Detecting Social Desirability Bias with Human-Computer Interaction: A Mouse-Tracking Study

Paul A. Weisgarber
The University of Arizona
paulweisgarber@arizona.edu

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

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

David Kim
The University of Arizona
davidkim7@arizona.edu

Manasvi Kumar
The University of Arizona
manasvik@arizona.edu

Abstract

Social desirability bias undermines self-report accuracy, necessitating novel approaches to detect and mitigate its impact. This study aimed to investigate the influence of social desirability on questionnaire responses by analyzing mouse cursor movements and answering behaviors. Respondents (n=238) completed a health and wellness questionnaire while their mouse cursor data was recorded. The results revealed that individuals under a higher social desirability treatment exhibited significantly longer response times and slower mouse cursor speeds, supporting the hypothesis that they may engage in more cautious and deliberate responding. However, no significant differences were found in terms of mouse cursor deviations or answer switches between the two groups. These findings suggest that analyzing mouse cursor movements can provide valuable insights into the influence of social desirability bias on questionnaire responses, offering a potentially scalable method for detection and future intervention.

Keywords: HCI dynamics, mouse cursor movements, online survey research, self-report data, social desirability response bias.

1. Introduction

Various data collection methods, including observations and surveys, have been employed by researchers and practitioners for decades to obtain information about individuals, groups, and contexts (McGrath et al., 1982). Among these methods, *observation* has been widely utilized across different disciplines to gather data on people, processes, and cultures (Kawulich, 2005). For example, observational studies and in-person checkups have been instrumental in identifying potential risk factors for various diseases and mental health disorders in medical research and practice (Ligthelm et al., 2007). Furthermore, observational studies have enabled researchers to gain profound insights into the psychological and social behavior of both individuals and groups (Shaughnessy et al., 2012). While observation offers a highly personalized perspective, it is not without limitations.

For example, observational data is confined to the behaviors and events that can be directly observed and recorded by the researcher, which may not always fully capture the underlying construct of interest (Straub et al., 2004). In addition, observational data is limited in its scalability, making it difficult to observe every person in large populations (Bhattacherjee, 2012). While reducing a study's sample to a subset of the overall population may be less resource-intensive, it may also result in an incomplete or biased understanding.

Surveys were developed as a means of collecting data beyond the restrictive limitations of observation – enabling greater external validity via data collection from the entire population (Fowler Jr., 2013). Additionally, the advent of the internet has made *online surveys* a scalable method to remotely collect data from large populations. However, online self-reported survey data also suffer from various limitations.

First, unmeasured biased responses, stemming from social desirability (SD) bias, can compromise both the internal validity of data collection (Bhattacherjee, 2012) and the accuracy of online questionnaires and survey results (Fowler Jr., 2013). This bias involves respondents under-reporting socially undesirable behaviors and over-reporting socially desirable behaviors. It is particularly evident with sensitive topics and contexts with perceived high stakes, such as health and wellness questionnaires that could impact one's occupation (Marquis et al., 1986). For example, Hoffman et al. (2023) found that 72% of U.S. military pilots under-reported their health and wellness behaviors due to fear of losing their flying status. As a result, ill-advised conclusions by not only researchers but also practitioners, managers, and healthcare providers may result when this response bias manifests itself in the data collection process (Kwak et al., 2019).

URI: https://hdl.handle.net/10125/106946
978-0-9981331-7-1

Second, survey data (particularly in the form of online questionnaires) have another inherent limitation: they fail to capture certain human-computer interaction (HCI) dynamics, which encompass respondents' digital records of their interactions with technology (Valacich et al., 2022). In other words, the data reflect solely the respondent's *submitted* answer, without capturing the observational behaviors exhibited *during* the answering process (e.g., hesitations, answer switches, etc.). This drawback leads to the loss of valuable real-time and contextualized behavioral data that is often rich with valuable insights (Weinmann et al., 2022).

To address the limitations of surveys and observations, our methodology combines the detailed insights of observational data with the broad reach of online surveys to create *observations at scale*. Using various metrics derived from HCI devices, our method covertly observed respondents' answering behaviors within a self-reported, survey-based health and wellness assessment. This included capturing a respondent's HCI dynamics (e.g., mouse movement metrics) before their final answer selection and submission, providing real-time insights into the cognitive and emotional state of the respondent *while* they formulated their responses. This approach seeks to overcome the limitations of previous research methods and provide a more accurate understanding of the state of the respondent. Ultimately, we explore the following research question: *Do respondents' mouse movements differ in health and wellness surveys when social desirability is manipulated through two experimental conditions*?

In this study, we analyze HCI dynamics in pursuit of two objectives. First, we demonstrate that mouse cursor dynamics can be used to identify respondents' SD bias while answering health and wellness questions. Second, we aim to extend the existing social science literature regarding SD response bias identification and mitigation techniques. As a result, we seek to advance the understanding of SD bias and improve the accuracy of self-reported health questionnaire data.

2. Literature Review

This section provides a broad understanding of measuring and minimizing SD bias in self-reported surveys. It reviews widely used SD measurement scales and their limitations, examines mitigation techniques to combat SD bias, and highlights the potential of HCI dynamics, specifically mouse cursor movements, for detecting subtle cognitive influences, including those associated with social desirability.

2.1. Social Desirability: Measurement Scales

SD response bias in self-reported questionnaires has been a topic of interest for researchers for decades. Early studies by Gordon and Edwards highlighted the impact of respondents' tendency to select socially acceptable alternatives instead of truthful options (Gordon, 1951; Edwards, 1957). The Marlowe-Crowne SD scale improved these early measurements of SD by focusing on everyday events (Crowne & Marlowe, 1960). Additional scales such as the Responding Desirably-16 (RD-16), Balanced Inventory of Desirable Responding (BIDR), and its variations, have also been developed over the years (Hart et al., 2015; Paulhus, 1988; Schuessler et al., 1978; Steenkamp et al., 2010).

These scales assume that respondents who endorse statistically infrequent statements (e.g., "I never resent being asked to return a favor") have a greater need for approval, given that the described behavior is highly unlikely or nearly impossible (Hart et al., 2015). As a result, a respondent's score on an SD scale can be useful to measure whether the content instrument was influenced by SD response bias (Paulhus, 1991).

However, these scales are seldom employed by researchers within the information systems domain – even within its subdisciplines that focus on sensitive topics (e.g., information security) (Gergely & Rao, 2016). Furthermore, these scales are even less utilized in non-academic settings (e.g., a medical provider's analysis of a patient's health and wellness survey, a marketing professional's analysis of consumer sentiments, etc.) (Larson, 2019). We posit that these trends may be attributed to various limitations of these scales. Table 1 summarizes the previously mentioned SD scales with a brief description of their limitations.

Reviewing these scales reveals some key trends that may explain their limited use in IS research and real-world contexts. First, the longer scales, such as the Edwards, Marlowe-Crowne, and BIDR, have been found to lead to respondent fatigue and reduce a respondent's motivation to answer truthfully (Fowler Jr., 2013). Second, the shorter scales, such as the RD-16, BIDR-20, and BIDR-16, do not always capture the full range of socially desirable behaviors or produce

Table 1. Social Desirability Measurement Scales

Reference	Scale	No. of Items	Limitations
Edwards, 1957	Edwards	39	Critiqued for its length and potential confounders.
Crowne & Marlowe, 1960	Marlowe-Crowne	33	Limited by its length and dated wording.
Schuessler et al., 1978	RD-16	16	Criticized due to its low reliability.
Paulhus, 1984	BIDR	40	Limited by its length.
Steenkamp et al., 2010	BIDR-20	20	Exhibited relatively low reliability.
Hart et al., 2015	BIDR-16	16	Exhibited relatively low reliability.

inconsistent results due to the limited item construction (Nederhof, 1985; Nunnally, 1978). As a result, these scales may lack practicality in accurately measuring social desirability response bias.

2.2. Social Desirability: Mitigation Techniques

To overcome the limitations associated with SD measurement scales, researchers have identified effective SD mitigation strategies to reduce this response bias. Two prominent mitigation techniques include anonymous survey administration and response motivations that prioritize honesty.

First, anonymity assurances and limited opportunities for identification have been found to decrease SD bias (Grimm, 2010). Negative judgments from others often drive socially desirable responses (Schaeffer, 2000), and – in settings where respondents are assured anonymity – they are less likely to provide socially desirable responses. Second, subtle cues in the survey's instructions, such as "we value your opinion" or explicit requests for honest feedback such as "tell us what you think," have been shown to increase the likelihood of obtaining accurate and less socially desirable responses (Podsakoff et al., 2012). For example, Gordon's (1987) experiment manipulated questionnaire instructions to mitigate SD bias. Standard instructions (i.e., the control group) for dental hygiene questions were modified (i.e., the treatment group) to emphasize anonymity, accuracy, and respondents' value as contributors. This adjustment in the instructions notably decreased the inclination to overreport socially desirable behavior in the treatment group.

Ultimately, these mitigation techniques aim to enhance the quality of the respondents' final, recorded responses by reducing the over-reporting of socially desirable behavior and under-reporting of socially undesirable behaviors. However, it is important to note that the effectiveness of these strategies can be influenced by specific contexts. For instance, in the case of online health questionnaires administered by a respondent's doctor, complete anonymity may not always be assured. Furthermore, for patients who are inclined to provide dishonest responses, regardless of the instructions, reminders of accurate response motivations may have a limited impact.

Despite the proven usefulness of these mitigation strategies in certain contexts, the IS discipline appears to underutilize them. A review conducted by Kwak et al. (2019) found that among the 1,679 papers published in the Basket of Six IS journals from 2011 to 2017, 26% used self-reported measures. Among the survey-based papers, only 5% attempted to address SD bias, with just 2% utilizing formal detection or control methods. These results indicate that SD bias is infrequently investigated in IS literature, and even when acknowledged, it may not be adequately mitigated.

Therefore, we believe that there is an opportunity to introduce a novel, non-invasive, and ubiquitous approach to measure social desirability response bias. By analyzing respondents' HCI dynamics, we can potentially provide a more comprehensive and accurate assessment of SD bias.

2.3. Measuring Cognitive Processes and Mouse Cursor Movements in Healthcare Surveys

The use of HCI dynamics as a methodology has gained popularity due to its potential to provide valuable insights into an individual's decision-making and psychological processes (Valacich et al., 2022). One method of collecting this data is through the tracking of mouse cursor movements. Freeman and colleagues (2011) demonstrated how the "movements of the hand…offer continuous streams of output that can reveal ongoing dynamics of processing, potentially capturing the mind in motion with fine-grained temporal sensitivity." Subsequently, mouse cursor tracking has been increasingly utilized in a broad range of cognitive and emotional research studies.

Although recent HCI research has explored related areas, such as response biases (Jenkins et al., 2017; Kumar et al., 2021), dishonest answering behaviors due to embarrassment (Masters et al., 2022), faked identity detection (Monaro et al., 2017), and faking good behavior of one's personality (Mazza et al., 2020), an opportunity remains to examine the impact of SD response bias on answering behaviors in health and wellness surveys. For example, Hoffman et al. (2023) revealed the prevalence of healthcare avoidance behaviors, finding that 72% of U.S. military pilots admitted to misreporting that their health status was socially acceptable to preserve their flying status.

Numerous studies, outside of the U.S. military, have corroborated these findings. Levy et al. (2018) discovered that a significant portion of patients – over 80% – falsify information about exercise frequency, dietary habits, and other health-related behaviors to avoid judgment from their healthcare providers. Additionally, in a more recent investigation, nearly half of the participants admitted to misrepresenting their adherence to COVID-19 public health measures (Levy et al., 2022). Thus, our study seeks to capitalize on HCI dynamics to examine SD response bias in the context of health and wellness surveys. We believe analyzing mouse movement in this context can deepen our understanding of individual responses and could potentially enhance healthcare screening tools.

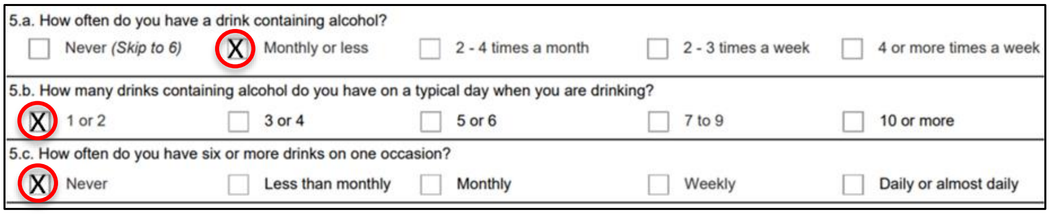
5.a. How often do you have a drink containing alcohol?
Never (Skip to 6) | X Monthly or less | 2 - 4 times a month | 2 - 3 times a week | 4 or more times a week
5.b. How many drinks containing alcohol do you have on a typical day when you are drinking?
X 1 or 2 | 3 or 4 | 5 or 6 | 7 to 9 | 10 or more
5.c. How often do you have six or more drinks on one occasion?
X Never | Less than monthly | Monthly | Weekly | Daily or almost daily

Figure 1. Hypothetical After-Submission Responses from an Online Health Assessment

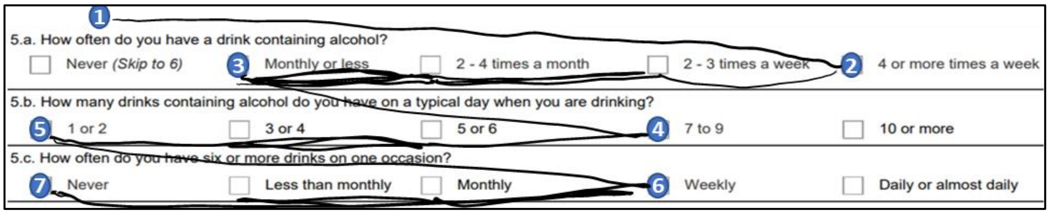
1
5.a. How often do you have a drink containing alcohol?
Never (Skip to 6) | 3 Monthly or less | 2 - 4 times a month | 2 - 3 times a week | 2 4 or more times a week
5.b. How many drinks containing alcohol do you have on a typical day when you are drinking?
5 1 or 2 | 3 or 4 | 5 or 6 | 4 7 to 9 | 10 or more
5.c. How often do you have six or more drinks on one occasion?
7 Never | Less than monthly | Monthly | 6 Weekly | Daily or almost daily

Figure 2. Hypothetical Behavioral, Real-Time Responses from an Online Health Assessment

To illustrate the potential of observing answering behavior using HCI dynamics, consider the hypothetical results from a selected portion of an online health assessment shown in Figure 1. Based solely on the final answers (X), a healthcare provider would likely categorize this patient's health as "Routine."

However, in Figure 2, we illustrate the informative potential of tracking answering behavior using mouse dynamics. This insight could lead providers to reconsider a respondent's health status classification, like shifting from "Routine" to "Priority." The figure depicts a respondent's navigation and selection process, highlighting moments where initial choices were revised to more socially desirable options. Such patterns are echoed in questions 5.b. and 5.c., revealing indecision and low confidence, evident in answer switches, navigation paths, hesitations, etc. These nuanced cues furnish observational data that can offer healthcare providers new insights beyond their current knowledge.

Overall, analyzing these mouse movement and cognitive processing metrics can enhance our understanding of how individuals respond to health and wellness questionnaires, leading to the development of more effective screening tools.

3. Theory Development, Research Question, and Hypotheses

Considering the practical limitations of social desirability response bias measurement scales and the contextual appropriateness of social desirability mitigation techniques, as well as the advancements in HCI dynamics, we situate these components within the construct of cognitive control to develop our research question and hypotheses.

3.1. Theory Development

While individuals generally strive to maintain a positive self-image (Paulhus, 1988), particularly when seeking social acceptance and moral approval (Aronson, 1969; Harris et al., 1976), one line of research suggests that a respondent's *initial* impulse is to answer objectively and honestly (Wright, 1994). Then, respondents use *cognitive control* to override their original answer to align with socially desirable norms (Baumeister et al., 2005). Cognitive control is slow and effortful, requiring sustained attention and working memory capacity to actively override impulse responses (Evans, 2008). As a result, it has been demonstrated that *cognitive control* is responsible for enabling respondents to behave in ways that maintain a positive perception by others, such as engaging in socially desirable behavior (Pitesa et al., 2013). Baumeister and Juola Exline likened cognitive control to "the moral muscle" (1999, p. 1165) that guides individuals toward socially desirable behavior, even when their natural impulses might lead to an initial, honest response.

Advances in measuring HCI dynamics enable researchers to infer differences in cognitive control. For example, in situations where users experience increased cognitive load, such as when they are concealing information (Jenkins et al., 2019), they tend to *take longer* and *move slower* while completing tasks that involve using a mouse or keyboard. Additionally, users may exhibit *increased deviations* in their mouse movements while processing inauthentic information (Jenkins et al., 2019). Therefore, we aim to expand upon this research by applying the theory of cognitive control to socially desirable contexts, specifically by measuring respondents' social desirability tendencies through their computer mouse movements.

3.2. Research Question and Hypotheses

Drawing upon cognitive control, we propose that when self-report survey instructions are designed to elicit a *higher* socially desirable response (by emphasizing non-anonymity and abstaining from honesty reminders), respondents will exhibit more deliberation and less confidence in their answering behaviors. Conversely, when instructions are designed to elicit a *lower* socially desirable response (by emphasizing anonymity and honest responses), we predict that respondents will exhibit less deliberation and more confidence. Based on these considerations, we investigate the following hypotheses:

H1: Respondents in the higher social desirability group will exhibit longer response times compared to the respondents in the lower social desirability group.

H2: Respondents in the higher social desirability group will exhibit greater mouse cursor deviations compared to the respondents in the lower social desirability group.

H3: Respondents in the higher social desirability group will exhibit slower mouse cursor speeds compared to the respondents in the lower social desirability group.

H4: Respondents in the higher social desirability group will exhibit more answer switches compared to the respondents in the lower social desirability group.

These hypotheses enable us to investigate how SD tendencies relate to respondents' cognitive control processes, as evidenced by their mouse movements.

4. Methodology

Our experimental study implemented a carefully designed set of instructions to induce either a higher or lower level of social desirability bias among respondents while they answered questions related to their health and wellness behaviors. Respondents were randomly assigned to one of two groups, each receiving different instructions. Group 1 was informed that their non-anonymous responses would be evaluated by the survey administrators, while Group 2 was encouraged to provide honest and anonymous responses. In the former group, the instructions aimed to increase the perceived social pressure to provide SD responses, leading to a heightened level of SD response bias. In the latter group, the instructions aimed to minimize any perceived social pressure to provide SD responses. The instructions were adapted from Gordon’s (1987) social desirability demonstration and edited by three senior IS scholars and three IS Ph.D. students.

4.1. Sample

The survey was initially sent to 427 students. Of these, 257 students started the survey, and ultimately, 238 students answered all manipulation check questions correctly and completed the entire survey. All respondents were junior-level business students at a U.S. public university, primarily aged 19 to 21 (90.8%) and with 56.3% identifying as female. As the present study aimed to investigate biased response behaviors related to health and wellness habits – a subject in which students may exhibit SD tendencies – and given that external validity was not the primary goal of this investigation (Compeau et al., 2012), a student sample was an appropriate choice for this research. Each respondent received extra credit for their participation.

4.2. Procedure

Respondents were randomly assigned into two groups: Group 1 (n = 113) or Group 2 (n = 125). Figure 3 presents the entire experiment protocol, while the rest of this section explains each phase.

Each group received distinct instructions before completing the same health-related questionnaire. In Group 1 (i.e., the higher SD bias group), we used the following instructions:

We require ***your name, student ID, and university email****, so we can* ***properly identify you*** *and* ***report your answers****. We will analyze each of your answers, so you must complete the entire survey to receive credit.*

In Group 2 (i.e., the lower SD bias group), we used the following instructions:

One of the ***most important*** *tasks in behavioral health care is* ***to collect accurate information****.*

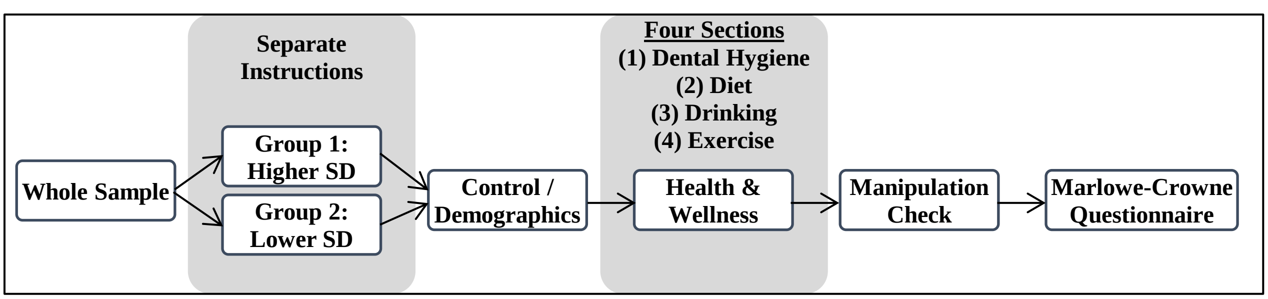

Figure 3. Experiment Protocol

Therefore, please answer the following questions ***honestly****. The results of the survey are* ***completely anonymous****. We will also ask for your student ID and university email only to ensure proper course credit. You must complete the entire survey to receive credit.*

After reading one of the two randomly selected sets of instructions, each respondent answered identical demographic questions followed by a series of identical health and wellness survey questions. These questions consisted of four sections (e.g., dental hygiene, exercise, diet, and drinking habits) with four items per section. An example question would be: "How often do I brush my teeth?" All answer choices for each question were presented using a five-point Likert scale. For example, the answer choices to the previous question would be "Less than once a day," "Once a day," "Twice a day," "Three times a day," and "More than three times a day." Attention-check questions were also included to monitor response quality.

To assess the internal validity of the study's design, we then conducted a manipulation check to evaluate the efficacy of our instructions of inducing differing levels of perceived anonymity between the two groups. We asked respondents from both groups the following: "I felt my answers were anonymous." Respondents were required to select one of five options, with (1) = "Strongly disagree" to (5) = "Strongly agree."

Next, each respondent completed the Marlowe-Crowne Social Desirability (MCSD) questionnaire (Crowne & Marlowe, 1960) to provide an additional measure of the respondents' SD tendencies. Despite ongoing debates about the most effective SD scales, the MCSD remains one of the most widely used scales (Barger, 2002). Scholars have maintained that the MCSD's popularity is due to its ability to accurately detect SD behaviors (Lambert et al., 2016) and its robust validity and reliability across diverse population groups (Loo & Loewen, 2004). Therefore, we employed the 33-question (True-False) MCSD scale to provide an additional measure of all respondents' SD tendencies.

Finally, following the completion of the experiment, the respondents were debriefed about the study's purpose and provided with assurance that their responses to all questions were completely anonymous.

4.3. Measures

During the experiment, we passively collected mouse movement data using an embedded JavaScript that recorded the x- and y-coordinates of each respondent's mouse movements. This method, which has been commonly employed in previous research (Mathur & Reichling, 2019; Valacich et al., 2022), allowed us to capture a wide range of interaction events such as mouse clicks and movements at millisecond precision. These captured events provided rich HCI dynamics that could be further processed to examine our hypotheses and obtain a more accurate and nuanced understanding of *how* SD response bias can be measured during online surveys.

Table 2. Mousing Dynamics

Metric	Description
Response Times (ms)	The time spent interacting with the target questions.
Mouse Cursor Deviations (px)	The difference between the actual distance covered by a respondent's mouse cursor and the distance it would have taken along the idealized response trajectory.
Mouse Cursor Speeds (px/ms)	The average speed of the mouse cursor while interacting with the target questions.
Answer Switches	The number of times the respondent switched between answers on the target questions.

Using the raw data from JavaScript, we adopted a comprehensive approach to calculate four metrics for each respondent: response times, mouse cursor deviations, mouse cursor speeds, and answer switches. While the details of extracting these metrics from Qualtrics are outside the scope of this paper, we followed the methodology and recommendations put forth by Valacich et al. (2022), which can be readily incorporated into an organization's survey platform to capture user behavior. Table 2 provides a summary of these metrics.

5. Results

5.1. Manipulation Check

We assessed the effectiveness of our instructions in inducing different levels of perceived anonymity. Group 1 had a mean score of 3.45 on the statement "I felt my answers were anonymous," while Group 2 had a mean score of 4.03. The difference was statistically significant ($t(236) = -4.32$; $p < .001$; Cohen's $d = 0.56$), indicating that Group 1 perceived their responses as less anonymous and may have been more influenced by SD.

5.2. Analysis and Results

We first examined the demographic characteristics between the two groups and found no significant differences in gender, age, first language, or marital status. Second, we calculated the median (M) and median absolute deviation (MAD) of each metric across all respondents for the 16 health and wellness questions. Following the approach of Kumar et al. (2021), we removed these behavioral outliers by excluding data points in which a respondent's metric exceeded M + 3 * MAD. Third, since the distributions for each of the four metrics in our hypotheses deviated from normality, a log transformation was applied to the data. This transformation was carried out with the aim of approximating normality, following the approach recommended by Raschka et al. (2022). Finally, as all respondents answered each of the 16 health and wellness questions, we used a linear mixed-effects model to examine our hypotheses – predicting each hypothesized metric based on the treatment group (0 = Group 1; 1 = Group 2) nested within each respondent.

5.2.1. Hypothesis 1 Results: Response Times

To test Hypothesis 1, we analyzed if respondents in the higher SD group (Group 1) exhibited longer response times compared to those in the lower SD group (Group 2). The results supported H1: respondents in the higher SD group demonstrated longer response times compared to respondents in the lower SD group (β = 0.041, t-value = 1.822; p = .035).

5.2.2. Hypothesis 2 Results: Mouse Cursor Deviations

To test Hypothesis 2, we analyzed if respondents in the higher SD group exhibited greater mouse cursor deviations compared to those in the lower SD group. The results failed to support H2: respondents in the higher SD group did not demonstrate greater mouse cursor deviations compared to respondents in the lower SD group (β = -0.016, t-value = -0.457; p = .324).

5.2.3. Hypothesis 3 Results: Mouse Cursor Speeds

To test Hypothesis 3, we analyzed if respondents in the higher SD group exhibited slower mouse cursor speeds compared to those in the lower SD group. The results supported H3: respondents in the higher SD group demonstrated slower mouse cursor speeds compared to respondents in the lower SD group (β = -0.023, t-value = -3.085; p = .001).

5.2.4. Hypothesis 4 Results: Answer Switches

To test Hypothesis 4, we analyzed if respondents in the higher SD group exhibited more answer switches compared to those in the lower SD group. The results failed to support H4: respondents in the higher SD group did not demonstrate greater answer switches compared to respondents in the lower SD group (β = -0.005, t-value = -0.194; p = .423). Table 3 summarizes the results of our hypotheses.

5.2.5. Exploratory Comparison: MCSD Scores

In addition to our hypothesis analysis, we performed an independent t-test to examine the potential difference in the Marlowe-Crowne Social Desirability (MCSD) scores between the two groups. The results indicated that there was no significant difference observed: $t(236) = 1.35$; p = .178; Group 1 mean = 16.73 (sd = 5.27); Group 2 mean = 15.81 (sd = 5.24).

Table 3. Summary of Hypotheses

Hypothesis	Result	p-value
H1: Respondents in the higher SD group will exhibit longer response times compared to the respondents in the lower SD group.	Supported	.035
H2: Respondents in the higher SD group will exhibit greater mouse cursor deviations compared to the respondents in the lower SD group.	Not Supported	.324
H3: Respondents in the higher SD group will exhibit slower mouse cursor speeds compared to the respondents in the lower SD group.	Supported	.001
H4: Respondents in the higher SD group will exhibit more answer switches compared to the respondents in the lower SD group.	Not Supported	.423

It is important to note that the MCSD scores were not a central aspect of our study; rather, they were included as an additional measure to explore potential influences on the observed results. Therefore, the lack of significant difference in MCSD scores does not invalidate the results of the aforementioned hypotheses.

6. Discussion

This research contributes to the development of a potentially scalable method for detecting social desirability bias in self-report surveys. By analyzing mouse cursor movements and answering behaviors, we gain insights into the impact of social desirability on participants' responses. This approach offers a novel way to identify and address the influence of social desirability on self-report measures.

6.1. Theoretical Implications

Our study contributes to the existing research on the construct of cognitive control and its association with socially desirable behavior. We provide empirical evidence that illuminates distinct response patterns among individuals who may be more prone to SD bias, as demonstrated in an experiment manipulating perceived anonymity levels. Specifically, we found that respondents in the higher SD group, where anonymity was reduced, exhibited marginally longer response times and significantly slower mouse cursor speeds. These findings suggest that individuals in this group adopted a more cautious and deliberate approach in their responses, potentially to maintain a positive perception among others. Thus, our study highlights the influence of SD on individuals' response behaviors in a health and wellness questionnaire context.

These findings align with the concept that respondents engage their cognitive control mechanisms to align their answers with socially desirable norms. The observed longer response times and slower mouse movements suggest that individuals in the higher social desirability group take additional time and exert conscious effort to override their initial responses and conform to socially desirable standards.

Importantly, our results extend the existing theory by providing objective evidence of cognitive control in the context of self-report measures. By analyzing mouse cursor movements, we gained insight into the cognitive processes underlying SD bias, offering a unique perspective on the impact of cognitive control on questionnaire responses.

Overall, our findings contribute to a deeper understanding of the mechanisms behind SD behavior and provide valuable insights for detecting potential biases in self-report measures.

6.2. Practical Implications

There are several practical implications for utilizing mouse cursor metrics as a measure of an individual's social desirability bias. First, this method offers a key strength in terms of *scalability*. Unlike traditional approaches that rely on lengthy or potentially invalid and unreliable scales, the use of objective measures like mouse cursor movements enables researchers to overcome these limitations and gather more reliable and objective data. This scalability allows for the application of the method in a wide range of contexts and settings, including large-scale studies and online assessments.

Second, the analysis of mouse cursor movements provides valuable information about participants' cognitive processes and decision-making strategies while answering questions (i.e., *how* they answered in addition to *what* they answered). For example, the findings related to response times and mouse cursor speeds reveal that individuals with higher SD tend to exhibit longer response times and slower mouse cursor speeds, indicating a more cautious and deliberate approach to answering questions. These metrics could then be used by researchers and practitioners (e.g., medical providers) to gain deeper insights into respondents' underlying motivations and decision-making processes. This, in turn, enables the development of more robust assessment tools, improves data quality, and enhances the understanding of respondents' true attitudes and behaviors.

Third, the use of mouse cursor metrics to detect SD bias offers an intriguing opportunity to integrate dynamic questionnaire features that adapt based on respondent's answering behaviors alongside their responses. Similar to current systems, in which follow-up questions appear based on specific responses, this method could introduce follow-up questions driven by respondents' mousing metrics. This innovative approach could potentially yield a more nuanced comprehension of respondents' attitudes and behaviors.

6.3. Limitations

Despite the insights provided by this study, several limitations should be acknowledged. First, the sample used in this research consisted of participants from a specific demographic (i.e., college students), which may limit the generalizability of the findings to other populations. Notably, the inherent trust disposition of a student population towards claims of anonymity might differ from other demographics. Future studies could aim for a more diverse and representative sample to enhance the external validity of the results.

Second, this study focused on a specific set of health and wellness questions; therefore, it would be beneficial to replicate the findings across different domains (e.g., sensitive political or social issues, cybersecurity policy compliance, etc.) and assess the robustness of the observed effects.

Third, while our study demonstrates the potential of mouse tracking to identify SD bias, we acknowledge the need for additional studies to rule out alternative explanations for the differences in the mousing dynamics that may go beyond SD tendencies (e.g., feelings of embarrassment, fear of retribution, etc.).

Fourth, it is important to note that the metrics employed in this study can be inherently variable across subjects. While this variance has proven to be highly salient across certain demographics (i.e., elderly population vs. a younger, tech-savvy population), this issue can be addressed by using a within-subjects design. Future studies could incorporate analyses that explore response time differences within *innocuous questions* (e.g., demographic questions) and *target questions* (e.g., health and wellness questions) of the same respondent, aiming to control how quickly a respondent answers in general. By addressing these limitations, future studies can build upon this research and provide a more comprehensive understanding of SD bias and its impact on questionnaire responses.

7. Conclusion

This study presented a novel and potentially scalable approach for detecting and addressing social desirability bias in self-report questionnaires. By analyzing mouse cursor movements and answering behaviors, we gain valuable insights into the influence of social desirability on participants' responses. The findings suggest that individuals with higher SD may exhibit significantly longer response times and slower cursor speeds, indicating a more cautious and deliberate approach to answering questions. However, no significant differences were observed in terms of mouse cursor deviations and answer-switching behaviors between the SD groups. These results highlight the importance of incorporating multiple behavioral indicators to capture the complex nature of social desirability bias. By utilizing objective measures, this method offers a promising avenue for improving the validity and reliability of screening questionnaires by addressing the impact of social desirability bias.

8. References

Aronson, E. (1969). The Theory of Cognitive Dissonance: A Current Perspective. In *Advances in Experimental Social Psychology* (Vol. 4, pp. 1–34). Elsevier. https://doi.org/10.1016/S0065-2601(08)60075-1

Barger, S. D. (2002). The Marlowe-Crowne Affair: Short Forms, Psychometric Structure, and Social Desirability. *Journal of Personality Assessment*, *79*(2), 286–305. https://doi.org/10.1207/S15327752JPA7902_11

Baumeister, R. F., DeWall, C. N., Ciarocco, N. J., & Twenge, J. M. (2005). Social exclusion impairs self-regulation. *Journal of Personality and Social Psychology*, *88*(4), 589–604. https://doi.org/10.1037/0022-3514.88.4.589

Baumeister, R. F., & Juola Exline, J. (1999). Virtue, Personality, and Social Relations: Self-Control as the Moral Muscle. *Journal of Personality*, *67*(6), 1165–1194. https://doi.org/10.1111/1467-6494.00086

Bhattacherjee, A. (2012). *Social Science Research: Principles, Methods, and Practices* (Second Edition). Textbooks Collection.

Crowne, D., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*(4), 349–354. https://doi.org/10.1037/h0047358

Edwards, A. L. (1957). The social desirability variable in personality assessment and research. *Dryden Press*.

Evans, J. St. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, *59*(1), 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Fowler Jr., F. J. (2013). *(2013). Survey Research Methods*. Sage Publications.

Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in Motion Reveals Mind in Motion. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00059

Gergely, M., & Rao, V. S. (2016). Social desirability bias in software piracy research: Evidence from pilot studies. *2016 12th International Conference on Innovations in Information Technology (IIT)*, 1–4. https://doi.org/10.1109/INNOVATIONS.2016.7880052

Gordon, L. V. (1951). Validities of the forced-choice and questionnaire methods of personality measurement. *Journal of Applied Psychology*, *35*(6), 407–412. https://doi.org/10.1037/h0058853

Gordon, R. A. (1987). Social Desirability Bias: A Demonstration and Technique for Its Reduction. *Teaching of Psychology*, *14*(1), 40–42. https://doi.org/10.1207/s15328023top1401_11

Grimm, P. (2010). *Social desirability bias*. Wiley international encyclopedia of marketing.

Harris, S., Mussen, P., & Rutherford, E. (1976). Maturity of Moral Judgment. *The Journal of Genetic Psychology*, *128*(1), 123–135. https://doi.org/10.1080/00221325.1976.10533980

Hart, C. M., Ritchie, T. D., Hepper, E. G., & Gebauer, J. E. (2015). The Balanced Inventory of Desirable Responding Short Form (BIDR-16). *SAGE Open*, *5*(4), 1–9. https://doi.org/10.1177/2158244015621113

Hoffman, W., Aden, J., Barbera, D., & Tvaryanas, A. (2023). Self-Reported Health Care Avoidance Behavior in U.S. Military Pilots Related to Fear for Loss of Flying Status. *Military Medicine*, *188*(3–4), e446–e450. https://doi.org/10.1093/milmed/usac311

Jenkins, J. L., Proudfoot, J. G., Valacich, J. S., Grimes, G. M., & Nunamaker, J. F. Jr. (2019). Sleight of Hand: Identifying Concealed Information by Monitoring Mouse-Cursor Movements. *Journal of the Association for Information Systems*, *20*, 1–32. https://doi.org/10.17705/1jais.00527

Jenkins, J. L., Valacich, J. S., & Williams, P. (2017). Human-Computer Interaction Movement Indicators of Response Biases in Online Surveys. *ICIS 2017 Proceedings*, 12.

Kawulich, B. B. (2005). Participant Observation as a Data Collection Method. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, *6*(2).

Kumar, M., Kim, D., Valacich, J., Jenkins, J., & Dennis, A. (2021). Improving the Quality of Survey Data: Using Answering Behavior as an Alternative Method for Detecting Biased Respondents. *SIGHCI 2021 Proceedings*, *13*. https://aisel.aisnet.org/sighci2021/13

Kwak, D.-H., Holtkamp, P., & Kim, S. S. (2019). Measuring and Controlling Social Desirability Bias: Applications in Information Systems Research. *Journal of the Association for Information Systems*, 317–345. https://doi.org/10.17705/1jais.00537

Lambert, C. E., Arbuckle, S. A., & Holden, R. R. (2016). The Marlowe–Crowne Social Desirability Scale outperforms the BIDR Impression Management Scale for identifying fakers. *Journal of Research in Personality*, *61*, 80–86. https://doi.org/10.1016/j.jrp.2016.02.004

Larson, R. B. (2019). Controlling social desirability bias. *International Journal of Market Research*, *61*(5), 534–547. https://doi.org/10.1177/1470785318805305

Levy, A. G., Scherer, A. M., Zikmund-Fisher, B. J., Larkin, K., Barnes, G. D., & Fagerlin, A. (2018). Prevalence of and Factors Associated With Patient Nondisclosure of Medically Relevant Information to Clinicians. *JAMA Network Open*, *1*(7), e185293. https://doi.org/10.1001/jamanetworkopen.2018.5293

Levy, A. G., Thorpe, A., Scherer, L. D., Scherer, A. M., Drews, F. A., Butler, J. M., Burpo, N., Shoemaker, H., Stevens, V., & Fagerlin, A. (2022). Misrepresentation and Nonadherence Regarding COVID-19 Public Health Measures. *JAMA Network Open*, *5*(10), e2235837. https://doi.org/10.1001/jamanetworkopen.2022.35837

Ligthelm, R. J., Borzì, V., Gumprecht, J., Kawamori, R., Wenying, Y., & Valensi, P. (2007). Importance of Observational Studies in Clinical Practice. *Clinical Therapeutics*, *29*(6), 1284–1292. https://doi.org/10.1016/j.clinthera.2007.07.004

Loo, R., & Loewen, P. (2004). Confirmatory Factor Analyses of Scores From Full and Short Versions of the Marlowe–Crowne Social Desirability Scale. *Journal of Applied Social Psychology*, *34*(11), 2343–2352. https://doi.org/10.1111/j.1559-1816.2004.tb01980.x

Marquis, K. H., Marquis, M. S., & Polich, J. M. (1986). Response Bias and Reliability in Sensitive Topic Surveys. *Journal of the American Statistical Association*, *81*. https://doi.org/10.2307/2289227

Masters, T. M., Keith, M., Hess, R., & Jenkins, J. L. (2022). Do privacy assurances work? A study of truthfulness in healthcare history data collection. *PLOS ONE*, *17*(11), e0276442. https://doi.org/10.1371/journal.pone.0276442

Mathur, M. B., & Reichling, D. B. (2019). Open-source software for mouse-tracking in Qualtrics to measure category competition. *Behavior Research Methods*, *51*(5), 1987–1997. https://doi.org/10.3758/s13428-019-01258-6

Mazza, C., Monaro, M., Burla, F., Colasanti, M., Orrù, G., Ferracuti, S., & Roma, P. (2020). Use of mouse-tracking software to detect faking-good behavior on personality questionnaires: An explorative study. *Scientific Reports*, *10*(1), 4835. https://doi.org/10.1038/s41598-020-61636-5

McGrath, J. E., Martin, J. M., & Kulka, R. A. (1982). *Judgment Calls in Research* (Vol. 2). SAGE Publications, Incorporated.

Monaro, M., Gamberini, L., & Sartori, G. (2017). The detection of faked identity using unexpected questions and mouse dynamics. *PLOS ONE*, *12*(5), e0177851. https://doi.org/10.1371/journal.pone.0177851

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, *15*(3), 263–280. https://doi.org/10.1002/ejsp.2420150303

Nunnally, J. C. (1978). An Overview of Psychological Measurement. In B. B. Wolman (Ed.), *Clinical Diagnosis of Mental Disorders: A Handbook* (pp. 97–146). New York: Plenum Press. https://doi.org/10.1007/978-1-4684-2490-4_4

Paulhus, D. L. (1988). Balanced Inventory of Desirable Responding. *Acceptance and Commitment Therapy. Measures Package*, *41*. https://doi.org/10.1037/t08059-000

Paulhus, D. L. (1991). Measurement and Control of Response Bias. In *Measures of Personality and Social Psychological Attitudes* (pp. 17–59). Elsevier. https://doi.org/10.1016/B978-0-12-590241-0.50006-X

Pitesa, M., Thau, S., & Pillutla, M. M. (2013). Cognitive control and socially desirable behavior: The role of interpersonal impact. *Organizational Behavior and Human Decision Processes*, *122*(2), 232–243. https://doi.org/10.1016/j.obhdp.2013.08.003

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of Method Bias in Social Science Research and Recommendations on How to Control It. *Annual Review of Psychology*, *63*(1), 539–569. https://doi.org/10.1146/annurev-psych-120710-100452

Raschka, S., Liu, Y. H., Mirjalili, V., & Dzhulgakov, D. (2022). *Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python*. Packt Publishing Ltd.

Schaeffer, N. C. (2000). Asking questions about threatening topics: A selective overview. *The Science of Self-Report: Implications for Research and Practice*, 105–121.

Schuessler, K., Hittle, D., & Cardascia, J. (1978). Measuring Responding Desirably with Attitude-Opinion Items. *Social Psychology*, *41*(3), 224. https://doi.org/10.2307/3033559

Shaughnessy, J. J., Zechmeister, E. B., & Zechmeister, J. S. (2012). *Research Methods in Psychology* (9th ed.). McGraw-Hill.

Steenkamp, J.-B. E. M., De Jong, M. G., & Baumgartner, H. (2010). Socially Desirable Response Tendencies in Survey Research. *Journal of Marketing Research*, *47*(2), 199–214. https://doi.org/10.1509/jmkr.47.2.199

Straub, D., Boudreau, M.-C., & Gefen, D. (2004). Validation Guidelines for IS Positivist Research. *Communications of the Association for Information Systems*, *13*, 380–427. https://doi.org/10.17705/1CAIS.01324

Valacich, J. S., Jenkins, J. L., & Cisic, D. (2022). Digital Behavioral Biometrics and Privacy: Methods for Improving Business Processes without Compromising Customer Privacy. *2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO)*, 1019–1024. https://doi.org/10.23919/MIPRO55190.2022.9803621

Weinmann, M., Valacich, J., Schneider, C., Jenkins, J., & Hibbeln, M. (2022). The Path of the Righteous: Using Trace Data to Understand Fraud Decisions in Real Time. *MIS Quarterly*, *46*(4), 2317–2336. https://doi.org/10.25300/MISQ/2022/17038

Wright, R. (1994). *The moral animal: Why we are, the way we are: The new science of evolutionary psychology*. New York: Vintage Books.