

# TRUSTING AUTONOMOUS SYSTEMS – DESIGNING FOR BYSTANDERS AND SECONDARY USERS

Elise Christin Lind HJEMLY and Ole Andreas ALSOS  
NTNU Norwegian University of Science and Technology

## ABSTRACT

Knowledge and awareness of the relationship and trust between man and technology is important for users to accept artificial intelligence (AI) and autonomous systems. While research has been conducted on the interaction between operators or passengers and robotic vehicles, there is not much research on secondary users, i.e., users who are not directly interacting with the system, such as pedestrians or bystanders in the robot's environments. This paper summarizes and synthesizes the existing literature relevant to secondary users' trust toward autonomous systems. The focus is (1) to find out how the secondary user experience in human-robot interaction is affected by trust, and (2) to see how trust is designed in autonomous systems. By designing for appropriate levels of trust, the designer can empower both primary and secondary users to accept the information provided by the robot and follow its suggestions.

*Keywords: Trust, autonomous system, human-robot interaction, design, secondary user experience*

## 1 AUTONOMOUS SYSTEMS AND THE USER EXPERIENCE OF BYSTANDERS

Artificial intelligence (AI) in the form of autonomous systems, such as robots and vehicles, are introduced to our everyday life in a wide variety of domains, including transportation, military, scientific applications, entertainment, and home use. With advances in the development of autonomous systems and autonomous transport solutions, the relationship and trust between man and machine is important for humans to accept the technology [1]. This paper explores trust-related challenges in human-robot interaction (HRI) by outlining how the secondary user experience in HRI is affected by trust.

### 1.1 The secondary user experience of autonomous systems

To achieve full and successful implementation of autonomous systems, it is essential to understand the notions of user acceptance and intended behaviour in relation to them. However, most studies target the buyers, main users, or operators of robots. These studies do not consider 3rd persons or bystanders who do not explicitly interact with the robots. However, these persons can nevertheless interfere with and be affected by the robot's actions. Alsos and Svanæs [2] refer to these as *secondary users* [2]. Examples in the context of autonomous systems are pedestrians walking in front of a self-driving vehicle, a kayaker encountering an autonomous passenger ferry in a canal, a person getting in the way of an autonomous warehouse robot co-worker, or someone meeting a self-driving cleaning robot in a hallway. Further Alsos and Svanæs define the *secondary user experience* of a system as “the part of the overall experience of the secondary user that can be attributed to (1) the primary user's interaction with the system, or (2) the secondary user's interaction with the system with the primary users as an intermediary [2]”. In this context, it is the user experience that passengers, pedestrians, or bystanders have when experiencing an autonomous system.

In fully autonomous systems, the robot is technically independent and will make decisions that require no primary user interaction. However, the primary user (the passenger or operator), will still most likely monitor and interact with the autonomous system through a user interface or a remote-control room, while having the possibility to intervene in the event of malfunction or an emergency. Secondary users, however, can only experience the autonomous system's internal state and future intention through its behaviour or feedback from the operator. As Figure 1 illustrates, the secondary user experience is affected by the overall experience of the primary users and vice versa. The primary and secondary users interact through the autonomous system, based on their interaction with the robot. Interaction between

the autonomous system and the secondary users could take place via external features indicating that the pedestrian should cross the road or not, or by the pedestrian signalling to the autonomous systems that they want to cross. Actions taken by either the robotic system or the primary or secondary user will affect the co-experience, which is the user experience created in social interaction with the presence of a system or product [3].

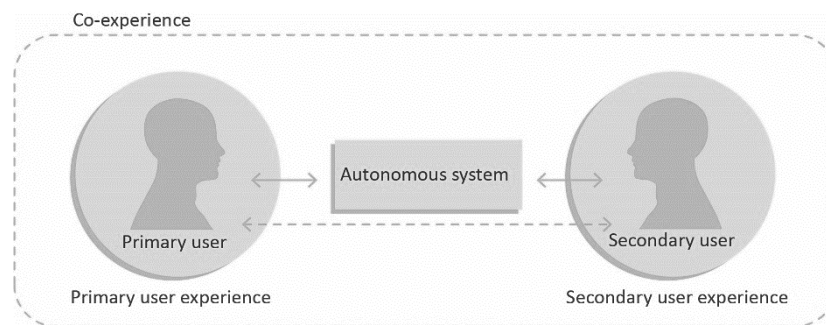


Figure 1. The solid arrows indicate direct interaction, while the dotted arrow indicates indirect interaction (adapted for autonomous systems from [1])

## 1.2 The importance of the secondary user experience

Before crossing the street, pedestrians tend to look for eye contact or any form of human communication with drivers. A study conducted by Lin et al [4] on the interaction between drivers and pedestrians at signalized intersections demonstrated that pedestrians feel more comfortable when they can interact with drivers before crossing a road. However, in the case of pedestrian-autonomous vehicle interaction, the interaction is between the robotic vehicle and the pedestrian. This could cause confusion since the robotic vehicle's design would probably be very similar to the current vehicle design, or, as in the case of self-driving cars, give the driver the opportunity to engage and disengage the autonomous system. This means that with the current seating arrangement, one could end up with a passenger in the conventional driver's seat. Unless the vehicle provides the pedestrian with adequate information, the pedestrian could become confused about whether to interact with the vehicle or the person.

In line with the study by Alsos and Svanæs [2], the co-experience will have an impact on the overall experience of a system. Based on the robot's features and communication skills, the primary and secondary users can make choices affecting the co-experience. For instance, if a pedestrian is unsure about whether to interact with the car or the passenger sitting in the driver's seat, the pedestrian could cross the road at the wrong time, creating dangerous scenarios. Today, robotic vehicles are programmed to stop immediately if something appears in front of their sensors. This sudden stop could affect both the passenger experience and their trust in the system, while at the same time lead to a possible accident with surrounding vehicles. Moreover, as stated by Wickens [5], humans are less forgiving of mistakes made by robots than by humans. This could again affect user trust and the willingness of the customer to buy or use the product.

## 2 DESIGNING FOR TRUST

In human-robot interaction (HRI), trust is considered essential in establishing and maintaining appropriate use and expectations, as it is closely associated with persuasiveness. Trust can directly affect people's willingness to accept the information communicated by the robot and to follow its suggestions [5]. Trust in HRI is closely related to research conducted on trust in automation in general. Although there are several studies on trust and automation [6-8], previous research on interaction between pedestrians and vehicles is based primarily on behaviour that involves traditional human-driven vehicles. More recent research, however, has started to explore pedestrian interaction with autonomous vehicles [9-11]. Nevertheless, few studies have focused on why it is important to design for trust for the secondary users in other fields, such as the interaction between workers and robots in a warehouse, or patients and robotic doctors/nurses in a hospital.

### 2.1 Three layers of trust

For automated systems, Marsh and Dibben [12] identify trust as something based on three factors – *dispositional trust*, *learned trust*, and *situational trust*. Dispositional trust represents the individual's

tendency to rely on automation, situational trust depends on the specific context the person finds themselves in, while learned trust is based on previous experience with a product or system. In other words, trust varies depending on the user's personality, background, experience, and the context-of-use. This makes trust an individual experience that will differ from person to person, making it a difficult factor to measure.

Based on Marsh and Dibben's [12] three layers of trust, Hoff and Bashir [6] created a theoretical model for automatic systems. Their study indicates three main factors for building trust between an operator and robot: *Initial trust*, *dynamic trust* and *situational trust*. Initial trust can directly affect both the dynamic trust and the performed human actions, as shown in Figure 2. Action performed by the secondary user affects action performed by the robot, which again affects the operator. How the robot chooses to respond and communicate its intentions based on the human actions, will in turn affect the dynamic trust. At the end of the interaction, the dynamic trust will transfer to existing knowledge and experience about the system.

With the correct use and understanding of the factors in Figure 2, one will be able to design for the right amount of trust. The right amount of trust could again lead to a better match between the expected and actual outcome, which in turn will contribute to a user experience that promotes feelings of security, safety and predictability. This will influence both the user experience, the secondary user experience and the shared co-experience.

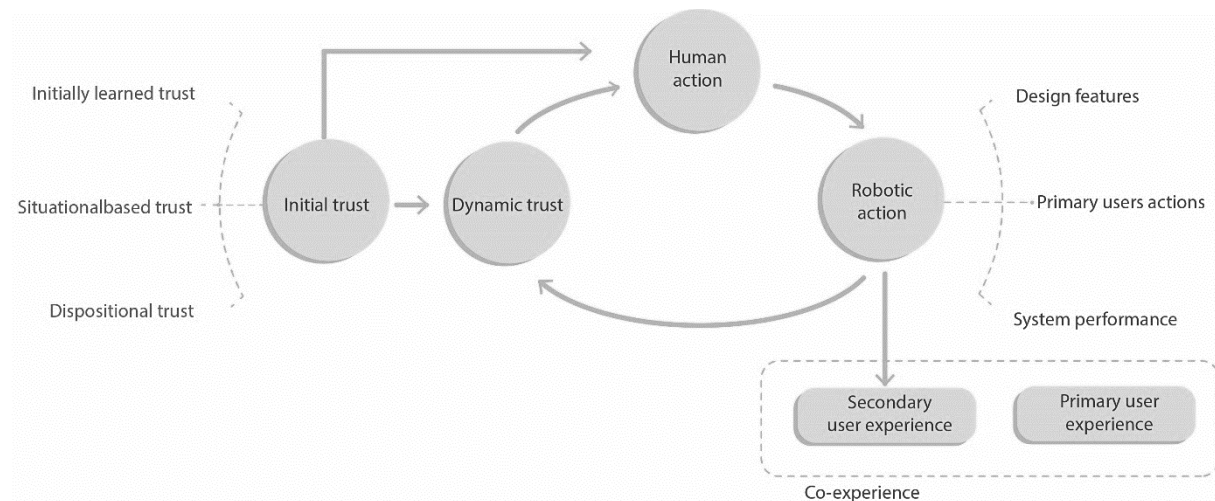


Figure 2. Factors that affect trust before and during interaction with robots. The authors' interpretation on Hoff and Bashir's research, and how it can be translated to secondary users

## 2.2 Misuse and disuse

The relationship between robots with human-like qualities and trust is uncertain, but research demonstrates that humans tend to trust robots in the same way they trust other humans. This can lead to people underestimating or overestimating the risk associated with the robots' decisions and actions in the form misuse (over-trusting) or disuse (under-trusting), which can lead to negative consequences [7,13, 14]. The less an individual trusts a robot, the faster they will intervene as it progresses towards the completion of the task [15].

There have been several serious accidents due to misplaced trust in automatic systems. An example is the Costa Concordia cruise ship disaster in 2012 which killed 32 passengers and may have been a result of the captain handing over the ship's navigation system in favour of manual control due to lack of trust in the technology [16]. A more recent accident resulted in the first pedestrian death associated with self-driving technology [17]. The autonomous taxi was a test vehicle with a safety operator behind the wheel. Several factors are thought to have played a role in the accident, but excessive trust in the system could have led the driver to allow himself to watch a video on his phone instead of monitoring the car.

Automated systems are still in the testing phase, which increases the likelihood of more accidents. Although AI-based technologies must provide adequate guarantees to minimize the risk of harm, current regulations cannot completely exclude the possibility of damage resulting from the operation of these

technologies. As stated in *Liability for AI and other new digital technologies* [18], such technologies are characterized by complexity, modifications through updates or self-learning during operation, limited predictability, and vulnerability to cybersecurity threats. One thing is the factors making it more difficult to claim compensation for victims, another that it makes it significantly more important to design for the right amount of trust. By creating an understanding of the product being under development and training, one can create an expectation which correspond to the result. Designing for the right amount of trust is a crucial part of the challenges when designing for new services such as robots and should be given more thought during the design process.

### **3 ROBOTS AND THE ROLE OF THE DESIGNER**

How do we create meaningful experiences of autonomous systems that leaves users feeling in control, and respect their values, goals, and attention? As shown in Figure 2, trust is dependant on the dispositional qualities of the user and the situation the user and robot is in, creating the need for different guidelines for different kinds of robots and systems. Robots having social interactions with humans, such as healthcare/care-taker robots, could benefit a more human-likeness approach due to the ability of establishing deeper connections with users [19]. On the other hand, robots used in industries could benefit having a machine-like design since it gives the feeling of higher levels of responsibility, possibly preventing misconceptions about the intelligence of the robot [20]. In any case, humanness can be added in the movement, appearance, sound, verbal communication, and non-verbal communication. However, human characteristics should be handled cautiously in the design process, as people anthropomorphize rather quickly [19].

A general rule is the need for a balance between the seamlessness of the interaction with the need for transparency of the AI system. In the case of non-humanoid robots, this means that secondary users are left with analysing cues such as lights, sounds, movements, and looks, in addition to analysing their environments. In other words, designing for trust is a way of designing how the robot is communicating with the world.

#### **3.1 Trust as a part of the design process**

For an automated system to work well, both primary and secondary users must be included and considered in the development process. Involving those who must deal with the future system in the design process creates better user interactions [21] and helps to optimize HRI and minimize human error [22]. Designers are trained to focus on the human in a technological system and play an important role as an intermediary between technology and psychology. However, trust is dynamic, and autonomous systems are constantly evolving, making it important to be flexible and efficient. Although designers are trained to handle future unknown problems, using different tools such as system-oriented design, design thinking and human centred design, are we using them on the right things?

While designers have limited influence in the actions of organizations for which they work, they can raise awareness of social responsibility within a company and make their voices heard. As very few products are designed and built alone, awareness of secondary users can be brought into cross disciplinary teams. One way of ensuring responsible innovation is by using The British Standards Institution's framework [23], including elements companies should consider when introducing new products, services, or processes to market. While this framework brings into the attention the importance of societal and environmental elements, and the requirement of engaging stakeholders such as end-users, it lacks the awareness of the secondary users. Both primary and secondary users must be recognized as individuals with needs, but still be part of a coherent and functional system. As people have different needs, and trust is personal, it is important to have an approach that can understand this complexity. Knowing what the different users need for information, and how to present it, creates better interaction between humans and machines and prevents disuse and misuse. Designers need to not only learn about how to design for trust, but how to design for the right amount of trust.

#### **3.2 Communicating autonomous systems intention to secondary users through external features**

Trust is reflected on how the autonomous system is communicating and interpreted by the world, making external features important in ensuring secondary user trust. The challenge lies in the ability to communicate the desired information to different kinds of users, regardless of their physical abilities, background and/or cultures. A survey from Deb et al. on interaction between pedestrians and drivers

indicated that pedestrians preferred visual signs, such as lights and screens, to indicate the vehicle's intention to stop at a crosswalk [24]. Following the study, several researchers and studies have proposed different strategies and design features to ensure the safety of pedestrians in HRI. One of those being 'Blink' from The Imperial College of London and the Royal College, allowing a two-way communication recognizing the pedestrians hand gestures to stop or keep moving [25]. However, this relies heavily on the pedestrians' learned trust and their ability to make hand gestures. On the other hand, proposals such as Simcons' prototype 'smiling car' [26], emphasizes/relies on the use of sight where the robotic vehicle will light up in a smile to indicate that the car will stop, and the pedestrian can cross the road. Since future autonomous systems will operate in public places among other technological systems, humans and animals, the systems do not only have to speak to each other but ensure a universal design able to communicate in similar manners to reduce the chances of confusion.

### 3.3 Trust as part of the Design Education

The traditional user interface through screens and physical buttons and knobs will go through a transformation with the introduction of autonomous systems and robots. This will require a different take on design education. As we have argued above, trust is an important aspect of the interaction between users and AI-based technologies. A designer needs to understand how to design for trust in AI-systems and robots through means such as lights, sound, movement, non-verbal and verbal communication, etc. This must be an important part of the designer's curriculum.

Even though we might not know what kind of autonomous systems and robots the future will bring, we do know that they will interact and operate among humans. To create a more humane future, we need a design education with humans in mind. This includes the understanding of the bigger picture; how the use of autonomous systems/technology has financial, ecological, and social consequences, and how psychological elements such as trust plays an important role in the user acceptance of autonomous systems.

## 4 CONCLUDING REMARKS

A fully autonomous system is technically independent, but still – as in the case of robotic vehicles – depends on people's trust to function properly, both primary users (i.e., operators), and secondary users (i.e., pedestrians). Appropriate design features need to be developed to ensure secondary user trust, and scientific evaluation of their effectiveness must be carried out. Trust in autonomous systems will improve if humans understand how robots are going to behave in their presence, for example by communicating actions through external features such as movement, appearance, sound, verbal and non-verbal communication. By designing for the right amount of trust, one can prevent users from getting the impression that the system works better than it does, which can have a negative effect on user experiences and in the worst-case lead to dangerous situations. With the right amount of trust, human-robot interactions will improve, thus making the system more understandable and safer, while reducing human error for both primary and secondary users.

## REFERENCES

- [1] Freedy A., Devisser E., Weltman G., and Coeyman N. Measurement of trust in human-robot collaboration. In, *International Symposium on Collaborative Technologies and Systems*, Orlando, May 2007.
- [2] Alsos O. A. and Svanæs D. Designing for secondary user experience. In *INTERACT 2011: Human-Computer Interaction*, part IV, pp 84-91.
- [3] Battarbee K. Defining Co-Experience. In *Proceedings of DPPI*. ACM, New York, June 2003.
- [4] Lin P. S., Wang Z., Guo R. and Kourtellis A. A pilot study on interactions between drivers and pedestrian features at signalized intersections – using the SHRP2 naturalistic driving study data. In *Proc. 11th Asia Pac. Transp. Develop. Conf. 29th ICTA Annu. Conf.*, 2016, pp. 70-77.
- [5] Wickens C. D., Lee J., Gordon-Becker S., and Liu Y. D. (2004). *An introduction to human factors engineering*. Pearson Education International, 2004 (Pearson, Upper Saddle River, NJ).
- [6] Hoff K. A. and Bashir M. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. In *Human Factors: The Journal of the Human Factors and Ergonomics Society*, September 2014, pp. 407–434.
- [7] Lee J. D. and See K. A. Trust in automation: Designing for appropriate reliance. In *Human Factors: The Journal of the Human Factors and Ergonomics Society*, volume 46, March 1, 2004,

pp 50–80.

- [8] Parasuraman R, Sheridan T. B. and Wickens C. D. Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs. *Journal of Cognitive Engineering and Decision Making*, volume 2, 2008, pp. 140–160.
- [9] Pillai A. Virtual reality-based study to analyse pedestrian attitude towards autonomous vehicles. *Master of Science thesis, KTH Royal institute of technology*, 2017.
- [10] Deb S., Rahnman M. M., Strawderman L. J., and Garrison T-M. Pedestrians' receptivity toward fully automated vehicles: research and roadmap for future research. In *IEEE Transaction on human-machine systems*, 48, 2018, pp 279-290.
- [11] Fuest T., Michalowski L., Tr'aris L., Bellem H., and Bengler K. Using the driving behaviour of an automated vehicle to communicate intentions – A wizard of Oz study. In *21st International Conference on Intelligent Transportation Systems (ITSC)*, Maui, USA, November 2018.
- [12] Marsh S. and Dibben M. R. The role of trust in information science and technology. *Annual Review of Information Science and Technology*, 2003, 37, pp. 465–498.
- [13] Parasuraman R. and Manzey D. Complacency and bias in human use of automation: An attentional integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 2010, volume 52, pp. 381–410.
- [14] Parasuraman R. and Riley V. Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 1997, 39(2): pp. 230-253.
- [15] Steinfeld A., Fong T., Kaber D., Lewis M., Scholtz J., Schultz A., and Goodrich M. Common metrics for human-robot interaction. In *Proceedings of 2006 ACM Conference on Human-Robot Interaction*, 2006, pp. 33–40.
- [16] Schröder-Hinrichs J. U., Hollnagel E., and Baldauf M. From Titanic to Costa Concordia—a century of lessons not learned. *WMU J Marit Affairs* 11: pp- 151–167 (2012).
- [17] Wakabayashi D. Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam. Available: <https://www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html> [Accessed on 2020, 24 September] (2018, 19 March).
- [18] Directorate-General for Justice and Consumers. *Liability for Artificial Intelligence and Other Emerging Digital Technologies*, 2019 (Publications Office of the European Union).
- [19] Darling K. 'Who's Johnny?' Anthropomorphic framing in human-robot interaction, integration, and policy, *Robot ethics 2nd edition ds. P. Lin, G. Bekey, K. Abney, R. Jenkins, Oxford University Press, 2017, Forthcoming*.
- [20] Hinds P., Roberts T. and Jones H. Whose job is it anyway? A study of human-robot interaction in a collaborative task. In *Human Computer Interaction*, 2004, pp. 151-181.
- [21] Marsh S. and Meech J. Trust in design, in *CHI '00 Extended Abstracts on Human Factors in Computing Systems*. 2000, ACM: The Hague, The Netherlands. pp. 45-46.
- [22] Oviatt S. Human-centred design meets cognitive load theory: designing interfaces that help people think, In *Proceedings of the 14th ACM international conference on Multimedia*. 2006, ACM: Santa Barbara, CA, USA. pp. 871- 880.
- [23] British Standard Institution 2020, *Responsible Innovation Guide*, PAS 440:2020, BSI Standards Limited 2020, viewed 3 May 2021, retrieved from BSI group.
- [24] Deb S., Warner B., Poudel S., and Bhandari S. Identification of external design preferences in autonomous vehicles. In *Proc. IIE Res. Conf.*, 2016, pp. 69-44.
- [25] Peters A., *What if driverless cars let you cross the street when you wave at them?* Available: <https://www.fastcompany.com/3068166/what-if-driverless-cars-let-you-cross-the-streetwhen-you-wave-at-them> [Accessed on 2020, October 18] (2017, 16. February).
- [26] Snyder J. B. *This self-driving car smiles at pedestrians*. Available: <https://www.autoblog.com/2016/09/16/thisself-driving-car-smiles-at-pedestrians> [Accessed at 2020, October 15] (2016, 16 September).