Are Scrutiny and Mistrust Related? An Eye-Tracking Study

Danushka Bandara[1] (✉) and Sambhab Sau[2]

[1] Fairfield University, Fairfield, CT 06824, USA
dbandara@fairfield.edu
[2] Georgia Institute of Technology, Atlanta, GA 30332, USA

Abstract. Eyes are an important organ in both information processing and communicating. Eye gaze contains rich information about our attention and internal state. It has previously been studied as a measure of trust in various contexts such as air traffic control, driving, and online shopping. The study uses fixation and saccadic measurements to obtain a viable measure of human trust. The experiment involved nine participants and their trust level was indicated by whether they accepted or rejected a decision made by an Artificial Intelligence (AI). A Tobii pro nano eye-tracker and psychopy software were used to track the participants' eye gaze and responses. Results indicate that saccade count shows a statistically significant variation between trust and mistrust conditions with $p < 0.05$, while the fixation count showed a variation at $p < 0.2$. Through this study, we show that the count of saccades is a viable measure of a human's mistrust in an AI system.

Keywords: Eye tracking · Trust · Human-AI interaction

1 Introduction

Eye gaze tracking is a data-rich behavioral measure for Human Computer Interaction (HCI) researchers. It is used as a usability metric and a control mechanism [4]. In this study, we are interested in delineating the connection between eye gaze and perceived level of trust in a human-AI interaction scenario. Several research areas have used eye gaze as a measure of trust. It has been used in air traffic controller scenarios where it was used to measure stress-related task overload [1] because stress is connected to low levels of trust; eye tracking can be an indicator of trust levels. Researchers have shown in driving scenarios that there is a connection between the frequency and duration of eye glances and trust in the system [2, 3]. In a study examining the effect of human brands on consumers' trust levels in online shopping, it was ascertained that longer fixation durations on brand images, calculated from eye tracking data, indicated greater trust in specific products [12]. Moreover, in an unmanned aerial vehicle control study, participants were shown to trust high-reliability automation more than low-reliability automation with several key eye tracking metrics (total fixation duration, backtrack rate, scanpath length per second, etc.) being closely correlated with trust levels [14, 15].

C. Stephanidis et al. (Eds.): HCII 2023, CCIS 1832, pp. 539–545, 2023.
https://doi.org/10.1007/978-3-031-35989-7_68

Analysis of eye movements is typically conducted in terms of fixations (pauses over informative regions of interest) and saccades (rapid movements between fixations). Given the strong link between fixations and overt visual attention, fixations have been widely studied in experimental psychology to evaluate reading comprehension, facial processing, and online learning [18]. Typically, subjects exhibiting mistrust will pay greater attention to evaluating the system in question. Saccadic measurement has become prominent in many research fields as deficient saccadic function is helping in examining neurological activity and diagnosing various disorders [19]. Eye tracking metrics, such as fixations and saccades, extracted from raw eye movement data have been proven to work well as input to various modeling methods [16]. The parameters for identifying fixations and saccades are commonly either velocity-based or dispersion-based. For example, identifying a fixation might mean specifying the allowed dispersion from a single point in a subject's field of vision for a minimum time threshold. On the other hand, a saccade may be specified by defining a minimum velocity threshold for the pupil to be moving from one fixation to the next. A fixation and saccade can never occur at the same time. The exact parameters for identification are a matter of debate among researchers and are generally chosen based on what serves the model in question best [17]. This study explores the connection between eye tracking features and the trust level of a human. The trust and mistrust are measured by the humans' acceptance or rejection of decisions made by an Artificial Intelligence (AI) agent.

Research Hypothesis: Eye tracking features (fixations and saccades) can be used to distinguish trust and mistrust conditions.

2 Method

2.1 Experiment Design

We conducted a human subject study where nine subjects participated in a 'Human-AI Interaction' experiment. Participants completed a pre-experiment questionnaire to gauge their base level of trust in computers. We adopted the disposition to trust inventory [7] and adjusted the questions to reflect human-computer trust. When participants arrived for the experiment, they were seated in front of the computer and presented various images consecutively with 10 s of rest between each trial. Their job was to judge whether the images were authentic or 'doctored' (manipulated in any way). We used the CASIA [7] dataset as the source for the images. In addition to the images, participants were shown the AI decision of whether the image was authentic or not authentic. We used a 'wizard of oz' methodology for this study (i.e., we simulated the AI response) to show the participants a variety of AI responses. We also administered a trust measurement survey during the experiment to gauge how trust varies subjectively [9]. During the experiment, a Tobii pro nano eye-tracker was used to track their eye gaze continuously. The Tobii eye-tracker was calibrated to track the eye gaze location on the computer screen at 20 Hz. The user responses were tracked by the psychopy [10] software, which was also utilized to present the experiment materials; it tracked the time of presentation of each question as well. As indicated above, we combined eye tracking with the user responses to evaluate eye tracking data as trust indicators (Fig. 1).

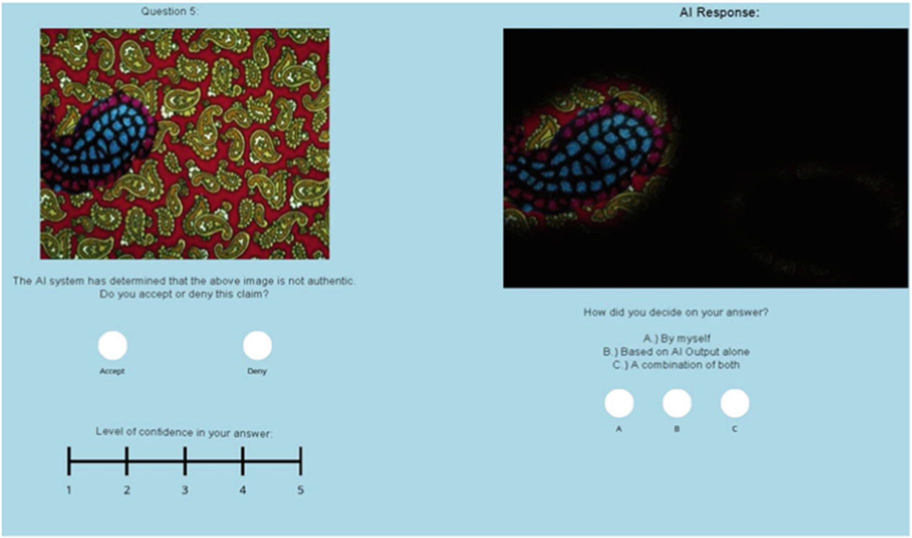

Fig. 1. An example screen that was shown to participants.

2.2 Data Collection

The participants completed the consent form and pre-experiment questionnaire before attending the experiment. When they arrived for the experiment, they were asked to sit in front of the computer comfortably. Then, the Tobii pro nano eyetracker was calibrated for their eye gaze. The eyetracking data was collected in a csv file along with timestamps. The psychopy experiment was run and collected the user response for each question along with the time stamps. The starting and ending timestamps for each question was used to extract the eye tracking data for each question. Then the eyetracking data was further processed to obtain features such as fixations and saccades.

2.3 Data Processing and Modeling

A fixation is defined here as a state in which the eye is focused on a point of interest with some maximum limit in variation and minimum time limit. For this study, the maximum allowed variation from the point of interest was 25 pixels and the minimum duration was 200 ms. A saccade is defined as a state in which the eye moves rapidly from one point of interest to another for some minimum time period, minimum velocity threshold, and minimum acceleration threshold. After doing a grid search over the parameters, these three parameters were chosen to be 5 ms, 60 pixels/second, and 150 pixels/seconds^2, respectively. Fixations and saccades were calculated for each question asked to participants. Each fixation was documented using the start time, end time, total duration (calculated as start time subtracted from end time), and x and y coordinates of the fixation. Each saccade was documented using the start time, end time, total duration, and two pairs of x and y coordinates representing the start and end of the saccade [11].

Incomplete data from one of the participants was discarded, leaving the eyetracking data and levels of trust of 8 participants for analysis. For the raw eyetracking data, (0,0) represents the top left of the screen, (0.5, 0.5) represents the center of the screen, and (1,1) represents the bottom right of the screen. Using the resolution information of the computer screen used (1920 x 1080), the raw eyetracking data was converted from these ratios to pixels where (0,0) still represents the top left of the screen, while (1920, 1080) represents the bottom right of the screen. Further, we only used the left eye's coordinates for feature extraction of fixations and saccades in this paper as using both eyes' coordinates was assumed to be redundant. Moreover, any rows of the eyetracker data that had 0s, assumed to be output errors of the Tobii eyetracker, were removed so as not to bias the detection of fixations and saccades.

From the fixation dataset, two additional features were collected which were the fixation counts and the average fixation durations per question. In the same way, the saccade dataset was used to collect two additional features representing saccade counts and average fixation durations per question. These two features from both fixations and saccades were then divided again into two datasets representing the questions where participants accepted the AI response and questions where participants denied the AI response. Overall, a total of 685 questions were asked to participants with 416 questions being accepted by participants and 269 of them being denied.

3 Results

Here we show the results of our analysis for the trust vs mistrust conditions. The mistrust group is defined as the set of questions where participants denied the AI's recommendation, and the trust group is defined as the set of questions where participants accepted the AI's recommendation. We have left out the fixation duration and saccade duration results because they did not show any variation between trust and mistrust conditions.

The top left boxplot in Fig. 2 represents the fixation counts for all subjects divided between mistrust and trust. The mean for the mistrust group was 29 fixations, while the mean for the trust group was 27. The interquartile range (IQR) for the mistrust group was 14, while the IQR for the trust group was 15. Upon conducting a one-way ANOVA test to compare the means of the trust and mistrust group for all subjects, the p-value was found to be approximately 0.155512.

The top left boxplot of Fig. 3 represents the saccade counts for all subjects divided between mistrust and trust. The mean for the mistrust group was 43 saccades, while the mean for the trust group was 38. The interquartile range (IQR) for the mistrust group was 26, while the IQR for the trust group was 24. Upon conducting a one-way ANOVA test to compare the means of the trust and mistrust group for all subjects, the p-value was found to be approximately 0.001995. Assuming an alpha value of 0.05, a statistically significant difference was observed between the two groups for saccadic counts ($0.001995 < 0.05$).

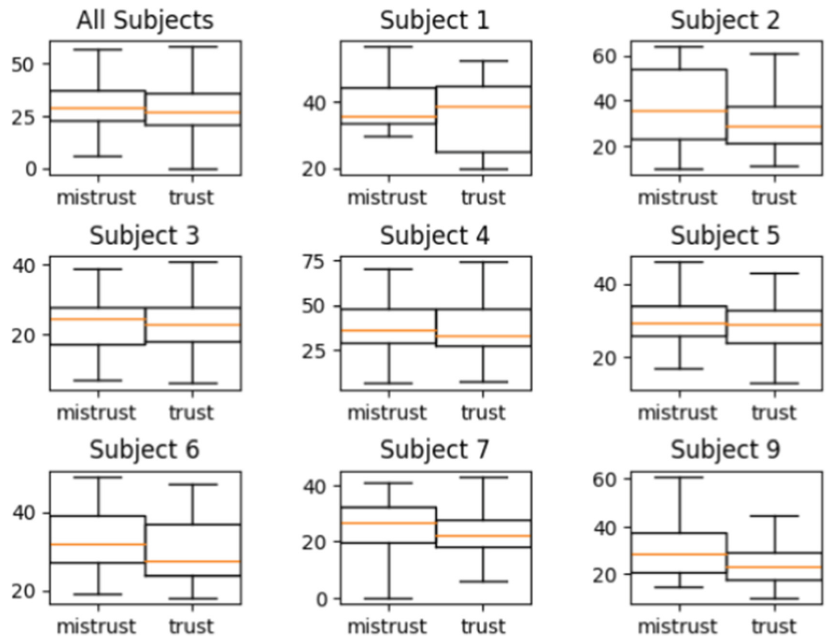

Fig. 2. (Top left) Per question average fixation counts for all subjects in the dataset. Fixation counts are higher for the mistrust condition across all subjects. (the rest of the plots) Per question average fixation counts for each individual subject.

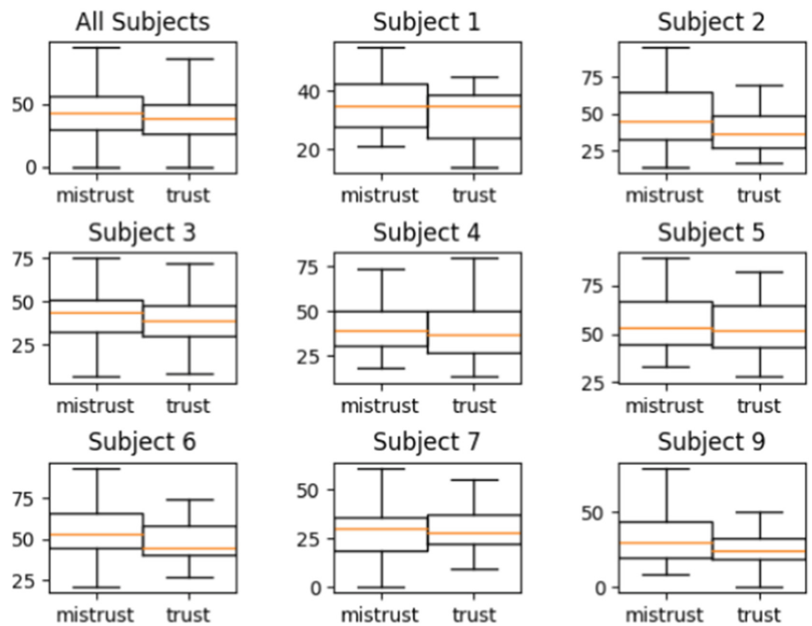

Fig. 3. (Top left) Per question average saccade counts for all subjects in the dataset. Saccade counts are higher for the mistrust condition across all subjects. (the rest of the plots) Per question average saccade counts for each individual subject.

4 Discussion

The results show that the count of saccades and fixations was higher in the mistrust condition which confirms our hypothesis that a trust and mistrust can be distinguished by saccade and fixation features. We did see a variation between subjects of how large this effect was. Indicating that there is inter subject variation on the effect of trust level on eye gaze. Saccadic counts provided a statistically significant discrimination between trust and mistrust conditions than fixations and, therefore, could be a novel complementary feature to accurately predict trust levels. In conclusion, our study shows the utility of eye tracking for evaluating trust and mistrust conditions in human computer interaction.

References

1. Prinet, J.C., Sarter, N.B.: The effects of high stress on attention: a first step toward triggering attentional narrowing in controlled environments. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 59(1), pp. 1530–1534 (2015). https://doi.org/10.1177/1541931215591331
2. Geitner, C., et al.: A link between trust in technology and glance allocation in on-road driving. In: Driving Assessment Conference, vol. 9 (2017). https://doi.org/10.17077/drivingassessment.1645
3. Hergeth, S., Lorenz, L., Vilimek, R., Krems, J.F.: Keep your scanners peeled: gaze behavior as a measure of automation trust during highly automated driving. Hum Factors. **58**(3), 509–519 (2016). https://doi.org/10.1177/0018720815625744. Epub 2016 Feb 3 PMID: 26843570
4. Jacob, R.J., Karn, K.S.: Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In: The mind's eye, North-Holland, pp. 573–605 (2003)
5. Freeman, J.B.: Doing psychological science by hand. Curr. Dir. Psychol. Sci. **27**(5), 315–323 (2018). https://doi.org/10.1177/0963721417746793
6. Maldonado, M., Dunbar, E., Chemla, E.: Mouse tracking as a window into decision making. Behav. Res. Methods **51**(3), 1085–1101 (2019). https://doi.org/10.3758/s13428-018-01194-x
7. Disposition to trust inventory [McKnight, D.H., Choudhury, V., Kacmar, C.: Developing and validating trust measures for e-commerce: an integrative typology. Inf. Syst. Res. **13**(3), 334–359 (2002).] modified to cater to trust in AI instead of trust in people
8. Sutherland, S.C., Harteveld, C., Young, M.E.: the role of environmental predictability and costs in relying on automation. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI' 2015). Association for Computing Machinery, New York (2015). https://doi.org/10.1145/2702123.2702609
9. Madsen, M., Gregor, S.D.: Measuring Human-Computer Trust (2000)
10. Peirce, J.W.: PsychoPy--Psychophysics software in Python. J. Neurosci. Methods **162**(1–2), 8–13, 15 May 2007. https://doi.org/10.1016/j.jneumeth.2006.11.017. Epub 2007 Jan 23. PMID: 17254636; PMCID: PMC2018741
11. Ghose, U., Arvind, A.S.: Titoghose/Pytrack. GitHub. https://github.com/titoghose/PyTrack. Accessed 15 Mar 2023
12. Wook Chae, S., Chang Lee, K.: Exploring the effect of the human brand on consumers' decision quality in online shopping. Online Inf. Rev. **37**(1), 83–100 (2013). https://doi.org/10.1108/14684521311311649
13. Jenkins, Q., Jiang, X.: Measuring trust and application of eye tracking in human robotic interaction. In: Industrial Engineering Research Conference (2010)
14. Lu, Y.: Detecting and overcoming trust miscalibration in real time using an eye-tracking based technique (dissertation) (2020)

15. Lu, Y., Sarter, N.: Eye tracking: a process-oriented method for inferring trust in automation as a function of priming and system reliability. IEEE Trans. Hum.-Mach. Syst. **49**(6), 560–568 (2019). https://doi.org/10.1109/thms.2019.2930980
16. Lu, Y., Sarter, N.: Modeling and inferring human trust in automation based on real- time eye tracking data. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 64(1), pp. 344–348 (2020). https://doi.org/10.1177/1071181320641078
17. Salvucci, D.D., Goldberg, J.H.: Identifying fixations and saccades in eye-tracking protocols. In: Proceedings of the Symposium on Eye Tracking Research & Applications - ETRA 2000 (2000). https://doi.org/10.1145/355017.355028
18. Steil, J., Huang, M.X., Bulling, A.: Fixation detection for head-mounted eye tracking based on visual similarity of gaze targets. In: Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications (2018). https://doi.org/10.1145/3204493.3204538
19. Stuart, S., et al.: Eye-tracker algorithms to detect saccades during static and dynamic tasks: a structured review. Physiol. Meas. **40**(2) (2019). https://doi.org/10.1088/1361-6579/ab02ab