

Thinking Fast or slow? Understanding Answering Behavior Using Dual-Process Theory through Mouse Cursor Movements

David Kim
The University of Arizona
davidkim7@arizona.edu

Joseph S. Valacich
The University of Arizona
valacich@arizona.edu

Jeffrey L. Jenkins
Brigham Young University
jjenkins@byu.edu

Manasvi Kumar
The University of Arizona
manasvik@email.arizona.edu

Alan R. Dennis
Indiana University
ardennis@indiana.edu

Abstract

Users' underlying cognitive states govern their behaviors online. An extreme cognitive burden during live system use would negatively influence important user behaviors such as using the system and purchasing a product. Thus, inferring the user's cognitive state has practical significance for the commercialized systems. We use Dual-Process Theory to explain how the mouse cursor movements can effectively measure cognitive load. In an experimental study with five hundred and thirty-four subjects, we induced cognitive burden then monitored mouse cursor movements when the participants answered questions in an online survey. We found that participants' mouse cursor movements slow down when engaged in cognitively demanding tasks. We further derived new measures to infer the state of heightened cognitive load with an overall accuracy of 70.22%. The results enable researchers to measure users' cognitive load with more granularity and present a new, theoretically sound method to assess the user's cognitive state.

1. Introduction

Understanding the key behaviors of online users is becoming increasingly important as it allows an organization to make a personalized recommendation to each user [1,2], predict future sales [3], and attract and retain customers [4]. Yet, the analysis and the usage of the online data still pose a challenge as user behaviors are influenced by underlying cognitive processes, which is difficult to infer without setting up a theoretical boundary. Specifically, online user behaviors involve numerous decision-making processes that pertain to a specific goal. Users may navigate and search the web to retrieve information, make selections and decisions based on the retrieved information (e.g., online shopping), share new information, form new social ties, communicate with

others, or even alter their offline behavior based on information retrieved online [5].

As Internet Technology and e-commerce continues to develop more rapidly, Information Systems (IS) researchers searched for ways to effectively assess the users' attitudes and behaviors. The most significant example in the IS discipline is the Technology Acceptance Model (TAM) [6]. TAM states that users' perceive that the ease of use of a system is mainly dependent on the interactive relationships between the user's abilities (i.e., knowledge and working memory) and the system-specific factors (e.g., the usability of the user interface, visual layout of the page, and interaction design) [7]. These factors influence the user's cognitive load while using the system and influence important user behaviors.

Though the TAM model well explains user behavior in the context of intention to use the system, identifying what drives specific online user behaviors (i.e., using the system vs. churn) remains a challenge. Specifically, measuring the level of cognitive load and deriving models is exhaustive since different online tasks (i.e., some complex tasks) exhibit distinctly different usage behaviors. Further, the existing methods of assessing and evaluating user's cognitive states heavily rely on the design and tool factors (e.g., surveys that only pertain to a specific context, timed responses, mock experiments).

To address this challenge, researchers are using Human-Computer Interaction (HCI) devices to measure and infer the underlying cognitive processes of online users. HCI studies have shown that various physiological reactions to the stimulus could be captured and analyzed using noninvasive, ubiquitous devices such as computer mice [8,9,10,11], face recognition sensors [12], and sound detection sensors [13]. Prior research in this area had examined various online users' behaviors, including habituation [14], concealing information [11] and, fraud [8,15].

This paper focuses on theoretically explaining how the underlying cognitive processes manifest as the difference in mouse movement speeds. Prior literature had also examined such a relationship, but the guiding theories across these studies vary significantly. For instance, Byrd (2018) had examined the impact of countermeasures in HCI-Based Deception Detection methods using computer mice and Signal Detection Theory [16]. Jenkins (2019) paired Response Activation Model [11] with mouse-cursor movements to identify movement characteristics of individuals that are concealing information. Lastly, Hibbeln (2017) drew from Attention Control Theory [10,17] to explain how negative emotions influence mouse cursor movements. Our research builds and extends this work by providing a new, theoretical explanation of how responses to stimuli manifest in mouse usage metrics at a primitive level. In doing so, we provide new methods and measures applicable in various situations.

First, we draw from Dual-Process Theory (DPT) to explain how our new methods can effectively capture the fluctuations in users' cognitive load exhibited in their device usage behaviors. DPT asserts that humans leverage two distinct cognitive processes for any given task: (1) intuitive and (2) deliberative. In terms of system usage, an *intuitive system* is utilized by default. When confronted with a more difficult task, a user then activates a *deliberative system* to complete the task. Second, we merge DPT with research demonstrating how changes in cognitive load influence device usage patterns. Thus, we provide a theoretical explanation of how the user's cognitive process, guided by DPT, influences device usage patterns to answer our research question: *Can the user's device usage behaviors be utilized to measure cognitive load?* In doing so, we introduce a new way to derive HCI metrics from the user's response behaviors. Third, we evaluate our new measures' ability to differentiate between cognitively demanding tasks (i.e., System 2 related tasks) vs. simple tasks (i.e., System 1 related tasks) with multiple predictive models. This allows us to answer our second research question: *Given the behavioral metrics of a user, can we identify whether the tasks were cognitively demanding (i.e., likely to activate System 2) for the user?*

We contribute to the literature in various ways. First, we present a theoretical explanation of why the response to a stimulus would manifest as a psychophysiological response. Specifically, we demonstrate that DPT explains the relationship between cognitive loads and physiological responses. Second, we provide a practical measure that can be used for capturing and measuring a broad range of

online behaviors. We demonstrate that difference in users' response behaviors, when engaged with tasks with varying levels of complexity, provides powerful signals of the underlying cognitive processes. Lastly, we empirically validate our measures with predictive models to demonstrate the effectiveness of the newly derived measure.

2. Background

2.1. Cognitive Load in Information Systems Research

The influence of cognitive load on important user behaviors has been a popular topic in behavioral data analytics. Specifically, in prior Information Systems (IS) research, the relationships between users' cognitive load and behavior have been studied under the context of perceived ease of use [6,7], user satisfaction [18,19], and reuse intentions [20]. This prior research typically leverages experimental designs and analytical methods to examine hypotheses and impose underlying assumptions. For instance, in a simple experimental design, subjects are randomly assigned into one of two conditions with varying levels of mental workload. Then, the subjects are presented with sets of questions designed to measure a specific construct (e.g., user satisfaction).

Emerging HCI research combines behavioral science and design science approaches to derive new cognitive load measures. Like traditional behavioral studies, HCI studies leverage stimuli that alter the level of cognitive load of participants (e.g., screening process). The change in participants' cognitive load is accompanied by physiological responses that are being captured using a variety of devices and approaches including, conversational agents [12], eye-tracking technology [2], facial and voice recognition systems [21], mobile devices [22,23], and computer mouse movements [8,10,15,21,24].

We leverage the computer mouse as a tool to measure user's internal cognitive processes. Numerous studies have demonstrated that hand movements were predictive of cognitive processing and heightened cognitive load [10,24,21,25,26].

Specifically, when cognitive load is heightened, users take additional time to complete tasks while interacting with a device, such as moving a mouse or typing in a response on a keyboard. We argue that such behaviors will manifest as significant changes in mouse movement speeds. We build and extend the prior literature by examining the mouse movements in the context of DPT.

2.2. Dual-Process Theory (DPT)

DPT is a widely accepted model that explains human reasoning at a primitive level [27,28,29,30,31,32,33]. Precisely, it posits that human thinking consists of two distinct types: System 1 (also referred to as intuitive and heuristic) and System 2 (deliberative and analytic). Though there are many different extensions of DPT, a dominant framework in the psychology field is the Default Interventionist (DI) Dual-Process model. The DI model extends DPT by laying a serial view on the interaction between System 1 and System 2 [29,33]. The model states that when humans are given a reasoning problem, they will rely on System 1 by default, with System 2 being activated when evaluating or correcting System 1 output as needed [29,30].

2.2.1. System 1 – Intuitive System. System 1 comprises a set of sub-systems that operate with autonomy [28,33]. The behavioral output from System 1 is instinctive, immediate, effortless, and fast. When given a reasoning task, the user first utilizes System 1 by default to produce an output. For instance, when presented with a survey question asking the respondent's age, a respondent would read the question and decide on an answer without requiring extensive cognitive resources. In this case, a user will first mentally prepare a response then select a response target without re-evaluating the answers by activating System 2. Hence, the underlying cognitive process that falls under System 1 would be exhibited as fast, immediate, and autonomous behaviors.

2.2.2. System 2 – Analytic System. System 2 allows us to engage in many of the behaviors bounded by the rules of reasoning and evidence [34,35]. For instance, when humans are involved in mathematical computations, they encode reality into abstract symbols and numbers then follow a specific process to produce an outcome. In this example, the abstract symbols and processes (i.e., guided by mathematical theories) would be equivalent to rules of reasoning and evidence. When faced with such tasks that demand high cognitive resources, System 2 will activate to evaluate System 1 responses. When System 2 is active, a person will exhibit response behaviors that are conscious, analytical, effortful, relatively slow, and deliberate [28, 33].

2.3 Dual-Process Theory and Mouse Cursor Movements

We next describe how hand movements can be predictive of underlying cognitive processes within

the context of DPT. Specifically, to explain how System 1 and System 2 responses manifest in different types of HCI device usage. First, the DI model explains how competing systems (e.g., intuitive and analytic systems) influence an individual's cognitive load during the task execution [30,33]. Studies examining the DI model typically leverage the Cognitive Reflection Test (CRT) to empirically demonstrate that System 2 processing is slower than System 1 processing [36,37]. Specifically, a recent study that leveraged CRT and a mouse tracking methodology and a bat-and-ball problem, where the participants were given a question stating: "a bat and a ball together costs £1.10. A bat costs £1 more than a ball. How much does a ball cost?" On the same screen, participants were given four options placed on each corner of the computer screen. There were four options that a user could choose, including an intuitively appealing response (e.g., 0.10) and a choice that requires deliberative reasoning (e.g., 0.05). The results suggested that the participants who chose the intuitively appealing response (i.e., favoring System 1) had higher mouse cursor speed than those who chose the correct options [37].

Thus, when the users are engaged in simple tasks online using a computer mouse, their response behaviors will be faster, direct, and immediate as they are less likely to activate System 2 to evaluate System 1 output. The user's response behaviors will be slower for complex tasks with higher cognitive demands as tasks may require System 2 to assess the System 1 output. We argue that DPT aligns with Prior HCI research that provided preliminary support that cognitive and motor systems are closely intertwined [8,14,15]. In summary, we propose that:

H1. Users will exhibit slower mouse cursor speed when executing higher complexity tasks.

Though prior literature suggests System 2 related thinking is significantly slower than System 1 related thinking, Individual differences in cognitive processing capacity can influence response behaviors. Specifically, the processing capacity of individuals is systematically different [38,41] due to reasons such as aging, lack of experience, surrounding environments [40,41,42]. Thus, when the users are faced with similar tasks, the magnitude of changes in users' mouse cursor speed will vary depending on the individual's cognitive processing capacity. For example, Hibbeln (2018) randomly manipulated negative emotions in three experimental studies to decrease attentional control (i.e., cognitive capacity) to demonstrate that negative emotions influence the mouse cursor distance and speed.

We argue that even when considering individual differences in cognitive processing capacity, the

behavioral gap between System 1 and System 2 still provides valuable signals for the user's internal cognitive processes. Specifically, when the users' cognitive processing capacity is negatively impacted, they are still likely to be completing simpler tasks faster (i.e., we still expect the overall movement patterns associated with System 2 tasks to be faster and shorter). Thus, we propose:

H2. Users will still exhibit slower mouse cursor speed when executing higher complexity tasks regardless of differences in cognitive processing capacity.

3. Measures

3.1. Mouse Movement Metrics

We collected mouse-movement data when users answered the demographics questions and the main survey questions related to personality traits and work experience. The embedded JavaScript captures the x-coordinates and y-coordinates of the mouse movement at millisecond precision. The captured mouse coordinates are then used to map both actual trajectories to calculate the cursor distance per question.

The distance between two points $P_t(x_t, y_t)$ and $P_{t-1}(x_{t-1}, y_{t-1})$ is given by,

$$d(P_t, P_{t-1}) = \sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2}$$

where t denotes timestamp in milliseconds. Similarly, the average velocity between data points is derived by,

$$\bar{v} = \frac{\Delta d}{\Delta t}$$

where Δd denotes the change in distance and Δt denotes the change in time. To assure that we only capture the mouse movements, we have controlled and excluded mobile devices for this study.

3.2. Feature Level Normalization

To derive new metrics and examine Hypothesis 1, we performed feature scaling and normalization on the variables of interest [43]. We compare each velocity value to other velocity values throughout the entire survey. In doing so, we derive a Feature-Level-Normalized Score (FLNS) that reflects variability in response behaviors (i.e., velocity) across different tasks. Thus, feature level normalizations allow us to examine how the mouse cursor speed differs per task that varies in complexity.

For computer mice, movement data is captured in pixel units at millisecond precision. Thus, when there

is a sudden movement within a short amount of time, the feature value (e.g., velocity) can rapidly increase as well. In terms of velocity, when there is a minimal movement within a short amount of time, the recorded feature value will stay near 0. Since the difference in range of features can interfere with the performance of machine learning algorithms that uses gradient descent as an optimization technique (e.g., logistic regression), we performed min/max normalizations to normalize the mouse cursor velocity to the range [0,1]:

$$\tilde{v} = \max [0, \min \left(1, \frac{v_i - [v]}{[v] - [v]} \right)]$$

v_i is the i^{th} raw velocity feature and $[v]$ and $[v]$ are lower and upper bounds values for normalization. The normalization bounds are calculated by:

$$[v] = \mu_v - H_f \sigma_v$$

$$[v] = \mu_v + H_f \sigma_v$$

H_f is a free parameter that determines the value of standard deviation. In this study, the parameter was set to 1. The resulting feature scores are robust against outliers as the scores are retained within an interval [0,1].

3.3. Subject Level Normalization

We combine DPT with the existing subject level normalization method [39] to derive a Subject-Level Normalized Score (SLNS). The purpose of deriving an SLNS is to account for the differences in individual cognitive capabilities. We first define the questions (or tasks) pertaining to System 1 of DPT (e.g., answering a set of demographics questions). Then, use the derived FLNS (i.e., score for velocity) from System 1 tasks to normalize the other FLNS. Using participants' own response behaviors in innocuous demographics questions (i.e., System 1 related questions) has three key benefits:

1. Using FLNS allows us to preserve information regarding task-specific variability.
2. It allows us to capture the user-specific differences (i.e., across subject differences).
3. The approach also captures the difference in response behaviors between System 1 tasks and other tasks.

In summary, SLNS derived from subject-level normalizations captures the user-specific characteristics (e.g., user-specific behavior, cognitive capacity). The SLNS is given by:

$$\tilde{s} = \max [0, \min \left(1, \frac{s_{u,i} - [s_{u,sys1}]}{[s_{u,sys1}] - [s_{u,sys1}]} \right)]$$

$[s_{u,sys1}]$ and $[s_{u,sys1}]$ pertain to upper bounds and lower bounds derived from System 1 related tasks.

Subscript u pertains to each user, while $sysl$ denotes that the boundaries are derived from System 1 tasks. We derive the boundaries by:

$$\begin{aligned} [s_{u,sys1}] &= \mu_{u,sys1} - H_s \sigma_{s_{u,sys1}} \\ [s_{u,sys1}] &= \mu_{u,sys1} + H_s \sigma_{s_{u,sys1}} \end{aligned}$$

For each user, $\mu_{u,sys1}$ and $\sigma_{u,sys1}$ denote the average and a standard deviation of a feature score across System 1 tasks.

4. Methodology

We derived and tested new metrics with an experimental study. The study included a set of questions and conditions that are designed to heighten cognitive load.

4.1. Procedure and Manipulation

4.1.1. Baseline Condition. In the first part of the survey, all participants were asked to complete a demographic questionnaire. The Big Five Inventory questionnaires consisting of 39 questions that measure the Big Five personality traits (e.g., conscientiousness, extraversion, neuroticism, agreeableness, openness to experience) were subsequently presented along with an attention check question [44].

4.1.2. Manipulation (System 2) Condition. We manipulate the activation level of System 2 with two approaches. First, we primed the participants by letting them think that their answers to the second part of the study were important. Participants were given an option to voluntarily participate in the second part of the survey for additional payout (e.g., a total of \$1.00). The instruction for the second part stated, “the second part of this survey is an application for a follow-on study that will pay \$10 for 10 minutes of work. We will only select a few people to do this follow-on study. If we select you, you can refuse to participate; there is no obligation to do it. We will pay you an extra \$0.50 to complete the application, regardless of whether we select you or whether you agree to do the follow-on study if selected.” We further stated that “the follow-on study will have you use Excel to do some data analysis, so we are looking for people who have experience with Excel. Experience with Excel’s math and statistical functions is a plus but is not required.”

Second, we included a question that asks for the subject’s experience related to a Non-existent Excel tool (i.e., we named this extension, StatView). The follow-up survey consisted of multiple subparts

related to the user’s background and experience. The participants were asked to answer the questions related to their work hours and occupation (e.g., Are you currently employed in a full-time job?). When participants answered questions regarding their computer skills, we asked all participants to rate their skills on a non-existent Excel Plugin (i.e., StatView). The range for all the experience-related questions was from 0 (e.g., beginner) to 10 (e.g., expert).

4.2. Participants

534 subjects were recruited from Amazon Mechanical Turk. Participants’ age was diverse. About 56% of the participants belonged to a younger crowd between 18-34 years old, with 42% reported being 35 years or older. Of the recruited participants, 41% were female. Of those participants, three hundred and sixty-two subjects completed the second part of the survey. Attention check question was given to all participants before proceeding to the second part.

As we are mainly interested in how people answer cognitively demanding questions, we retained all the samples regardless of whether a person completed part 2 of the survey or not. This resulted in the final sample size of five hundred and thirty participants with raw mouse movement data of 2.7 million observations.

5. Results

We first evaluate the formal hypothesis H1 and then conduct a series of predictive analyses to examine H2. To examine H1, we compare the FLNS across the demographic’s questions and cognitively demanding questions (e.g., attention check question and fake software question). For H2, we examine the overall trends of SLNS and examine its relationship to FLNS. All analyses were conducted using the R statistical software package [45].

5.1. Hypothesis 1

Hypothesis 1 stated that Users involved in higher complexity tasks (i.e., questions) would exhibit slower mouse cursor speed. To examine the differences in mouse cursor velocity across all questions, we examine the FLNS. Before comparing if the FLNS vary across the questions, we first tested the homogeneity of variance assumption using Brown-Forsyth type Levene’s test. The Brown-Forsythe type Levene’s test 1 was significant on the FLNS,

variances to ensure the homogeneity of variance assumption is not

¹ Brown-Forsythe test is a statistical test to compare the group

indicating that the classic parametric F-test for ANOVA can lead to severely biased results [46]. Thus, the score means were further examined using Welch's ANOVA [47,48]. Welch's ANOVA is designed to test the group means when multiple groups are compared, and the homogeneity of variance assumption is violated [49]. Welch's ANOVA result for the FLNS is reported in Table 1.

Table 1. FLNS and SLNS Welch's ANOVA results

Measure	F	Num DF	Den DF	
FLNS	44.739	6.0	1228.2	***
SLNS	55.408	6.0	1461.9	***
Key: * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$				

Table 2. Exemplary Results for Pairwise Comparisons

Group 1	Group 2	Estimate	P val adj.	
ATTN CHECK	FAKE SOFT	-0.01	0.995	
LANGU-AGE	FAKE SOFT	-0.11	5.01 E-10	***
Key: * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$				

The result of Welch's ANOVA suggests that the average FLNS for at least one of the examined questions is statistically different. Thus, further evaluations are needed.

FLNS Pairwise Comparisons – Games Howell Post-hoc Test. For further evaluations, a series of pairwise Games-Howell Post-hoc Tests² [50] was performed to compare all possible pairs of means. As the Games-Howell test does not assume equal variances and sample sizes, the test is suitable in situations where the homogeneity of variance assumption is violated.

First, the difference in means between the cognitively demanding questions (e.g., Attention check and StatView question) was found to be minimal ($p = 0.99$). Second, the mean FLNS for the cognitively demanding questions was lower than the other demographics question. The overall results suggested that the mean differences between the System 1 questions and System 2 questions were directionally correct and were significant. Further, the Games-Howell Test comparing the differences in means between the two cognitively demanding questions failed to attain significance (see Table 2). Thus, we conclude that Hypothesis 1 was supported. We did not include the results in this paper as a total

violated.

² Games and Howell Post-hoc Tests compare all combinations

of 21 pairs of averages were compared (i.e., it takes up an entire page). The results and the plot of the averages are available upon request.

5.2. Hypothesis 2

Hypothesis 2 stated that the users will still exhibit slower mouse cursor speed when executing higher complexity tasks regardless of differences in cognitive processing capacity. Though it is difficult to perfectly measure an individual's cognitive processing capacity as various factors influence it, SLNS accommodate user-specific characteristics (i.e., including the differences in cognitive capacity).

As the Brown-Forsythe type Levene's test was significant on the SLNS, the scores were further examined using Welch's ANOVA [47]. The ANOVA result of the SLNS is shown in Table 1.

The result of Welch's ANOVA suggests that the difference between SLNS means some of the questions are statistically significant.

SLNS Pairwise Comparisons – Games Howell Post-hoc Test. To further examine Hypothesis 2, a series of pairwise Games-Howell Post-hoc Tests were performed to examine pairs of SLNS means. Even when considering user-specific information, the overall results were consistent with those from analyzing FLNS: (1) mean differences between the System 1 questions and System 2 questions were directionally correct and were significant. (2) The Games-Howell Test comparing the differences in means between the two cognitively demanding questions failed to attain significance (e.g., Attention check and Fake Software Question). Thus, we conclude that Hypothesis 2 was supported. Similar to the FLNS pairwise comparison, a total of 21 pairs of averages were compared. The results and the plot of the averages are also available upon request.

5.3. Predictive Models (10-fold cross-validation)

We chose to validate the performance of our score measures by using a subset of aggregated data (i.e., at a question level). There are several reasons why we do this. First, as the scores vary across several demographics' questions, the inclusion of all the questions would, in fact, negatively impact the performance of the predictive model. Second, as the models we are constructing are simple models with

of group differences when homogeneity of variance assumption is violated.

only two score features, including all variables would cause class imbalance problems (i.e., 5 to 2 ratio). We chose the language question of all the demographics questions as it had the highest average for both FLNS and SLNS. A total of 970 data points were used in the models, with 586 being cognitively demanding questions (i.e., attention check question and StatView question) and 384 data points belonging to language questions. Thus, the base rate for the accuracy of all models is 60.41%.

We formulated a typical binary classification problem by creating a dependent variable with labels “Language” and “Demanding”,

$$y_i = \beta_0 + \beta_1 \tilde{s}_i + \beta_2 \tilde{v}_i + \epsilon$$

where \tilde{s}_i and \tilde{v}_i are SLNS and FLNS, respectively.

The Logistic regression model, a simple but widely used model, was selected as the baseline model for this study. Other popular classification models, such as Support Vector Machine (SVM), Random Forest (Bagging), and XGBoost (Boosting), were also used for comparison. All models using original samples have been trained with 10-fold cross-validation. Average AUC, accuracy, sensitivity, and specificity across the folds are reported in Table 3 and Table 4.

Table 3. 10-fold cross-validation results (Logistic Regression and SVM)

	Logistic Regression	SVM
AUC (SD)	0.7705 (0.035)	0.7696 (0.068)
Accuracy (SD)	70.01% (6.00%)	70.22% (3.57%)
Sensitivity	0.8090	0.8173
Specificity	0.5341	0.5132

All the models examined surpassed the base rate of 60.41% accuracy, with Random Forest being the lowest (67.62%). SVM was the top-performing model, with an AUC of 0.7696, an accuracy of 70.22%, and the lowest standard deviation for accuracy (3.57%). We further examined the relationship between our newly derived measures. By constructing a plot of the variables, we were able to identify two distinct patterns that align with our hypotheses (See Figure 1). For cognitively demanding questions, both FLNS and SLNS were relatively lower than the language questions. Precisely, the patterns signal that the users exhibit slower mouse movements when engaged with more cognitively demanding tasks. Further, even when controlling for user-specific factors (i.e., including SLNS), we observe an evident pattern that suggests that users' mouse cursor movements are slower for cognitively demanding tasks.

Table 4. 10-fold cross-validation results (Random Forest and XGBoost)

	Random Forest (mtry = 2)	XGBoost
AUC (SD)	0.7170 (0.042)	0.7580 (0.046)
Accuracy (SD)	67.62% (4.43%)	69.84% (3.5%)
Sensitivity	0.7200	0.7724
Specificity	0.5653	0.5854



Figure 1. Plot of FLNS and SLNS

6. Discussion

We introduced a new measure to assess the underlying cognitive processing of an individual during task execution. The results show that users have slower mouse cursor speed when executing higher complexity tasks (H1), and such a trend is still evident even when considering user-specific differences (H2). The findings have implications for several research disciplines that examine the relationship between users' cognitive processing and their online behaviors.

6.1. Implications for Research

We contribute to the literature by theoretically explaining and empirically validating how the heightened cognitive load slows the mouse cursor speed. Specifically, we apply DPT to examine the relationship between the user's device usage patterns and the underlying cognitive processes in a broader context. We extend DPT by explaining how the activation of System 2 (i.e., deliberate, effortful, and slow) will result in slower mouse movement patterns.

Second, we present a new method to account for individual differences in cognitive capacity. DPT

helps to identify types of questions that are less dependent on System 2. We account for individual-specific factors using the user's response behaviors from these questions to normalize FLNS. In doing so, we provide a new measure indicative of the user's underlying cognitive processes (i.e., cognitive load).

Our method can potentially be applied to a wide variety of research disciplines. Specifically, our measures are helpful for various studies that examine individuals' underlying cognitive processes. For instance, our method can be applied to the IS studies that examine the relationships between cognitive load and key user behaviors (e.g., Perceived Ease of Use, User Satisfaction, Reuse Intention). Our method can also be leveraged for psychology studies that manipulate participants' cognitive resources [27,51,52,53]. In summary, we provide new methodological tools, accompanied by theory, to investigate online user behaviors in various contexts.

6.2. Implications for Practice

There are several practical implications for using a mouse cursor to measure an individual's cognitive load. Specifically, the cognitive load is strongly related to predicting customer behaviors, including acceptance of advertising [54], customer attention and attraction [55], perceptions and recall of advertising content [56], and brand attitude [57]. Using the methodology described in the paper, practitioners can design a real-time system that assesses a user's cognitive state. Our methodology also helps practitioners to establish foundations for implementing theory-driven design strategies when designing their websites. In terms of usability, managers now have a tool to identify sections that impose cognitive burdens on the users and make changes. For instance, managers may assess website churning behavior by identifying pages and interaction points with lower velocity scores. Further, negative feedbacks can be assessed along with the mouse movement data to gain more insight into where the negative interactions occurred on the website.

Velocity scores can also be analyzed to assess cognitive states that drive various online purchasing behaviors. When a user purchases on a website, the user's navigational path and overall speed at the time of purchase can be analyzed to identify a user's cognitive state (e.g., System 1 vs. System 2). The cognitive state can then be assessed to make further inferences (e.g., whether a purchase made was impulsive or not). On the contrary, practitioners can also use velocity scores to analyze whether the customer has a negative experience on the website. For instance, if the website is too difficult to navigate, it

may impose a cognitive burden on the users (i.e., users are likely to activate System 2 to figure out how to search and purchase the item). Specifically, velocity scores could potentially be used to detect which pages seem to be imposing a cognitive burden on the users.

The velocity scores could help advertisers to test the performance of their ads. Advertisers spend a considerable amount of effort to capture user attention and memory [58]. However, real-time assessment of the effectiveness of online advertisements still poses a challenge. Practitioners can use mouse movement speed to perform real-time analysis of the users' cognitive states when the users are engaged with an advertisement.

6.3. Limitations

Like all research, our work has limitations. First, our study only examined mouse cursor speeds in an online survey. As responding to an online survey consist of simple, goal-oriented tasks, we need to explore if our results would generalize in a broader context (e.g., web browsing, navigating within a system, online shopping). Second, users use multiple devices to connect to the internet. Thus, the underlying cognitive processes associated with System 1 and System 2 would manifest as different usage behaviors, requiring further explorations. Future research should examine whether our metrics can be applied to a broader range of tasks and devices. Third, our measures can benefit from outlier analysis and the data cleaning process. As the normalized measures are directly derived from the raw mouse data captured in millisecond precision, the sudden movements (e.g., satisficing and speeding through the survey) can significantly inflate the overall means. Further data cleaning work can potentially improve classification results. Finally, the DI model is still an actively evolving model. For instance, the Hybrid Dual-Process model, one of the latest developments in the psychology field, further develops the DI model and posits that a response generated under the influence of System 1 undergoes two different intuitive reasoning processes to produce (1) traditional heuristic intuitive response and (2) logical intuitive response [27,59]. As the DI model is still actively studied, future research should examine how and where DPT can enhance our understanding of a broad range of HCI contexts.

7. Conclusion

Capturing and assessing user's cognitive states had been an essential yet challenging task in an online setting. We explained how the high cognitive load manifests as the slower mouse movement speeds by

leveraging DPT. Our findings suggest that the analysis of mouse cursor speeds may enable research that could not have been easily conducted using a traditional measuring approach. The proposed method can be easily applied by both practitioners and researchers in various contexts as computer mice is a ubiquitous HCI tool that users across the globe widely adopt.

8. References

- [1] Knijnenburg, B.P., M.C. Willemsen, Z. Gantner, H. Soncu, and C. Newell, “Explaining the user experience of recommender systems”, *User Modeling and User-Adapted Interaction* 22(4–5), 2012, pp. 441–504.
- [2] Proudfoot, J.G., J.L. Jenkins, J.K. Burgoon, and J.F. Nunamaker, “More Than Meets the Eye: How Oculometric Behaviors Evolve Over the Course of Automated Deception Detection Interactions”, *Journal of Management Information Systems* 33(2), 2016, pp. 332–360.
- [3] Yuan, H., W. Xu, and M. Wang, “Can online user behavior improve the performance of sales prediction in E-commerce?”, *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, (2014), 2347–2352.
- [4] Chen, P.-Y. (Sharon), and L.M. Hitt, “Measuring Switching Costs and the Determinants of Customer Retention in Internet-Enabled Businesses: A Study of the Online Brokerage Industry”, *Information Systems Research* 13(3), 2002, pp. 255–274.
- [5] Erete, S.L., “Engaging Around Neighborhood Issues: How Online Communication Affects Offline Behavior”, *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ACM (2015), 1590–1601.
- [6] Davis, F.D., “Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology”, *MIS Quarterly* 13(3), 1989, pp. 319.
- [7] Venkatesh, V., “Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model”, *Information Systems Research* 11(4), 2000, pp. 342–365.
- [8] Byrd, M., “An Empirical Exploration OF COUNTERMEASURES IN HCI-BASED DECEPTION RESEARCH”, 2018.
- [9] Grimes, M., J. Jenkins, and J. Valacich, “Exploring the Effect of Arousal and Valence on Mouse Interaction”, 2015, pp. 15.
- [10] Hibbeln, M., J.L. Jenkins, Brigham University, et al., “How Is Your User Feeling? Inferring Emotion Through Human-Computer Interaction Devices”, *MIS Quarterly* 41(1), 2017, pp. 1–21.
- [11] Jenkins, J.L., J.G. Proudfoot, Bentley University, et al., “Sleight of Hand: Identifying Concealed Information by Monitoring Mouse-Cursor Movements”, *Journal of the Association for Information Systems* 20, 2019, pp. 1–32.
- [12] Nunamaker, J.F., D.C. Derrick, A.C. Elkins, J.K. Burgoon, and M.W. Patton, “Embodied Conversational Agent-Based Kiosk for Automated Interviewing”, *Journal of Management Information Systems* 28(1), 2011, pp. 17–48.
- [13] Burgoon, J., W.J. Mayew, J.S. Giboney, et al., “Which Spoken Language Markers Identify Deception in High-Stakes Settings? Evidence From Earnings Conference Calls”, *Journal of Language and Social Psychology* 35(2), 2016, pp. 123–157.
- [14] Anderson, B.B., C.B. Kirwan, J.L. Jenkins, D. Eargle, S. Howard, and A. Vance, “How Polymorphic Warnings Reduce Habituation in the Brain: Insights from an fMRI Study”, *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ACM (2015), 2883–2892.
- [15] Hibbeln, M., J. Jenkins, C. Schneider, J. Valacich, and M. Weinmann, “Investigating the Effect of Insurance Fraud on Mouse Usage in Human-Computer Interactions”, 2014, pp. 16.
- [16] McNicol, D., *A primer of signal detection theory.*, Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, 2005.
- [17] Eysenck, M.W., N. Derakshan, R. Santos, and M.G. Calvo, “Anxiety and cognitive performance: attentional control theory.”, *Emotion* 7(2), 2007, pp. 336.
- [18] Hu, P.J.-H., H. Hu, University of Nevada, Las Vegas, X. Fang, and University of Delaware, “Examining the Mediating Roles of Cognitive Load and Performance Outcomes in User Satisfaction with a Website: A Field Quasi-Experiment”, *MIS Quarterly* 41(3), 2017, pp. 975–987.
- [19] Schmutz, P., S. Heinz, Y. Métrailler, and K. Opwis, “Cognitive Load in eCommerce Applications—Measurement and Effects on User Satisfaction”, *Advances in Human-Computer Interaction 2009*, 2009, pp. 121494.
- [20] Mălăescu, I., and S.G. Sutton, “The effects of decision aid structural restrictiveness on cognitive load, perceived usefulness, and reuse intentions”, *International Journal of Accounting Information Systems* 17, 2015, pp. 16–36.
- [21] Freeman, J.B., and N. Ambady, “Hand movements reveal the time-course of shape and pigmentation processing in face categorization”, *Psychonomic Bulletin & Review* 18(4), 2011, pp. 705–712.
- [22] Chu, H.-C., “Potential Negative Effects of Mobile Learning on Students’ Learning Achievement and Cognitive Load—A Format Assessment Perspective”, 2021, pp. 14.
- [23] Valacich, J.S., M.D. Byrd, M. Kumar, J.L. Jenkins, D. Kim, and P.A. Williams, “Using accelerometer and gyroscope data in common mobile devices to assess credibility”, *26th Americas Conference on Information Systems, AMCIS 2020*, Association for Information Systems (2020).
- [24] Freeman, J.B., and N. Ambady, “MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method”, *Behavior Research Methods* 42(1), 2010, pp. 226–241.
- [25] Dale, R., and N.D. Duran, “The Cognitive Dynamics of Negated Sentence Verification”, *Cognitive Science* 35(5), 2011, pp. 983–996.
- [26] McKinstry, C., R. Dale, and M.J. Spivey, “Action Dynamics Reveal Parallel Competition in Decision Making”, *Psychological Science* 19(1), 2008, pp. 22–24.
- [27] Bago, B., and W. De Neys, “Fast logic?: Examining the time course assumption of dual process theory”, *Cognition* 158, 2017, pp. 90–109.

- [28] Evans, J.St.B.T., “In two minds: dual-process accounts of reasoning”, *Trends in Cognitive Sciences* 7(10), 2003, pp. 454–459.
- [29] Evans, J.St.B.T., “Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition”, *Annual Review of Psychology* 59(1), 2008, pp. 255–278.
- [30] Evans, J.St.B.T., and K.E. Stanovich, “Dual-Process Theories of Higher Cognition: Advancing the Debate”, *Perspectives on Psychological Science* 8(3), 2013, pp. 223–241.
- [31] Smith, E.R., and J. DeCoster, “Dual-Process Models in Social and Cognitive Psychology: Conceptual Integration and Links to Underlying Memory Systems”, *Personality and Social Psychology Review* 4(2), 2000, pp. 108–131.
- [32] Sowden, P.T., A. Pringle, and L. Gabora, “The shifting sands of creative thinking: Connections to dual-process theory”, *Thinking & Reasoning* 21(1), 2015, pp. 40–60.
- [33] Kahneman, D., *Thinking, fast and slow*, Macmillan, 2011.
- [34] Stanovich, K.E., *Who is rational?: Studies of individual differences in reasoning*, Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, 1999.
- [35] Stanovich, K.E., and R.F. West, “Individual differences in reasoning: Implications for the rationality debate?”, *Behavioral and brain sciences* 23(5), 2000, pp. 645–665.
- [36] Frederick, S., “Cognitive Reflection and Decision Making”, *Journal of Economic Perspectives* 19(4), 2005, pp. 25–42.
- [37] Travers, E., J.J. Rolison, and A. Feeney, “The time course of conflict on the Cognitive Reflection Test”, *Cognition* 150, 2016, pp. 109–118.
- [38] Barrett, L.F., M.M. Tugade, and R.W. Engle, “Individual Differences in Working Memory Capacity and Dual-Process Theories of the Mind.”, *Psychological Bulletin* 130(4), 2004, pp. 553–573.
- [39] Scandura, J.M., “Deterministic Theorizing in Structural Learning: Three Levels of Empiricism”, *Journal of Structural Learning*, 1991, pp. 21–53.
- [40] Halford, G.S., *Children’s understanding: The development of mental models*, Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US, 1993
- [41] Murphy, G.L., and J.C. Wright, “Changes in conceptual structure with expertise: Differences between real-world experts and novices.”, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10(1), 1984, pp. 144–155.
- [42] Wickens, C.D., R. Braune, and A. Stokes, “Age differences in the speed and capacity of information processing: I. A dual-task approach.”, *Psychology and Aging* 2(1), 1987, pp. 70–78.
- [43] Monaco, J.V., “Robust Keystroke Biometric Anomaly Detection”, *arXiv:1606.09075 [cs]*, 2017.
- [44] Chmielewski, M.S., and T.A. Morgan, “Five-Factor Model of Personality”, In M.D. Gellman and J.R. Turner, eds., *Encyclopedia of Behavioral Medicine*. Springer New York, New York, NY, 2013, 803–804.
- [45] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021.
- [46] Delacre, M., C. Leys, Y.L. Mora, and D. Lakens, “Taking Parametric Assumptions Seriously: Arguments for the Use of Welch’s *F*-test instead of the Classical *F*-test in One-Way ANOVA”, *International Review of Social Psychology* 32(1), 2019, pp. 13.
- [47] Moder, K., “How to keep the Type I Error Rate in ANOVA if Variances are Heteroscedastic”, *Austrian Journal of Statistics* 36(3), 2007.
- [48] Moder, K., “Alternatives to F-Test in One Way ANOVA in case of heterogeneity of variances (a simulation study)”, 2010, pp. 11.
- [49] Liu, H., “Comparing Welch’s ANOVA, a Kruskal-Wallis test and traditional ANOVA in case of Heterogeneity of Variance”, 2015.
- [50] Games, P.A., and J.F. Howell, “Pairwise Multiple Comparison Procedures with Unequal N’s and/or Variances: A Monte Carlo Study”, 1976, pp. 14.
- [51] Miyake, A., N. Friedman, D. Rettinger, P. Shah, and M. Hegarty, “How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analyses”, *Journal of Experimental Psychology: General* 130, 2001, pp. 621–640.
- [52] Newman, I.R., M. Gibb, and V.A. Thompson, “Rule-based reasoning is fast and belief-based reasoning can be slow: Challenging current explanations of belief-bias and base-rate neglect.”, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43(7), 2017, pp. 1154–1170.
- [53] Pennycook, G., J.A. Fugelsang, and D.J. Koehler, “Are we good at detecting conflict during reasoning?”, *Cognition* 124(1), 2012, pp. 101–106.
- [54] Wright, P.L., “The Cognitive Processes Mediating Acceptance of Advertising”, *JOURNAL OF MARKETING RESEARCH*, 1973, pp. 10.
- [55] Wojdyski, B.W., and H. Bang, “Distraction effects of contextual advertising on online news processing: An eye-tracking study.”, *Behaviour & Information Technology* 35(8), 2016, pp. 654–664.
- [56] Nasco, S.A., and G.C. Bruner, “Perceptions and Recall of Advertising Content Presented on Mobile Handled Devices”, *Journal of Interactive Advertising* 7(2), 2007, pp. 51–62.
- [57] Pantoja, F., P. Rossi, and A. Borges, “How Product-Plot Integration and Cognitive Load Affect Brand Attitude: A Replication”, *Journal of Advertising* 45(1), 2016, pp. 113–119.
- [58] Kong, S., Z. Huang, N. Scott, Z. Zhang, and Z. Shen, “Web advertisement effectiveness evaluation: Attention and memory”, *Journal of Vacation Marketing* 25(1), 2019, pp. 130–146.
- [59] Bago, B., and W.D. Neys, “Advancing the specification of dual process models of higher cognition: a critical test of the hybrid model view”, *Thinking & Reasoning* 26(1), 2020, pp. 1–30.