



## The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle



Adam Waytz<sup>a,\*</sup>, Joy Heafner<sup>b</sup>, Nicholas Epley<sup>c</sup>

<sup>a</sup> Northwestern University, USA

<sup>b</sup> University of Connecticut, USA

<sup>c</sup> University of Chicago, USA

### HIGHLIGHTS

- Anthropomorphism of a car predicts trust in that car.
- Trust is reflected in behavioral, physiological, and self-report measures.
- Anthropomorphism also affects attributions of responsibility/punishment.
- These findings shed light on human interaction with autonomous vehicles.

### ARTICLE INFO

#### Article history:

Received 7 August 2013

Revised 11 January 2014

Available online 23 January 2014

#### Keywords:

Anthropomorphism

Mind perception

Trust

Moral responsibility

Human–computer interaction

Dehumanization

### ABSTRACT

Sophisticated technology is increasingly replacing human minds to perform complicated tasks in domains ranging from medicine to education to transportation. We investigated an important theoretical determinant of people's willingness to trust such technology to perform competently—the extent to which a nonhuman agent is anthropomorphized with a humanlike mind—in a domain of practical importance, autonomous driving. Participants using a driving simulator drove either a normal car, an autonomous vehicle able to control steering and speed, or a comparable autonomous vehicle augmented with additional anthropomorphic features—name, gender, and voice. Behavioral, physiological, and self-report measures revealed that participants trusted that the vehicle would perform more competently as it acquired more anthropomorphic features. Technology appears better able to perform its intended design when it seems to have a humanlike mind. These results suggest meaningful consequences of humanizing technology, and also offer insights into the inverse process of objectifying humans.

© 2014 Elsevier Inc. All rights reserved.

### Introduction

Technology is an increasingly common substitute for humanity. Sophisticated machines now perform tasks that once required a thoughtful human mind, from grading essays to diagnosing cancer to driving a car. As engineers overcome design barriers to creating such technology, important psychological barriers that users will face when using this technology emerge. Perhaps most important, will people be willing to trust competent technology to replace a human mind, such as a teacher's mind when grading essays, or a doctor's mind when diagnosing cancer, or their own mind when driving a car?

Our research tests one important theoretical determinant of trust in any nonhuman agent: anthropomorphism (Waytz, Cacioppo, & Epley, 2010). Anthropomorphism is a process of inductive inference whereby people attribute to nonhumans distinctively human characteristics,

particularly the capacity for rational thought (agency) and conscious feeling (experience; Gray, Gray, & Wegner, 2007). Philosophical definitions of personhood focus on these mental capacities as essential to being human (Dennett, 1978; Locke, 1997). Furthermore, studies examining people's lay theories of humanness show that people define humanness in terms of emotions that implicate higher order mental process such as self-awareness and memory (e.g., humiliation, nostalgia; Leyens et al., 2000) and traits that involve cognition and emotion (e.g., analytic, insecure; Haslam, 2006). Anthropomorphizing a nonhuman does not simply involve attributing superficial human characteristics (e.g., a humanlike face or body) to it, but rather attributing *essential* human characteristics to the agent (namely a humanlike mind, capable of thinking and feeling).

Trust is a multifaceted concept that can refer to belief that another will behave with benevolence, integrity, predictability, or competence (McKnight & Chervany, 2001). Our prediction that anthropomorphism will increase trust centers on this last component of trust in another's competence (akin to confidence) (Siegrist, Earle, & Gutscher, 2003; Twyman, Harvey, & Harries, 2008). Just as a patient would trust a

\* Corresponding author at: Northwestern University, 2001 Sheridan Rd, Evanston, IL 60208, USA.

E-mail address: [a-waytz@kellogg.northwestern.edu](mailto:a-waytz@kellogg.northwestern.edu) (A. Waytz).

thoughtful doctor to diagnose cancer more than a thoughtless one, or would rely on mindful cab driver to navigate through rush hour traffic more than a mindless cab driver, this conceptualization of anthropomorphism predicts that people would trust easily anthropomorphized technology to perform its intended function more than seemingly mindless technology. An autonomous vehicle (one that drives itself) for instance, should seem better able to navigate through traffic when it seems able to think and sense its surroundings than when it seems to be simply mindless machinery. Or a “warbot” intended to kill should seem more lethal and sinister when it appears capable of thinking and planning than when it seems to be simply a computer mindlessly following an operator’s instructions. The more technology seems to have humanlike mental capacities, the more people should trust it to perform its intended function competently, regardless of the valence of its intended function (Epley, Caruso, & Bazerman, 2006; Pierce, Kilduff, Galinsky, & Sivanathan, 2013).

This prediction builds on the common association between people’s perceptions of others’ mental states and of competent action. Because mindful agents appear capable of controlling their own actions, people judge others to be more responsible for successful actions they perform with conscious awareness, foresight, and planning (Cushman, 2008; Malle & Knobe, 1997) than for actions they perform mindlessly (see Alicke, 2000; Shaver, 1985; Weiner, 1995). Attributing a humanlike mind to a nonhuman agent should therefore more make the agent seem better able to control its own actions, and therefore better able to perform its intended functions competently. Our prediction also advances existing research on the consequences of anthropomorphism by articulating the psychological processes by which anthropomorphism could affect trust in technology (Nass & Moon, 2000), and by both experimentally manipulating anthropomorphism as well as measuring it as a critical mediator. Some experiments have manipulated the humanlike appearance of robots and assessed measures indirectly related to trust. However, such studies have not measured whether such superficial manipulations actually increase the attribution of essential humanlike qualities to that agent (the attribution we predict is critical for trust in technology; Hancock et al., 2011), and therefore cannot explain factors found ad-hoc to moderate the apparent effect of anthropomorphism on trust (Pak, Fink, Price, Bass, & Sturre, 2012). Another study found that individual differences in anthropomorphism predicted differences in willingness to trust technology in hypothetical scenarios (Waytz et al., 2010), but did not manipulate anthropomorphism experimentally. Our experiment is therefore the first to test our theoretical model of how anthropomorphism affects trust in technology.

We conducted our experiment in a domain of practical relevance: people’s willingness to trust an autonomous vehicle. Autonomous vehicles—cars that control their own steering and speed—are expected to account for 75% of vehicles on the road by 2040 (Newcomb, 2012). Employing these autonomous features means surrendering personal control of the vehicle and trusting technology to drive safely. We manipulated the ease with which a vehicle, approximated by a driving simulator, could be anthropomorphized by merely giving it independent agency, or by also giving it a name, gender, and a human voice. We predicted that independent agency alone would make the car seem more mindful than a normal car, and that adding further anthropomorphic qualities would make the vehicle seem even more mindful. More important, we predicted that these relative increases in anthropomorphism would increase physiological, behavioral, and psychological measures of trust in the vehicle’s ability to drive effectively.

Because anthropomorphism increases trust in the agent’s ability to perform its job, we also predicted that increased anthropomorphism of an autonomous agent would mitigate blame for an agent’s involvement in an undesirable outcome. To test this, we implemented a virtually unavoidable accident during the driving simulation in which participants were struck by an oncoming car, an accident clearly caused by the other driver. We implemented this to maintain experimental

control over participants’ experience because everyone in the autonomous vehicle conditions would get into the same accident, one clearly caused by the other driver. Indeed, when two people are potentially responsible for an outcome, the agent seen to be more competent tends to be credited for a success whereas the agent seen to be less competent tends to be blamed for a failure (Beckman, 1970; Wetzel, 1982). Because we predicted that anthropomorphism would increase trust in the vehicle’s competence, we also predicted that it would reduce blame for an accident clearly caused by another vehicle.

## Experiment

### Method

One hundred participants (52 female,  $M_{age} = 26.39$ ) completed this experiment using a National Advanced Driving Simulator. Once in the simulator, the experimenter attached physiological equipment to participants and randomly assigned them to condition: Normal, Agentic, or Anthropomorphic. Participants in the Normal condition drove the vehicle themselves, without autonomous features. Participants in the Agentic condition drove a vehicle capable of controlling its steering and speed (an “autonomous vehicle”). The experimenter followed a script describing the vehicle’s features, suggesting when to use the autonomous features, and describing what was about to happen. Participants in the Anthropomorphic condition drove the same autonomous vehicle, but with additional anthropomorphic features beyond mere agency—the vehicle was referred to by name (Iris), was given a gender (female), and was given a voice through human audio files played at predetermined times throughout the course. The voice files followed the same script used by the experimenter in the Agentic condition, modified where necessary (See Supplemental Online Material [SOM]).

All participants first completed a driving history questionnaire and a measure of dispositional anthropomorphism (Waytz et al., 2010). Scores on this measure did not vary significantly by condition, so we do not discuss them further.

Participants in the Agentic and Anthropomorphic conditions then drove a short practice course to familiarize themselves with the car’s autonomous features. Participants could engage these features by pressing buttons on the steering wheel. All participants then drove two courses each lasting approximately 6 min. After the first course, participants completed a questionnaire (all on 0–10 scales, see SOM for all items) that assessed anthropomorphism, liking, and trust.

### Perceived anthropomorphism

Four items measured anthropomorphism, defined as attributing humanlike mental capacities of agency and experience to it (Epley, Waytz, & Cacioppo, 2007; Gray et al., 2007; Waytz et al., 2010). These asked how *smart* the car was, how well it could *feel* what was happening around it, how well it could *anticipate* what was about to happen, and how well it could *plan* a route. These items were averaged into a composite ( $\alpha = .89$ ).

### Liking

Four items measured liking: how enjoyable their driving was, how comfortable they felt driving the car, how much participants would like to own a car like this one, and what percentage of cars in 2020 they would like to be [autonomous] like this one. These items were standardized and averaged to form a single composite ( $\alpha = .90$ ).

### Self-reported trust

Eight items measured trust in the vehicle: how safe participants felt they and others would be if they actually owned a car like this one, how much they trust the vehicle to drive in heavy and light traffic conditions, how confident they are about the car driving the next course safely, and their willingness to give up control to the car. These items were standardized and averaged to form a single composite ( $\alpha = .91$ ).

Download English Version:

<https://daneshyari.com/en/article/947765>

Download Persian Version:

<https://daneshyari.com/article/947765>

[Daneshyari.com](https://daneshyari.com)