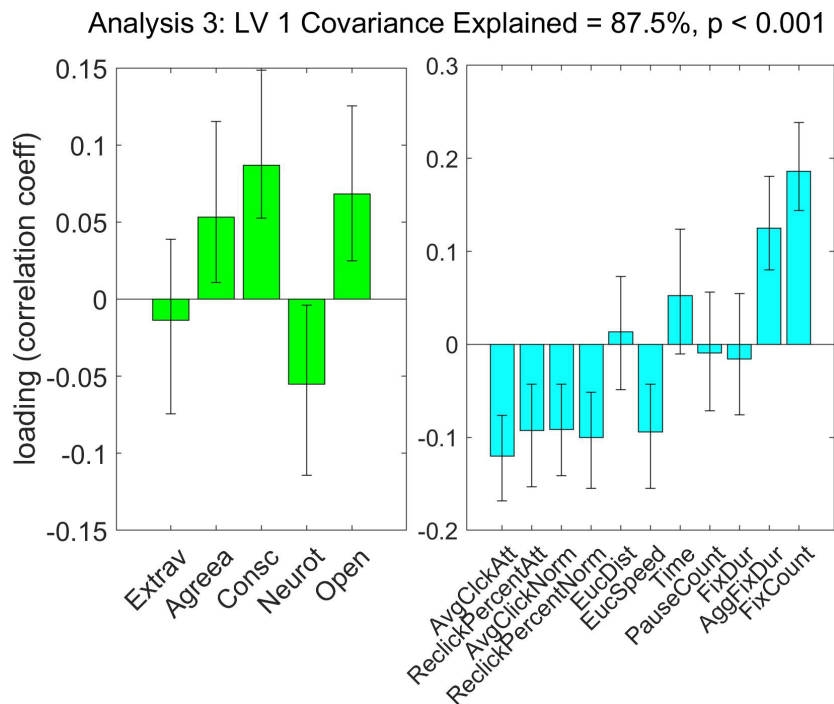
TABLE 3 Effect sizes for PLS analyses using model r (correlation between U-scores and V-scores) and e^2 (pseudo R^2)

Analysis	Correlation between U-score & V-score, r	Personality measures, e^2	Mouse features, e^2
1. Big Five ~ Mouse Movements & AUC	0.29	0.89	0.72
2. Big Five with Demographics ~ Mouse Movements & AUC	0.38	0.83	0.70
3. Big Five ~ Mouse Movements	0.22	0.92	0.64

sizes as measured by r (between 0.22 and 0.38) reflect small-to-medium effect sizes on average using Cohen's original guidelines for r (0.1 indicates a small effect, 0.3 indicates a medium effect; Cohen, 1992), and medium-to-large effect sizes using updated guidelines for individual differences research (0.1 indicates small, 0.2 indicates medium, and 0.3 reflects a large effect; Gignac & Szodorai, 2016).

Additionally, the effect sizes, e^2, which reflect the average "length" or "amount" of data that can be explained by the latent variable, scaling similarly to a traditional R^2 value, were between 0.83 and 0.92 for the personality side of the LVs and between 0.64 and 0.83 for the mouse movements side of the LVs. In other words, about 64% to 83% of the mouse-related features data and 83% to 92% of the personality-related data can be explained by the LVs in the PLS analysis. The full table of e^2 values is reported in Table 3.

4 | DISCUSSION

The goal of this work was to determine whether mouse movements, on a simple judgment task, could be used to predict individual differences in personality traits as well as compliance and attentiveness in an online experiment. The key finding of a set of multivariate analyses was that a pattern of mouse movement features that are reasonably interpretable as indicating greater care and attention to the task were associated with the expected personality traits of Conscientiousness, Agreeableness, and Openness, and negatively related to Neuroticism. Importantly, by approaching this question via a data-driven, multivariate analysis, a clearer and more robust pattern of results was found than what individual OLS regressions showed.

A key advantage of this multivariate approach is that it does not treat the Big Five traits as independent factors, but rather leverages the presence of known intercorrelations between these factors (Soto & John, 2012; van der Linden et al., 2010). In particular, Agreeableness and Conscientiousness tend to be positively interrelated and show negative relationships with Neuroticism (DeYoung et al., 2002; Digman, 1997). It is proposed that the presence of these intercorrelations reflects a set of higher-order factors or a single, general factor of

personality, although whether higher-order factor(s) result from a common mechanism or reflect a measurement artifact remains contested (DeYoung et al., 2002; Digman, 1997; van der Linden et al., 2021). While the ultimate cause of these intercorrelations is outside of the scope of the current work, the results of the current study suggest that examining linear combinations of the five factors (akin to a higher-order factor) will facilitate prediction from measures that are not explicitly created for personality measurement.

The inclusion of demographic variables (Age and Gender) changed the PLS results somewhat, although the overall pattern of effects stayed the same. Specifically, including age created stronger links to timing-related mouse features, and lessened the role of Agreeableness and Neuroticism. Previous studies suggest that, in adulthood, Agreeableness increases with age, and Neuroticism decreases with age (Soto et al., 2011). Consistent with this, in our sample, age was also positively correlated with Agreeableness ($r = 0.21$) and negatively related to Neuroticism ($r = -0.20$). Interestingly, Conscientiousness was positively correlated with age in our sample ($r = 0.27$) and in previous research (Soto et al., 2011; Soto & John, 2012), but remained a reliable loading on the personality side of this PLS analysis. Thus, while these results remained reasonably consistent across all analyses, this analysis demonstrates the utility of a multiverse approach to determine the constraints of an observed statistical relationship and suggests that Conscientiousness is more readily predicted by mouse movements even when Age and Gender are taken into account.

Although the general approach described here can be used by researchers interested in inferring personality traits or other internal characteristics from mouse movement features, the specific mouse movement patterns or traits of interest may be different depending on the task context and the aims of the researcher. As such, in practice, we propose consumer researchers take a multi-step approach. First, we recommend collecting both mouse movements and trait measures of interest (via questionnaire) from a subset of users or participants. Next, we recommend conducting a PLS analysis to identify which mouse movement features predict a profile of the trait-level factors, each in the form of a linear combination of mouse features. The analysis script used in the current work is accessible at https://osf.io/fr74q/ and can be easily modified to account for different datasets as inputs for the model(s). Based on the results of this PLS analysis, researchers may choose to include or exclude some of the trait-level attributes examined to determine if they are influenced by other data collected, such as demographic information or other user-level characteristics. Researchers may need to continue to collect data until their results stabilize to a point where the permutation tests on the latent variable(s) reach a desired threshold (e.g., $p < 0.05$). The factor loadings (correlation coefficients) of the mouse movement features can subsequently be used as weights to predict the trait profile of interest in future users. While the current work only examined personality as measured by the Big Five Inventory (John & Srivastava, 1999), it is very likely that other individual differences may show interesting relationships when examined in this way. Additionally, an exciting future direction would be to relate this to state-level differences in affect, fatigue, or other factors of interest to researchers.

Additionally, we find significant negative correlations between our proposed measures of atypical, inattentive responding (*Abs_Area_Under_Curve*) and all click-related features at $\alpha = 0.0045$. While work is necessary to determine whether a lower *Abs_Area_Under_Curve* does indeed reflect more random, inattentive responding rather than something more general about looseness and deviation from normative behaviors (Gelfand et al., 2011), these results provide preliminary evidence that click rate alone may be a useful metric of inattentiveness. Intuitively, more clicking can mean impatience or random clicking, which in turn leads to more inattentive or random responding, which combined with these preliminary results, support the potential use of mouse movements to filter out inattentive responding in online research.

While this work shows the feasibility of inferring personality from mouse movements in a simple image rating task, future work should further examine this in different tasks to gain more comprehensive insights on how to accurately infer internal traits based on cursor movement features. Additionally, as the study was not initially designed for mouse movement analysis, we did not record the different types of cursor devices used by our participants (e.g., touchpad vs. an actual mouse). Different types of devices might lead to variation in the raw data captured. Future work should make note of this nuisance and separate users' devices for further analysis.

Lastly, although the current work employs two metrics for calculating the effect size of an overall PLS model (overall model r and newly proposed e^2), the magnitude and interpretations of these two effect sizes are quite different. Specifically, the overall model r focuses on how related the two sets of data are at the participant level. In contrast, e^2 is potentially a better metric of effect size in a multivariate analysis, as it is providing additional information on the predictive power of viewing the relationship of two datasets in latent space across all participants at once. However, future work is necessary to examine the extent to which the overall model r, the newly proposed e^2,

or a combination of the two provide the most reliable and accurate measure of PLS analysis effect size.

A compelling implication from this work is that researchers may be able to infer individuals' personality characteristics without explicitly asking about them, even in tasks that are not designed to evaluate personality, such as the simple image choice task used here. Going forward, this work demonstrates researchers across academic and consumer sectors may be able to leverage multivariate analyses with mouse movements to infer a variety of individual differences.

AUTHOR CONTRIBUTIONS

Conceptualization: Tianyue Niu and Kyoung W. Choe. **Data curation:** Kimberly L. Meidenbauer and Tianyue Niu. **Formal analysis:** Kimberly L. Meidenbauer, Tianyue Niu, Kyoung W. Choe, Andrew J. Stier, and Marc G. Berman. **Funding acquisition:** Marc G. Berman. **Project administration:** Kyoung W. Choe. **Resources:** Kyoung W. Choe and Marc G. Berman. **Software:** Kimberly L. Meidenbauer, Tianyue Niu, and Kyoung W. Choe. **Visualization:** Kimberly L. Meidenbauer and Tianyue Niu. **Writing – original draft:** Kimberly L. Meidenbauer and Tianyue Niu. **Writing – review & editing:** Kimberly L. Meidenbauer, Tianyue Niu, Kyoung W. Choe, Marc G. Berman, and Andrew J. Stier.

CONFLICTS OF INTEREST
The authors declare no conflicts of interest.

ETHICS STATEMENT
All participants gaveonline informed consent before participation and experimental procedures wereapproved by the University of Chicago's Committee for Institutional ReviewBoard (IRB).

ORCID
Kimberly L. Meidenbauer https://orcid.org/0000-0001-9135-6130
Kyoung Whan Choe https://orcid.org/0000-0002-2138-9807
Andrew J. Stier https://orcid.org/0000-0002-2863-5978
Marc G. Berman https://orcid.org/0000-0002-7087-3697

REFERENCES

Arapakis, I., Lalmas, M., & Valkanas, G. (2014). Understanding within-content engagement through pattern analysis of mouse gestures. In *Proceedings of the 23rd ACM International Conference on conference on Information and Knowledge Management* (pp. 1439–1448). Association for Computing Machinery. https://doi.org/10.1145/2661829.2661909

Arapakis, I., & Leiva, L. A. (2016, July). Predicting user engagement with direct displays using mouse cursor information. In *Proceedings of the 39th International ACM IGIR conference on Research and Development in information retrieval* (pp. 599–608). Association for Computing Machinery. https://doi.org/10.1145/2911451.2911505

Barends, A. J., & de Vries, R. E. D. (2019). Noncompliant responding: Comparing exclusion criteria in MTurk personality research to improve data quality. *Personality and Individual Differences*, *143*, 84–89. https://doi.org/10.1016/j.paid.2019.02.015

Berry, K., Rana, R., Lockwood, A., Fletcher, L., & Pratt, D. (2019). Factors associated with inattentive responding in online survey research. *Personality and Individual Differences*, *149*, 157–159.

Brebner, J. (1983). A model of extraversion. *Australian Journal of Psychology*, *35*(3), 349–359. https://doi.org/10.1080/00049538308258748

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Costa, P. T., Terracciano, A., & Mccrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, *81*(2), 322–331. https://doi.org/10.1037/0022-3514.81.2.322

Dalmaijer, E. S., Mathôt, S., & Van der Stigchel, S. (2014). PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eye tracking experiments. *Behavior Research Methods*, *46*, 913–921. https://doi.org/10.3758/s13428-013-0422-2

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2002). Higher-order factors of the Big Five predict conformity: Are there neuroses of health? *Personality and Individual Differences*, *33*(4), 533–552.

Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology*, *73*(6), 1246–1256.

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W. H. (2018). Attention and information acquisition: Comparison of mouse-click with eye-movement attention tracking. *Journal of Eye Movement Research*, *11*(6), 1–27.

Freeman, J. B. (2018). Doing psychological science by hand. *Current Directions in Psychological Science*, *27*(5), 315–323.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., Duan, L., Almaliach, A., Ang, S., Arnadottir, J., Aycan, Z., Boehnke, K., Boski, P., Cabecinhas, R., Chan, D., Chhokar, J., D'Amato, A., Ferrer, M., Fischlmayr, I. C., ... Yamaguchi, S. (2011). Differences between tight and loose cultures: A 33-nation study. *Science*, *332*(6033), 1100–1104.

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78.

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

John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (Vol. *2*, pp. 102–138). Guilford Press.

Khan, I. A., Brinkman, W. P., Fine, N., & Hierons, R. M. (2008). Measuring personality from keyboard and mouse use. In *Proceedings of the 15th European conference on cognitive ergonomics the ergonomics of cool interaction - ECCE 08, Article 38* (pp. 1–8). Association for Computing Machinery. https://doi.org/10.1145/1473018.1473066

Koppensteiner, M. (2013). Motion cues that make an impression. *Journal of Experimental Social Psychology*, *49*(6), 1137–1143. https://doi.org/10.1016/j.jesp.2013.08.002

Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *PLOS ONE*, *9*(9), e105825.

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Meade, A. W., & Pappalardo, G. (2013, April). Predicting careless responses and attrition in survey data with personality. Paper presented at the 28th Annual Meeting of the Society for Industrial and Organizational Psychology, Houston, TX.

Orchard, L. J., & Fullwood, C. (2010). Current perspectives on personality and internet use. *Social Science Computer Review*, *28*(2), 155–169.

Rodden, K., Fu, X., Aula, A., & Spiro, I. (2008). Eye-mouse coordination patterns on web search results pages. In *CHI'08 extended abstracts on human factors in computing systems* (pp. 2997–3002). Association for Computing Machinery. https://doi.org/10.1145/1358628.1358797

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612.

Soto, C. J., & John, O. P. (2012). Development of big five domains and facets in adulthood: Mean-level age trends and broadly versus narrowly acting mechanisms. *Journal of Personality*, *80*(4), 881–914.

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology*, *100*(2), 330–348.

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712.

Stolier, R. M., & Freeman, J. B. (2016). Neural pattern similarity reveals the inherent intersection of social categories. *Nature Neuroscience*, *19*(6), 795–797.

Tzafilkou, K., & Nicolaos, P. (2018). Mouse behavioral patterns and keystroke dynamics in end-user development: What can they tell us about users' behavioral attributes? *Computers in Human Behavior*, *83*, 288–305. https://doi.org/10.1016/j.chb.2018.02.012

Tzafilkou, K., Nicolaos, P., & Charalampos, Y. (2014). Mouse tracking for web marketing: Enhancing user experience in web application software by measuring self-efficacy and hesitation levels. *International Journal on Strategic Innovative Marketing*, *1*, 233–247. https://doi.org/10.15556/ijsim.01.04.005

van der Linden, D., Dunkel, C. S., & Wu, P. (2021). Is there a meaningful general factor of personality? *The Spanish Journal of Psychology*, *24*, e9.

van der Linden, D., te Nijenhuis, J., & Bakker, A. B. (2010). The general factor of personality: A meta-analysis of Big Five intercorrelations and a criterion-related validity study. *Journal of Research in Personality*, *44*(3), 315–327.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.