Do We Blindly Trust Self-Driving Cars?

Kamilla Egedal Andersen, Simon Köslich, Bjarke Maigaard Kjær Pedersen & Bente Charlotte Weigelin
Mærsk Mc-Kinney Møller Institute, University of Southern Denmark

Lars Christian Jensen
Department of Design & Communication, University of Southern Denmark

ABSTRACT

Trust is an essential factor in ensuring robust human-robot interaction. However, recent work suggests that people can be too trusting of the technology with which they interact during emergencies, causing potential harm to themselves. To test whether this "over-trust" also extends to normal day-to-day activities, such as driving a car, we carried out a series of experiments with an autonomous car simulator. Participants (N=73) engaged in a scenario with no, correct or false audible information regarding the state of traffic around the self-driving vehicle, and were told they could assume control at any point in the interaction. Results show that participants trust the autonomous system, even when they should not, leading to potential dangerous situations.

Keywords

Human Robot Interaction; HRI; Trust; Artificial Intelligence; AI; Autonomous System; AS; Self-Driving Car; SDC;

1. INTRODUCTION

Autonomous systems (AS), robots and AI's, such as Google, Siri and Cortana, find their ways into more and more areas of human everyday life. Also self-driving cars (SDC) will be commonplace in the foreseeable future. Currently, SDCs operate on the premise of the driver, who is still responsible for taking over control, should the car not handle a situation properly. For people to relinquish control of their car to an AS they need to trust that the system works. However, over-trusting technology to the point of blind trust can lead to potentially dangerous situations. In a recent study, Robinette et al. [1] show that people put far too much faith in a robot in an emergency evacuation scenario. The current paper expands on this work and asks whether this over-trusting effect also expands to interactions that are not set in an emergency setting. In particular, we are interested in how transparency influences the interaction between users and a SDC and to what degree the car can provide users with wrong information before the user assumes control by braking the car. The hypothesis is that participants will overlook lesser errors, and only recognize larger ones. Consequently, they will not assume control when lesser errors occur, and thus place themselves in potentially dangerous situations.

2. RELATED WORK

Waytz et al. [2] show that people are more likely to trust ASs with anthropomorphic conditions, such as a name, gender and a voice. Work by Helldin et al. [3] suggests that drivers, who receive information about an SDCs uncertainty on how to act in a given situation, tend to trust the car less than drivers who did not receive such information. In their study, informed drivers were faster to assume control of the car when a dangerous situation occurred. Koo et al. [4] state that providing the driver with information on "why" and "how" an autonomous vehicle acts, is important to maintain a safe driving performance. Whereas Richards, D., & Stedmon, A. [5] suggest that a certain amount of system transparency has to be maintained in order for the user to fully understand what an AS can and cannot do. Robinette et al. [1] found that people are very likely to blindly trust an AS even when it is making an obvious mistake. The research carried out in this paper, builds upon the findings of Waytz et al., Koo et al. and Richards et al. in regards to the applied interface of the simulated SDC. It differs from Helldin et al. by not expressing uncertainty, in an effort to simulate actual sensor malfunctions, and from Robinette et al. by moving the scenario away from an emergency situation and into an everyday one.

3. METHODS

To carry out the study we set up a between-subject experiment with four experimental conditions, in which participants (N=73) interacted with a car simulator from Oktal[6] Participants are mostly students (87.7%), and mostly men (79.5%), who have experience with programming and/or robotics (79.5%). Age ranges between 19 and 64.

3.1 Procedure

In all conditions participants were placed in the simulator in which an SDC would encounter three main events: An intersection where four out of six cars are controlled by humans, a crossing pedestrian and a second intersection with five other cars, all of which are SDCs. The first condition provides no audible information. In the other scenarios the interfaces continuously commented on what the SDC did and why, through computer synthesis speech. These conditions (2-4) differed only at the second intersection, where they announced a different number of registered SDCs - five being the correct number (see table 1).

	First cross	Pedestrian	Second cross
C1 (n=15)	None	None	None
C2 (n=15)	Inform	Inform	Correct, 5 cars
C3 (n=16)	Inform	Inform	False, 2 cars
C4 (n=27)	Inform	Inform	False, 0 cars

Table 1: Experimental conditions

HRI '17 Companion March 06-09, 2017, Vienna, Austria

ACM ISBN 978-1-4503-4885-0/17/03…$15.00
DOI: http://dx.doi.org/10.1145/3029798.3038428

3.2 Analysis

Each simulation took 1:57 minutes to run. Demographic questions were asked before the start of the simulation, and afterwards a set of predefined questions was asked in a structured interview. The collected data was then analyzed with chi-square tests.

4. RESULTS

A chi-square test between C1 and C2-4 shows no significant differences in the percentage of participants that assumed control, i.e. braked, in the first intersection (0.00%, 3.45%) and at the crossing pedestrian (13.33%, 12.07%). Another chi-square test between C1, C2, C3 and C4, however, shows significant differences in how many of the participants assumed control in the second intersection (see figure 1).

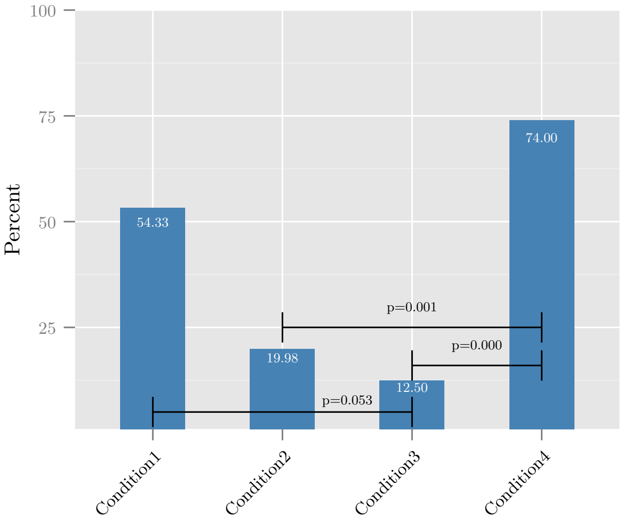

Figure 1: Percent of participants that assumed control

A third chi-square test between C3 and C4 shows no significant difference in how many of the participants detected that they were being wrongly informed in the second intersection. Yet another chi-square test shows a significant difference in how many of these still decided not to assume control (see table 2).

	C3	**C4**	**p**
Detected	43.75%	66.60%	0.148
Did not assume control	71.41%	11.12 %	0.001

Table 2: The percentage of participants who noticed that the information were wrong, and the percentage of these that after noticing it, still did not assume control

5. DISCUSSION

The results show that too little transparency in a SDCs actions, can cause situations where the driver has no understanding of what the car is currently doing (or why). This is especially likely to happen, when there are no apparent changes in behavior of the SDC, to indicate its current intentions. This can explain the similarity in the percentage of braking participants between C1 and C2-4 at the 1st intersection and at the crossing pedestrian. Here there is a clear physical feedback from the SDC about its intentions, due to the car slowing down well in advance and often before the participants themselves noticed the pedestrian. This stands in contrast to the percentage difference in participants who assumed control, between C1 and C2 at the second intersection. In C1 there are no clear indications of the SDCs doings or its capability to synchronize with the other SDCs, and thus pass them without slowing down. Should the participant suddenly brake in the middle of the second intersection in C1 – and thereby abruptly force the SDC to stray from its expected route - it would likely result in a dangerous situation. On the other hand, too much transparency and detailed information, can lead to smaller errors going undetected, being explained away and ultimately ignored, as seen in C3 – with possibly fatal consequences. ”I trust the programming more than the voice”, ”It’s a machine, it must know best”, are some of the quotes on why participants chose not to act, even though they noticed the mistake. Whereas in C4 there was a significant increase in participants who acted correctly by taking control of the SDC, after detecting the error in the given information.

Furthermore, the results clearly show that many participants were unable to detect errors when they occurred, no matter how apparent they were. Several participants commented on this being due to the car having done so well up until then. This had made them stop listening too closely to the information provided to them. ”I trust technology, but didn’t hear what the car said.”, ”Earlier, it knew exactly what was where”, so even though they did not hear what was said, they chose not to assume control anyway.

5.1 Limitations and Future Work

Since the simulation only lasted 1.57 minutes, it must be considered very unrealistic to expect that the same participants would notice an error, should one occur after a prolonged period of time. That the group of participants mainly consisted of young, male, engineering students, could prove a potential source of error. Also, had the participants faced any real danger, the results might have been different. Taking into consideration how important this matter will be in the future, it should be tested thoroughly. This could be carried out with more realistic simulations, or – if possible – controlled experiments in SDCs.

6. REFERENCES

[1] Robinette, P., Li, W., Allen, R., Howard, A. M., & Wagner, A. R. (2016, March). *Overtrust of robots in emergency evacuation scenarios.* In 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 101-108). IEEE.

[2] Waytz, A., Heafner, J., & Epley, N. (2014). *The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle.* Journal of Experimental Social Psychology, 52, 113-117.

[3] Helldin, T., Falkman, G., Riveiro, M., & Davidsson, S. (2013, October). *Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving.* In Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 210-217). ACM.

[4] Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., & Nass, C. (2015). *Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance.* International Journal on Interactive Design and Manufacturing (IJIDeM), 9(4), 269-275.

[5] Richards, D., & Stedmon, A. (2016). *To delegate or not to delegate: A review of control frameworks for autonomous cars.* Applied ergonomics, 53, 383-388.

[6] Oktal. *http://www.oktal.fr/en/automotive/range-of-simulators/software.* Accessed January 17th. 2017.