Your Eyes Never Lie:
A Robot Magician Can Tell if You Are Lying

Dario Pasquali[1,2,3]
dario.pasquali@iit.it

Alexander Mois Aroyo[4]
alexander.aroyo@iit.it

Jonas Gonzalez-Billandon[1,2]
jonas.gonzalez@iit.it

Francesco Rea[2]
francesco.rea@iit.it

Giulio Sandini[2]
giulio.sandini@iit.it

Alessandra Sciutti[4]
alessandra.sciutti@iit.it

ABSTRACT

Detecting lies in a real-world scenario is an important skill for a humanoid robot that aims to act as a teacher, a therapist, or a caregiver. In these contexts, it is essential to detect lies while preserving the pleasantness of the social interaction and the informality of the relation. This study investigates whether pupil dilation related to an increase in cognitive load can be used to swiftly identify a lie in an entertaining scenario. The iCub humanoid robot plays the role of a magician in a card game, telling which card the human partner is lying about. The results show a greater pupil dilation in presence of a false statement even if in front of a robot and without the need of a strictly controlled scenario. We developed a heuristic method (accuracy of 71.4% against 16.6% chance level) and a random forest classifier (precision and recall of 83.3%) to detect the false statement. Additionally, the current work suggests a potential method to assess the lying strategy of the partner.

KEYWORDS

Lie detection; humanoid robot; machine learning; human-robot interaction; pupillometry

ACM Reference format:

Dario Pasquali, Alexander Mois Aroyo, Jonas Gonzalez-Billandon, Francesco Rea, Giulio Sandini and Alessandra Sciutti. 2020. Your Eyes Never Lie: A Robot Magician Can Tell if You Are Lying. *In Proceedings' of HRI (HRI '20) Cambridge conference.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3371382.3378253

1 Introduction

Detecting lies during human-robot interaction (HRI) could be useful for many robotic applications in daily life like teaching [1], [2], rehabilitation [3], law enforcing [4] or even entertainment [5]. Lying can be related to a higher cognitive effort due to the need to create and maintain a plausible and coherent story [6]–[9]. Task-evoked cognitive load has proven to reflect into pupillary responses like mean pupil dilation, peak dilation and latency to peak [4], [10]–[13] measurable with minimally invasive devices [14]–[17]. Pupillometry has been proposed as a less invasive lie detection method than traditional approaches [6], [18], [19]. However, even with this technique, cognitive load assessment for lie detection is usually performed in highly controlled settings comparable to interrogatories [16], [20]. Recently researchers have studied cognitive load assessment during HRI with social robots [15], [21], [22] and its application for lie detection. For instance, Gonzalez-Billandon *et al.* [20] and Aroyo *et al.* [23] compared lie detection through pupillometry in HHI and HRI finding no significant differences. Finally, recent developments on RGB cameras [24] suggest the possibility to detect pupillometric features with traditional devices that are portable on robotic platforms. The current work explores the possibility to build a robot able to detect lies, based on evidences of pupillometric features related to an increase in cognitive load, during a quick and minimally invasive game-like HRI. As a proof of concept, we consider an entertainment application with a magician robot, played by iCub [25]. During a card game, we used pupillometric features, collected with an eyetracker, in order to enable iCub to guess when the participant is lying about one out of six playing cards. Results show the possibility to determine user-specific eye behavioral patterns and to relate them to cognitive load generated during lie fabrication with a quick process. We implemented a heuristic method with 71.4% of accuracy (against a chance level of 16.6%) and a random forest classifier with 83.3% of precision and recall. This shows the feasibility of porting the method on a practical HRI setting.

[1]DIBRIS, Università di Genova, Opera Pia 13, 16145, Genova, Italy
[2]RBCS, Istituto Italiano di Tecnologia, Enrico Melen 83, Bldg B, 16152, Genova, Italy
[3]ICT, Istituto Italiano di Tecnologia, Enrico Melen 83, Bldg B, 16152, Genova, Italy
[4]CONTACT, Istituto Italiano di Tecnologia, Enrico Melen 83, Bldg B, 16152, Genova, Italy

ACM ISBN 978-1-4503-7057-8/20/03.
DOI: https://doi.org/10.1145/3371382.3378253

2 Methodology

We asked 28 participant – all Italians, 9 males and 17 females, average age of 24 years old (SD=5) – to take part in the experiment. The research was approved by the Liguria Regional Ethical Committee; all participants provided their written informed consent and received a compensation of 15 euro. We removed 7 participants because Tobii eyetracker malfunctioned leading to substantial data loss (number of collected datapoints for at least one card lower than 35% of the expected number). The experimental room was arranged in order to have a chair for the participant and a table between the chair and the robot. The room was lit with artificial light and the windows blinders were closed to ensure the same light conditions for all the participants. Participants wore a Tobii Pro Glasses 2 eyetracker that recorded pupillometric features with a frequency of 100 Hz. For the experiment, we prepared 6 cardboard cards of 15x21 cm with different colored pictures over a white background. The iCub humanoid robot [25] controlled following the *Wizard of Oz* model [26], played the role of magician (Figure 1, left panel). Before the game, the experimenter asked the participants to draw out and memorize a random card (*target* card) out of the six cards, then he asked them to wear the Tobii glasses. Then the experimenter asked the participants to sit on the chair in front of iCub, put the six-cards deck on the table and left the room. ICub asked the participants to shuffle the cards, look at them one by one, shuffle them again and put them covering six marks on the table. ICub said that, in order to perform a magic trick, it was going to point each card one by one and it instructed the participants to take and describe the pointed card. It said: "*The trick is this: if the card you took was the one you drew out before the experiment, you should describe it in a deceitful and creative way. Otherwise, describe just what you see*". Finally, iCub, guided by the experimenter, guessed the participants' card and asked for a confirmation. The authors manually segmented three intervals for each card: POINT intervals (from the pointing end to the moment when the participant takes the card), REACT intervals (from the moment when the participant takes the card to the beginning of the description) and DESCR intervals (during the card description). Pupillometric features, collected with the eyetracker, were aggregated for each interval computing maximum, minimum, mean and standard deviation (in millimeters) for right and left pupil dilation. Pupil dilation features were normalized subtracting the average pupil dilation (for each eye) during the five seconds before the first pointing [27].

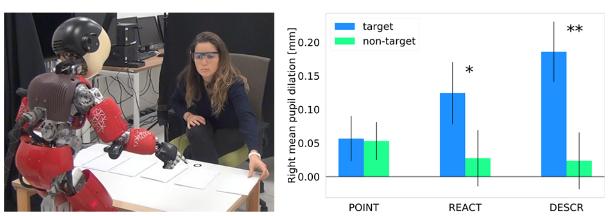

Figure 1: (Left panel) iCub pointing to a card during the game. (Right panel) Average right mean pupil dilation for *target* card and averaged *non-target* cards during the three intervals. Error bars represent standard error of the mean.

3 Results and Discussion

The main objective addressed in this work is to understand the feasibility of identifying a lie in the description of a card, based on participants' pupillometric features during a short and game-like interaction with a robot. We computed the average of each feature for the *non-target* card and compared it with the same feature for the *target* card. Shapiro-Wilk [28] and D'Agostino k-squared [29] tests showed that data were non normally distributed. Wilcoxon signed rank test [30] in the three intervals showed a highly significant difference in pupil dilation for the *target* card versus the others for both eyes during DESCR interval and a significant difference during REACT interval (Figure 1, right panel). We assume this factor represents a higher cognitive load due to the necessity to create a deceitful and creative description for the *target* card. Also, participants showed a high variability in the pupil dilation effect between the two intervals. This could represent different levels of lying expertise and, hence, induced cognitive load, like in [14]. We suggest the presence of two different lying strategies: *Premeditators* (71.4%, n=15) faced an increase in cognitive load before describing the card as if they were preparing the lie in advance; *Improvisers* (28.6%, n=6) showed a greater pupil dilation during the description as if the lie fabrication happened during the speech. In order to make iCub able to autonomously detect the *target* card we developed two models. We defined a simple heuristic function that, for each participant, classified as *target*, the card related to the higher right mean pupil dilation reaching an accuracy of 71.4% (against a chance level of 16.6%). However, the heuristic is applicable only if the presence and uniqueness of a deceitful description on a finite set of items is a priori known. Hence, we developed a random forest classifier able to classify a generic set of descriptions as true or false. We selected 13 features and applied a min max [31] per-subject normalization to all the features within the 6 cards. We performed a 4-fold grid search cross validation applying synthetic minority oversampling technique [32] inside the folds to balance the training set. The best classifier, optimized on the AUCROC score, has precision and recall of 83.3% and AUCROC score of 89.7% with SD=0.7. The random forest solves the lie uniqueness limitation of the heuristic. However, the per-subject normalization makes the model able to classify an item only among the observations of the same individual. The current implementation relies on manual post-hoc segmentation. However, preliminary analysis based on the current dataset showed that it is possible to discriminate the *target* card also with a less precise annotation: considering the intervals between two pointing gestures we obtained 62.9% of accuracy with the heuristic and precision and recall of 71.4% with the random forest. The extraction of these intervals can be performed autonomously by the robot based on the instants in which the pointing actions are initiated. We could also compute the features during the game by processing eyetracker data on-board. Finally, we could improve iCub's role making it able to actively propose the card in order to make the task like a real-world application. Currently, we are working on a new real-time version of the experiment useful to enlarge our dataset, improve performance and generalization of the proposed models and make a step toward a real-world application.

4 Conclusion

This work demonstrates the possibility to detect lies through the assessment of cognitive load during a quick and minimally invasive interaction with the iCub robotic platform. This ability is particularly relevant in the perspective of a future where robots will be part of our everyday life, especially in fields related to tutoring, caregiving, security and entertainment. For instance, a companion robot could detect if a child did the homework or a patient took the pills and act consistently. Our hope is that this ability could improve the social capabilities of robots making them able to better understand the true intentions of their partners.

REFERENCES

[1] A. Westbrook and T. S. Braver, "Cognitive effort: A neuroeconomic approach," Cognitive, Affective and Behavioral Neuroscience, vol. 15, no. 2. Springer New York LLC, pp. 395–415, 22-Jun-2015.

[2] J. Beatty, "Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources," 1982.

[3] A. Koenig et al., "Real-time closed-loop control of cognitive load in neurological patients during robot-assisted gait training.," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 19, no. 4, pp. 453–64, Aug. 2011.

[4] J. A. Stern, L. C. Walrath, and R. Goldstein, "The endogenous eyeblink.," Psychophysiology, vol. 21, no. 1, pp. 22–33, Jan. 1984.

[5] J. Fallon, "Box of Lies." [Online]. Available: https://www.youtube.com/watch?v=Md4QnipNYqM.

[6] C. R. Honts, D. C. Raskin, and J. C. Kircher, "Mental and physical countermeasures reduce the accuracy of polygraph tests.," J. Appl. Psychol., vol. 79, no. 2, pp. 252–9, Apr. 1994.

[7] B. M. DePaulo, J. J. Lindsay, B. E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper, "Cues to deception.," Psychol. Bull., vol. 129, no. 1, pp. 74–118, 2003.

[8] D. P. Dionisio, E. Granholm, W. A. Hillix, and W. F. Perrine, "Differentiation of deception using pupillary responses as an index of cognitive processing.," Psychophysiology, vol. 38, no. 2, pp. 205–11, Mar. 2001.

[9] S. M. Kassin, "On the Psychology of Confessions: Does Innocence Put Innocents at Risk?," Am. Psychol., vol. 60, no. 3, pp. 215–228, Apr. 2005.

[10] J. G. May, R. S. Kennedy, M. C. Williams, W. P. Dunlap, and J. R. Brannan, "Eye movement indices of mental workload," Acta Psychol. (Amst)., vol. 75, no. 1, pp. 75–89, 1990.

[11] M. Nakayama and Y. Shimizu, "Frequency analysis of task evoked pupillary response and eye-movement," in Proceedings of the Eye tracking research & applications symposium on Eye tracking research & applications - ETRA'2004, 2004, pp. 71–76.

[12] K. F. Van Orden, W. Limbert, S. Makeig, and T. P. Jung, "Eye activity correlates of workload during a visuospatial memory task.," Hum. Factors, vol. 43, no. 1, pp. 111–21, 2001.

[13] B. C. Goldwater, "Psychological significance of pupillary movements," Psychol. Bull., vol. 77, no. 5, pp. 340–355, May 1972.

[14] A. Szulewski, N. Roth, and D. Howes, "The Use of Task-Evoked Pupillary Response as an Objective Measure of Cognitive Load in Novices and Trained Physicians: A New Tool for the Assessment of Expertise," Acad. Med., vol. 90, no. 7, pp. 981–987, Jul. 2015.

[15] M. Ahmad, B. Jasmin, L. Katrin, and F. Eyssel, "Trust and Cognitive Load During Human-Robot Interaction," 2019.

[16] J. Klingner, "Measuring Cognitive Load During Visual Task by Combining Pupillometry and Eye Tracking," Ph.D. Diss., no. May, 2010.

[17] G. Hossain and M. Yeasin, "Understanding effects of cognitive load from pupillary responses using hilbert analytic phase," IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work., pp. 381–386, 2014.

[18] A. Gaggioli, "Beyond the Truth Machine: Emerging Technologies for Lie Detection," Cyberpsychology, Behav. Soc. Netw., vol. 21, no. 2, pp. 144–144, Feb. 2018.

[19] M. Gamer, "Detecting of deception and concealed information using neuroimaging techniques," in Memory Detection, B. Verschuere, G. Ben-Shakhar, and E. Meijer, Eds. Cambridge: Cambridge University Press, 2011, pp. 90–113.

[20] J. Gonzalez-Billandon et al., "Can a Robot Catch You Lying? A Machine Learning System to Detect Lies During Interactions," Front. Robot. AI, vol. 6, Jul. 2019.

[21] K. Kobayashi and S. Yamada, "Human-robot interaction design for low cognitive load in cooperative work," in RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759), pp. 569–574.

[22] S. M. Al Mahi, M. Atkins, and C. Crick, "Learning to assess the cognitive capacity of human partners," ACM/IEEE Int. Conf. Human-Robot Interact., pp. 63–64, 2017.

[23] A. M. Aroyo et al., "Can a Humanoid Robot Spot a Liar?," IEEE-RAS 18th Int. Conf. Humanoid Robot., 2018.

[24] C. Wangwiwattana, X. Ding, and E. C. Larson, "PupilNet, Measuring Task Evoked Pupillary Response using Commodity RGB Tablet Cameras," Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol., vol. 1, no. 4, pp. 1–26, Jan. 2018.

[25] G. Metta, G. Sandini, D. Vernon, L. Natale, and F. Nori, "The iCub humanoid robot," in Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems - PerMIS '08, 2008, p. 50.

[26] L. Riek, "Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines," J. Human-Robot Interact., vol. 1, no. 1, pp. 119–136, 2012.

[27] S. Mathôt, J. Fabius, E. Van Heusden, and S. Van der Stigchel, "Safe and sensible preprocessing and baseline correction of pupil-size data," Behav. Res. Methods, vol. 50, no. 1, pp. 94–106, 2018.

[28] S. Shapiro and M. Wilk, "An analysis of variance test for normality (complete samples)," Z. Bibliothekswes. Bibliogr., vol. 52, no. 6, pp. 311–319, 2005.

[29] R. B. D'Agostino, A. Belanger, and R. B. D'Agostino, "A Suggestion for Using Powerful and Informative Tests of Normality," Am. Stat., vol. 44, no. 4, p. 316, Nov. 1990.

[30] F. Wilcoxon, "Individual Comparisons by Ranking Methods," Biometrics Bull., vol. 1, no. 6, p. 80, Dec. 1945.

[31] S. G. K. Patro and K. K. sahu, "Normalization: A Preprocessing Stage," Iarjset, no. April, pp. 20–22, 2015.

[32] Nitesh V. Chawla, K. W. Bowyer, and L. O. Hall, "SMOTE: Synthetic Minority Over-sampling Technique Nitesh," J. Artif. Intell. Res., 2006.