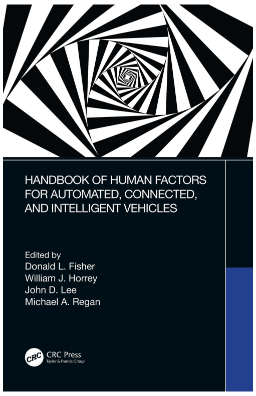

Handbook of Human Factors for Automated, Connected, and Intelligent Vehicles

Donald L. Fisher, William J. Horrey, John D. Lee, Michael A. Regan

Driver Trust in Automated, Connected, and Intelligent Vehicles

Publication details
https://test.routledgehandbooks.com/doi/10.1201/b21974-4
John D. Lee

4 Driver Trust in Automated, Connected, and Intelligent Vehicles

John D. Lee
University of Wisconsin-Madison

CONTENTS

KEY POINTS

- Trust complements the concepts of mental models and situation awareness in explaining how and when people rely on automation.
- Trust is a multi-faceted term that operates at timescales of seconds to years and mediates how people rely on, accept, and tolerate vehicle technology.
- Trust mediates micro interactions concerning how people rely on automation to engage in non-driving tasks to macro interactions concerning how the public accepts new forms of transport.

- Trust depends on surface features, such as the look and feel of the human–machine interface (HMI), as well as depth features, such as the reliability of the automation and alignment of the person's and automation's goals.
- The imperfect goal alignment between primary users (e.g., riders in an automated vehicle (AV)) and incidental users (e.g., pedestrians negotiating intersections with AVs) might undermine the trust and tolerance of incidental users.

4.1 INTRODUCTION

In the previous chapters, we described the different types of automated, connected, and intelligent vehicle technologies and the mental models that inform the driver's decisions and behaviors. The discussion of mental models (this Handbook, Chapter 3) shows that it is one thing to develop technologies that can greatly benefit the driver and society, but it is an entirely different matter for these technologies to be used appropriately. Like mental models, trust is one factor that guides how drivers use vehicle automation, as well as the public acceptance and tolerance of driverless vehicles. This chapter addresses how one goes about building, calibrating, and repairing trust for different types of automation.

This chapter begins by describing different types of automation and different roles of people from the perspective of trust. This is followed by a definition of trust, a description of the cognitive mechanisms that underlie it, and the need to promote appropriate trust. Design approaches to promote appropriate trust depend on various relationships between people and vehicle technology, such as drivers monitoring fallible self-driving vehicles, passengers riding in driverless vehicles, people relying on algorithms to choose ride-sharing partners, and even pedestrians who must negotiate intersections with automated vehicles (AV). These human–technology relationships are described in terms of general categories of vehicle technology—shared, traded, and ceded control. These categories are relevant for trust because they reflect increasing degrees to which people make themselves vulnerable to the technology. Considering these categories, we discuss the roles of people, why trust is relevant, the basis of trust, and the design choices that might promote appropriate trust and repair trust after automation mishaps. The chapter concludes by briefly addressing the ethical considerations associated with managing trust and creating trustworthy technology.

4.2 TRUST AND TYPES OF AUTOMATION

Technology is rapidly changing the relationship between people and vehicles. Increasingly, this technology is supporting drivers in new ways and is also taking responsibility for many of the aspects of driving that were once the sole responsibility of drivers. In the extreme, this has produced driverless vehicles that operate with no direct control input from people in the vehicle and the automation takes full responsibility for driving—SAE Level 5 automation. Vehicles with Level 5 automation might not even include a steering wheel or pedals and people *cede control* to the vehicle—drivers become riders. Even more extreme are vehicles that deliver goods and run errands without occupants. Less extreme are intermittently self-driving vehicles where drivers can *trade control* with automation—SAE Level 3 automation. Here, the vehicle might drive itself for part of the journey, but drivers would be able to take control and might be called upon by the vehicle to take control. Many vehicles contain less ambitious automation where drivers *share control* with the driver. Here the driver remains completely responsible for driving, but the automation eases the demands of driving and might allow for short periods of less vigilant attention to the roadway—SAE Levels 1 and 2 automation. Shared control includes adaptive cruise control (ACC) and lane-centering systems that the driver deliberately engages, as well as systems that engage automatically, such as electronic stability control and automatic emergency braking, which activate only in rare and extreme conditions. The terms ceded, traded, and shared are used because they reflect how automation puts people in different positions of vulnerability and uncertainty, which engages trust in guiding behavior (Mayer, Davis, & Schoorman, 1995). Whether in the form of ceded, traded, or shared control, vehicle automation is changing the role and responsibility of drivers, and the theoretical construct of trust will likely play a critical role in mediating how people adapt to this new technology (see also, this Handbook, Chapters 6, 8, 9).

The people for whom vehicle automation is designed—the drivers or riders—are the most obvious users whose trust in it will influence its success, but vehicle automation and people reside in a large interconnected network, see Figure 4.1 (Lee, 2018). In this context, even driverless vehicles, where drivers have ceded control to the vehicle, are not actually autonomous. Driverless vehicles will depend on an array of people to maintain it, repair it, and remotely operate it when the on-board automation encounters situations that exceed its capacity. Here trust might influence whether a remote operator chooses to intervene and "drive" the vehicle, much like trust influences those who manage remotely piloted drones (Guznov, Lyons, Nelson, & Woolley, 2016). Another critical trust relation emerges when people share rides in an AV. Similar to other situations in the sharing economy, trust can inhibit or facilitate people's willingness to accept the suggestion of an algorithm to share a ride with a stranger (Ert, Fleischer, & Magen, 2016; Payyanadan & Lee, 2018). Other nodes on this network that automation may affect are other road users, such as pedestrians and cyclists, as well as drivers of conventional vehicles. These other road users are not the primary users of automation, where the automation has been designed to achieve their goals, but are incidental or passive users, who did not choose to use the automation and whose goals may conflict with those of the AVs and their riders (Inbar & Tractinsky, 2009; Montague, Xu, & Chiou, 2014).

Changes to one node can ripple through the network and affect trust in unanticipated ways. For example, pedestrians who grow distrustful and intolerant of AVs might bully or sabotage them, undermining their efficiency—early demonstrations of self-driving vehicles have encountered hostile people who have vandalized them (Romero, 2018). The diminished efficiency might then undermine the trust of the riders when they fail to arrive at their destinations on time. The many nodes of the network that define the transportation ecosystem mean it is insufficient to consider only the relationship between the person in the vehicle and the vehicle technology (Woods, 2015).

Figure 4.1 shows many trust relationships linking the nodes of a multi-echelon network. The echelons range from the network of sensors and computers that reside

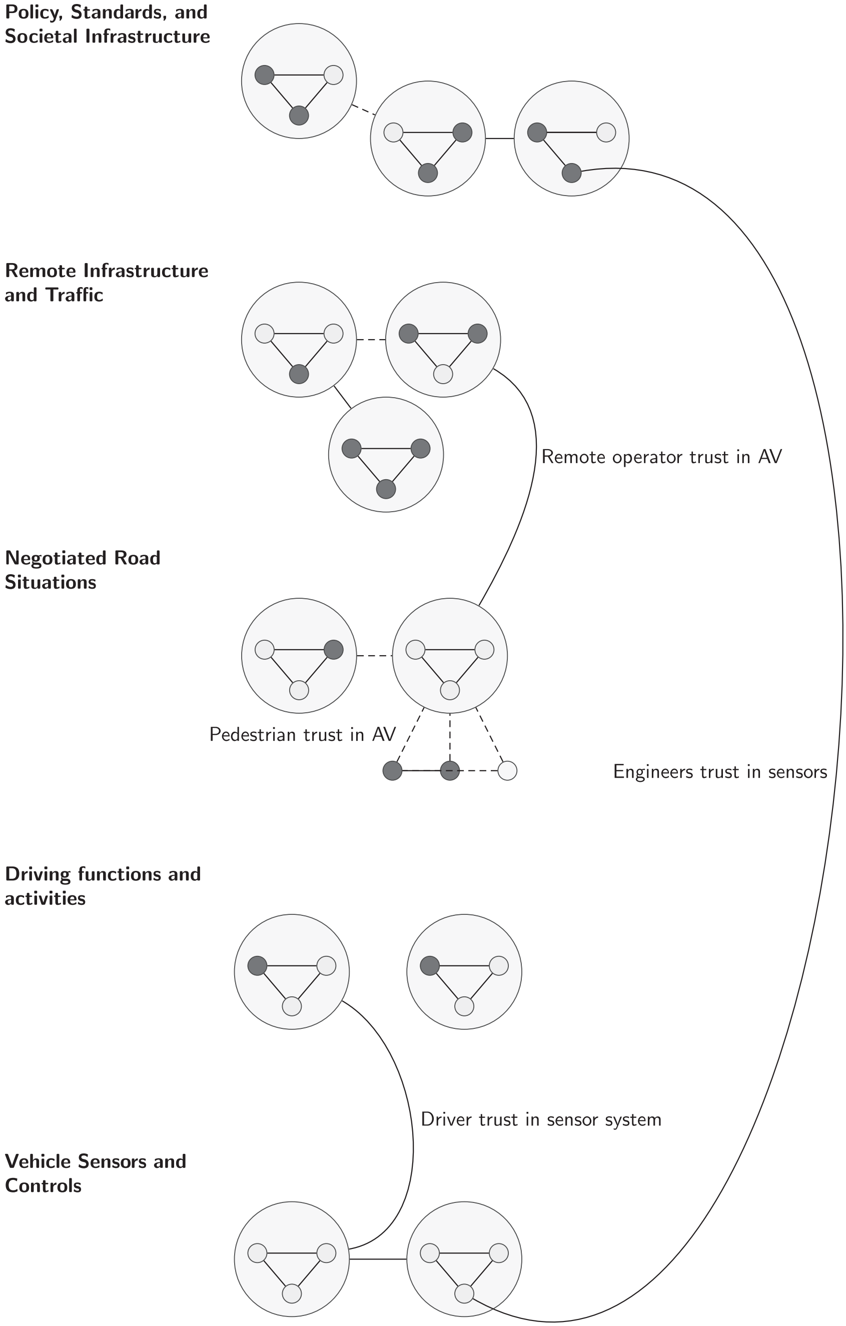

FIGURE 4.1 A multi-echelon network of trust relationships and vehicle technology. Filled circles indicate people and open circles represent automation elements. The lines indicate trust relationships, with the solid lines indicating likely goal alignment and dashed lines indicate relationships where goal alignment is less likely.

in the car to the network of organizations and societal infrastructure that guides technology development, deployment, and regulation. The influence of trust in these relationships depends on the type of vehicle technology and the role of the person. The most commonly studied trust relationships are those between people and the AV system, shown Figure 4.1 as the "Driving Functions and Activities." Other trust relationships span levels, such as between a remote operator of self-driving vehicles at the level of "Remote Infrastructure and Traffic" and the riders involved in the traffic situation at the level of "Negotiated Road Situations." Some trust relationships even span the extremes, as the engineering community at the level of "Policy, Standards, and Societal Infrastructure" and the behavior of sensors at the level of "Vehicle Sensors and Control." Generally, technology serves the goal of the person directly interacting with it, but in this network the goals of some people might be at odds with the automation—shown as dotted lines. For example, the goals of pedestrians negotiating an intersection will often conflict with those of an AV that is serving its rider's goal of minimizing travel time. This figure shows only a few of the many trust relationships that comprise the transportation ecology to highlight the range of situations where trust in technology plays a role.

These trust relationships can suffer from either over-trust or under-trust. For example, with shared control (SAE Levels 1 and 2), technology supports people in their role as drivers (SAE, 2016). Here the danger is that people will over-trust automation and rely on it to perform in ways the designers did not intend. As an example, some drivers over-rely on the Tesla automation, leading to extended periods of disengagement from driving and several fatal crashes (NTSB, 2016). For safety drivers and remote operators who supervise AVs and intervene when the automation encounters unexpected difficulties the situation is similar to shared control: the danger is that they will over-trust the technology and rely on it to perform safety critical driving functions when they should not.

Over-trust with shared control contrasts with under-trust with ceded control (SAE Level 5), where the person's role might be that of a passenger who is unable to intervene and take back the driving task. Here the danger is that people will under-trust the automation, reject it, and fail to enjoy its benefits. For example, people might not trust that self-driving vehicles can drive safely and opt to drive themselves. Trust is needed not just for riding in an AV, but also for requesting an AV, entering and initiating a trip, making trip changes while en route, and safely pulling over and exiting at the destination (Weast, Yurdana, & Jordan, 2016). Figure 4.1 highlights one such relationship, where riders will need to trust the vetting and matching algorithms associated with making shared rides safe. Here people must develop trust in their fellow passengers through the technology used to arrange the ride (Ert et al., 2016).

In contrast with riders and drivers who are primary users of technology, the goals of incidental users, such as pedestrians, might not align with the automation, which could undermine their trust of AVs even if they behave safely. For example, AVs might fail to give way to pedestrians waiting to cross the street, forcing the pedestrians to wait longer than they might with manually driven vehicles. Another class of incidental users is the general public who must share its transportation infrastructure with AVs. Public anger erupted when buses used to shuttle large technology employees in Silicon Valley clogged public bus lanes (De Kosnik, 2014). This anger is an indicator of potential distrust that might emerge if people see AVs as appropriating public resources. The factors affecting trust in technology differ in each of these situations, which can be revealed by considering the psychological mechanisms underlying trust.

4.3 DEFINITION AND MECHANISMS UNDERLYING TRUST

As with most sophisticated technology, acceptance of vehicle automation depends not only on an explicit rational consideration of its features and capabilities but also on an emotional or affective response to its behavior and branding (Lee, 2006; Norman, 2004). Central to this affective response is trust. Trust has emerged as an important concept in explaining how people adapt to and accept new technology in domains as diverse as industrial process automation (Muir & Moray, 1996; Sheridan & Ferrell, 1974; Zuboff, 1988), e-commerce (McKnight, Choudhury, & Kacmar, 2002), weapon systems (Dzindolet, Pierce, Beck, Dawe, & Anderson, 2001), and, at a more macro level, societal acceptance of technology risk (Slovic, 1993; 1999) and participation in the sharing economy (Ert et al., 2016). Many factors influence adaptation and acceptance, but trust is particularly critical in systems that take control in uncertain and risky situations (also see this Handbook, Chapter 5).

A comprehensive review of trust in automation considered the literature on human–human and human–technology trust to synthesize a definition of human–technology trust: "... the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability…" (Lee & See, 2004, p. 54). A more recent review of trust in automation reiterated this definition and emphasized that trust formation is a dynamic process (Hoff & Bashir, 2015). These reviews along with several meta-analyses document a substantial empirical base that shows that greater levels of trust lead to greater reliance on and compliance with the automation. These reviews also show that trust is not simply a reflection of the reliability of the automation (Schaefer, Chen, Szalma, & Hancock, 2016; Wu, Zhao, Zhu, Tan, & Zheng, 2011). Trust is an important influence on how people use, accept, and tolerate vehicle technology, but related concepts also play an important role.

Figure 4.2 shows trust in the context of Neisser's perceptual cycle. Trust directs trust-related actions (e.g., reliance, compliance, and tolerance), and these actions enable people to sample the characteristics of the automation (e.g., purpose, process, and performance), and these samples modify trust (through affective, analogical, and analytical cognitive mechanisms). This cycle repeats, with the modified trust, directing trust-related actions, which in turn modifies the level of trust. More specifically, reliability is an aspect of automation performance, which influences trust, and trust influences trust action, such as reliance. Trust depends on other factors than reliability and other factors beyond trust affect reliance (Lee & See, 2004). For example, trust might be influenced by an understanding of how the automation operates (i.e., the automation process) and reliance depends on self-confidence in the ability to perform the task manually. Trust is typically defined in terms of the automation achieving a particular goal of a person, which guides reliance and compliance. With incidental users, whose goals do not align with the automation, tolerance is the willingness to interact with automation that might

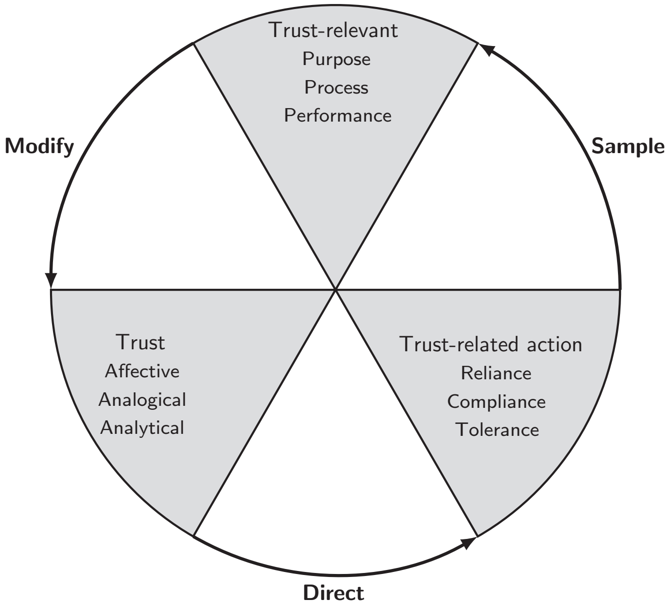

FIGURE 4.2 Trust-related actions, the trust-related information revealed by these actions, and the cognitive mechanisms underlying trust in the context of Neisser's perceptual cycle. (Neisser, 1976.)

interfere with achieving their goals. Tolerance is critical in considering how pedestrians and other drivers might respond to AVs. The balance of this section expands on the cognitive mechanisms of trust and by first relating trust to mental models and situation awareness.

Trust has close connections to other important concepts regarding driver interaction with automation, such as mental models (see this Handbook, Chapter 3) and situation awareness (see this Handbook, Chapter 7). One definition of mental models is an explicit verbalizable mental representation of a system that allows people to anticipate system behavior and to plan actions to achieve desired results (Norman, 1983; Rouse & Morris, 1986). Mental models support simulation of possible future states of the system to guide attention and action. According to this conception of mental models, trust complements them as an implicit, non-verbalizable, representation that guides behavior through an affective response to the system (Lee, 2006). For example, a feeling of diminished trust might lead the driver to suspect that the lane-keeping system is not operating properly and the driver's mental model might lead the driver to recognize that rain is interfering with the sensors. Trust also differs from mental models in that it guides interactions with intentional agents, or with agents whose complexity and autonomy makes them appear intentional, whereas mental models tend to guide interactions with inanimate systems—you might have a mental model of a toaster but trust a voice-based assistant. A toaster requires a mental model to guide specific interactions, but a voice-based assistant requires a belief that it will act and reliably provide information.

Situation awareness can be thought of as the elements of the mental models that are active in the moment, and so guides attention and behavior in a particular situation (Endsley, 2013). It also encompasses the perception, interpretation, and projection of the current system state (Endsley, 1995). All of these reflect the person's mental model and the assimilation of information regarding the situation. Situation awareness, as defined as an explicit awareness of the system, is often critical for guiding decision-making and effective behavior. However, much behavior is guided by implicit knowledge and affective processes that produce advantageous decisions without awareness for the cues that guide them (Bechara, Damasio, Tranel, & Damasio, 1997; Kahneman, 2011).

Figure 4.3 builds on Figure 4.2 to show how trust complements the influence of mental models and situation awareness to influence behavior. The cognitive state of the person, defined by trust, mental model, and situation awareness directs perceptual exploration, specific actions, and, more generally, behavior. This behavior samples the world and, more specifically, the operational domain and device, as well as the specific situation. These samples modify the cognitive state—trust, mental model, and situation awareness.

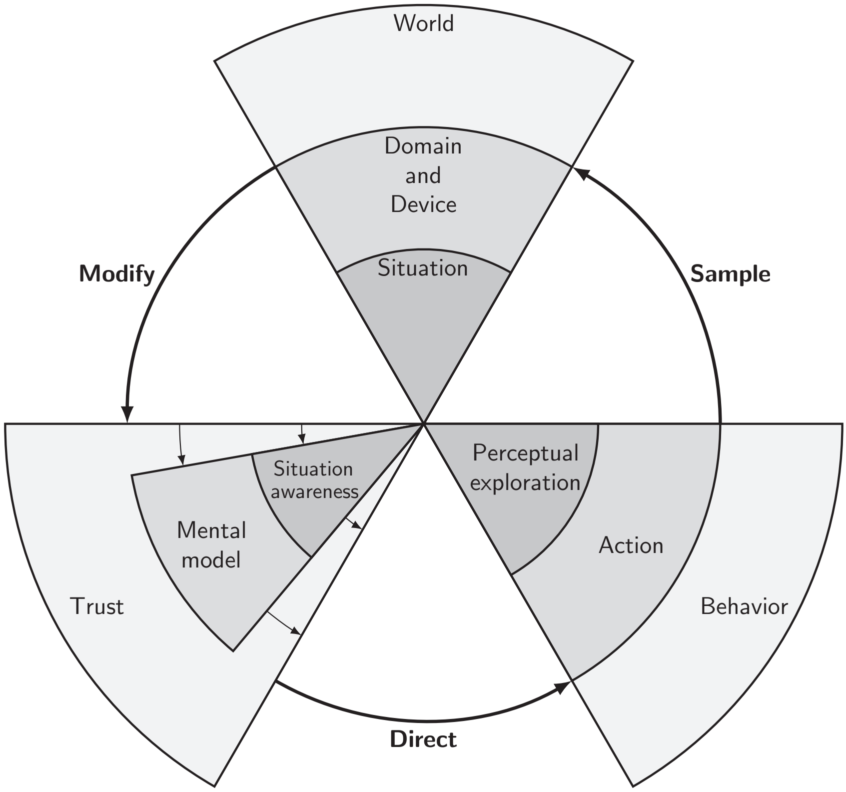

FIGURE 4.3 Links between situation awareness, mental models, and trust based on Neisser's perceptual cycle. (Neisser, 1976.)

Trust influences and is influenced by mental models and situation awareness through an affective process where attitudes and feelings guide response to technology (Gefen, Karahanna, & Straub, 2003; Lee & See, 2004). Trust in a system will shape the mental model and situation awareness of that system, and mental models and situation awareness affect trust. Figure 4.3 shows this joint influence with the small arrows from trust to mental models and situation awareness and from situation awareness and mental models to trust. Trust influences mental model and situation awareness by guiding attention to particular aspects of a system that might require the person's action, such as when the intermittent disengagement of ACC might guide attention of the driver to the display to understand why it disengages (Lee, 2006). This attention to the display contributes to a more complete mental model and better situation awareness. Figure 4.3 shows how mental models, situation awareness, and trust influence each other and jointly guide how people interact with automation.

The theoretical basis of trust in technology is grounded in research addressing trust between people (Muir, 1987; Sheridan & Ferrell, 1974). Trust mediates interactions between individuals and between individuals and organizations (Mayer et al., 1995). In interpersonal relationships, trust acts as a social lubricant that supports cooperation and exchanges in situations where it is not possible to exert complete control over those who are being relied upon to achieve a goal (Rotter, 1980). The substantial research concerning the role of trust in human-to-human relationships indicates that trust is a social emotion that underlies an affective process of responding to social uncertainty and risk (Adolphs, Tranel, & Damasio, 1998; Fehr & Camerer, 2007; Mikolajczak et al., 2010). As an affective process it can act pre-attentively to influence behavior without effort and in a way that can differ from the influence of conscious processes. As an example, a brain imaging study showed that different areas of the brain were active during explicit and implicit judgments of trustworthiness, with the brain activity associated with the implicit judgments indicating an affective process (Winston, Strange, O'Doherty, & Dolan, 2002). This affective process is partially governed by neurochemicals, such as oxytocin, that affect how we feel about others and how we respond to their behavior. Dosing people with oxytocin led to increased social risk-taking, but not risk-taking in general (Kosfeld, Heinrichs, Zak, Fishbacher, & Fehr, 2005). However subsequent studies have failed to replicate these findings, suggesting nasal administration of oxytocin has a small and variable effect on trust (Nave, Camerer, & McCullough, 2015). More generally, trust guides how people relate to other people, teams, organizations, and even governments to achieve their goals in situations characterized by risk and uncertainty (Rousseau, Sitkin, Burt, & Camerer, 1998; Slovic, 1993).

Despite the obvious differences between human–human relationships and human–technology relationships (Madhavan & Wiegmann, 2007), many studies show that people respond to technology in ways that parallel their response to people (Nass & Moon, 2000; Reeves & Nass, 1996). Typically, these responses are not based on people anthropomorphizing technology; but, instead, it seems that simple features of the technology, such as interactive complexity and voice interaction, lead people to respond to the technology as they would a person. Similar to the effect on human–human relationships, exposing people to oxytocin led people to trust and comply with an anthropomorphized aid—an avatar—but had little effect on how people responded to a computer aid without anthropomorphic features (de Visser et al., 2017). As with human–human relationships, trust in automation guides behavior where complexity, effort, and time constraints make it infeasible to completely understand the automation (Lee & See, 2004). The complexity of vehicle automation prevents people from developing a complete understanding of its capabilities which, combined with associated risk and uncertainty of the driving environment, make it very likely that trust will play an important role in how people respond to vehicle automation. Just as trust in people influences the degree to which people are willing to make themselves vulnerable to other people they are unable to monitor or control, trust in automation influences the degree people are willing to make themselves vulnerable to automated systems.

Trust becomes more relevant as technology displays greater degrees of agency. Agency concerns the degree to which technology exhibits behavior that is autonomous, responsive to other agents, and appears to be goal-directed. The perception of agency increases with the degree of autonomy—agents operating without the direct intervention of humans. The perception of agency also increases when technology is reactive, where agents perceive their environment and respond to changes. More than simply being reactive, proactive, and goal-directed, behavior is a particularly powerful indicator of agency (Wooldridge & Jennings, 1995). These characteristics all tend to induce people to consider software agents and similar sophisticated automation as an intentional agent (Lee, 2006; Meyer & Lee, 2013). Surface features of the technology can amplify the perception of automation agency. Even iconic representations of faces and eyes influence emotional response and visual attention (Driver et al., 1999; Langton & Bruce, 1999). Voices can also have a powerful influence. In one study, the emotional quality of the voice in an in-vehicle device interacted with the mood of the driver such that, when the voice matched the driver state (subdued for negative, enthused for positive), drivers drove more safely (Nass et al., 2005). In the context of vehicle technology, shared, traded, and ceded control describes vehicle automation that has increasing agency. Trust is important with highly agentic technology because the complexity of such technology makes a complete understanding difficult.

Agency can refer to a characteristic of technology, or to the sense of control people feel over their influence on the world. The agency of the automation can influence the agency that the person feels in joint activities, and this agency is profoundly important in describing the nature of the trust relationship. People's sense of agency is important because reduced agency can be accompanied by a sense of vulnerability and uncertainty that requires trust to overcome. With human–human relationships, people see the outcomes of interactions with others in terms of *I*, *We*, and *You* agency (Sebanz, Bekkering, & Knoblich, 2006). With *I* agency, people feel in control and take responsibility for outcomes. With *We* agency, people feel they are partnered with another and feel a sense of joint control and joint responsibility, as in a couple dancing. With *You* agency, control over outcomes and the associated responsibility are ascribed to the other. The sense of agency is composed of both a feeling of agency and an explicit judgment of agency, with the feeling of agency being a precondition of the judgment of agency (Haggard, 2017). The feeling of agency derives from the experience of fluent control, whereas the judgment of agency derives from the explicit interpretation of the action and outcomes relative to the goal of the action (Haggard & Tsakiris, 2009). Fluent control requires a direct connection between action and outcome, precise response to movement, and immediate feedback (Sidarus, Vuorre, Metcalfe, & Haggard, 2017). For example, people feel agency in controlling a conventional vehicle because a vehicle responds immediately and smoothly to a driver's movement of the steering wheel. Agency is greatest when people's actions clearly and immediately reflect a goal that they actively choose (Haggard, 2017).

The *I*, *We*, and *You* agencies are relevant to how people trust and respond to vehicle automation because agency influences the capacity to detect errors, accept responsibility, and feel in control (van der Wel, 2015). Through predictive processing of perceptual-motor actions, people are very sensitive to their own errors, in the case of *I* agency; or their partners' errors, in the case of *We* agency (Picard & Friston, 2014). Furthermore, people can predict the goals of others before the action ends, provided that the action conforms to the constraints of biological motion (Elsner, Falck-Ytter, & Gredebäck, 2012). Fluent interactions with automation and a sense of *We* agency depend on these predictive mechanisms that break down when biological motion rules are violated and the automation lacks anthropomorphic characteristics (Sahaï, Pacherie, Grynszpan, & Berberian, 2017). Observable and directable behavior that includes anthropomorphic control patterns will likely engage the mirror neuron system of the person and promote fluent control (Knoblich & Sebanz, 2008; Madhavan & Wiegmann, 2007). For specific examples of how to create such anthropomorphic control patterns see Lasseter (1987). Removing fluent control tends to shift people from *I* to *We* or *You* agency, resulting in diminished error detection, longer latency, and a diminished sense of control and responsibility. Ideally, the *I*, *We*, and *You* agencies should parallel the role of the person in shared, traded, and ceded control.

We agency is sometimes transformed into *I* agency when people develop a sense that they are responsible for control that is actually generated by another. This vicarious control emerges when people are deeply engaged by observation and actively predicting outcomes. Vicarious control tends to occur when instructions precede action but not when the instructions follow the action (Wegner, Sparrow, & Winerman, 2004). In the driving context, this might have implications for the timing of turn-by-turn navigation commands that are being followed by a car with traded control. Instructions that precede the response of the maneuver are likely to induce a greater degree of vicarious control and agency. The opposite of vicarious control—agency collapse—can also occur, leading people who are controlling the vehicle to move from a sense of *I* agency to *You* agency and ascribe agency and responsibility to the vehicle. This seems to occur in some cases of unintended acceleration, where people feel that the car is out of their control despite their continued depression of the accelerator pedal. Such agency collapse might become more prevalent with increasingly autonomous automation (Schmidt, 1989; Schmidt & Young, 2012).

SAE Level 2 automation poses a problem in this regard because removing the driver from the fluent, perceptual-motor, control might lead to a sense of *You* agency where people disengage from driving, fail to detect automation errors, and lose a sense that they are responsible to oversee the automation. A recent test-track study highlighted this possibility by exposing people to obstructions (e.g., a stationary vehicle and garbage bag) that require them to respond because the automation would not. Although people received instructions and had their eyes on the road, 28% failed to respond (Victor et al., 2018). The authors explained the results in terms of an expectation mismatch, which suggests a misunderstanding of system capability due to poor situation awareness and an inaccurate mental model (Engström et al., 2017). The concept of agency suggests a more fundamental mismatch where some people may have developed expectations based on *You* agency rather than the *I* or *We* agency appropriate for SAE Level 2 automation. Shifts of agency are likely driven, not so much by the explicit understanding associated with mental models of automation, but by developing trust based on experiencing the automation.

The research on agency and joint activity suggests that trust is rooted in perceptual-motor experience of acting together (Seemann, 2009). Trust that develops with *We* agency emerges as an attitude of the other that exists and is expressed in the bodily state of the other (embodied) and the intentions of the other can be read from actions of the other (enacted). Trust and *We* agency depends on a sense of joint control that depends on perceptual-motor coupling that some types of automation can enhance by providing additional feedback through the steering wheel and pedals, but that other types of automation can degrade or eliminate leading to *You* agency (Abbink et al., 2017; Mulder, Abbink, & Boer, 2012). More generally, these findings suggest assessing agency might be essential to define the nature of the trust relationship (Caspar et al., 2015; Haggard, 2017). Are people trusting the automation to drive—as in ceded control and *You* agency? Or are they trusting the automation to work with them to drive—as in shared control and *We* agency?

4.4 PROMOTING APPROPRIATE TRUST IN VEHICLE AUTOMATION

As with human-to-human relationships, higher levels of trust in human-to-technology relationships lead people to rely on technology and lower levels lead people to reject it (Lee & Moray, 1992; Muir & Moray, 1996). However, greater trust is not always better. Trusting too much leads to over-reliance and misuse and leaves people vulnerable to automation failures; trusting too little leads to disuse where people reject potentially useful technology (Parasuraman & Riley, 1997). Fostering appropriate trust poses a fundamental design challenge in creating safe and accepted vehicle automation.

4.4.1 Calibration and Resolution of Trust

Appropriate trust is the degree to which the capabilities of automation—trustworthiness—align with the attitude of the person regarding those capabilities—trust. Figure 4.4 shows the relationship between trust and automation capabilities. Well-calibrated trust lies on the diagonal where trust equals trustworthiness (Lee & See, 2004). High-resolution trust is sensitive to changes in the trustworthiness so that changes in trustworthiness lead to similar changes in trust. Poor resolution of trust would be a situation where changes in the capability of the automation do not lead to changes in trust, as shown by the relatively wide range of capability mapping onto a small range of trust.

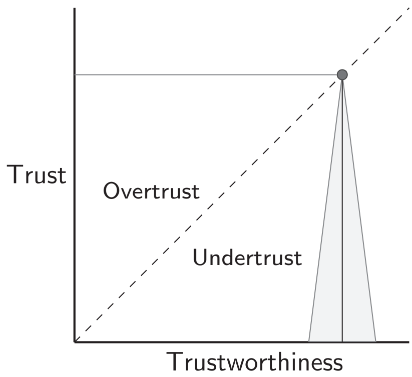

FIGURE 4.4 Calibration and resolution of trust in automation.

Much research examines how automation reliability affects trust and reliance, where reliance is the proportion of time people have the automation engaged. Reliability is often manipulated by inserting random faults in the automation that produce some overall level of reliability, such as 75% of the trials producing error-free performance (Hancock et al., 2011; Schaefer et al., 2016). This simple indication of performance often ignores the important effect of context (Bagheri & Jamieson, 2004). The context or operational domain refers to features of the situation that might influence the automation performance. To the extent that automation reliability depends on the context, the resolution of trust can be improved by highlighting situations where the automation performs well and situations where it does not. Improving the resolution of trust is particularly critical for SAE Level 2 and 3 automation. Such automation might work perfectly within the Operational Design Domain (ODD), which might cover 75% of the driving situations, and not elsewhere. This is a very different situation than the automation operating properly 75% in all driving situations. Designing an ODD that is easy for drivers to understand can greatly enhance the resolution of trust so that drivers trust and rely on the automation when it is within the ODD, and but not when it is not.

Ideally, trust would be highly resolved and well-calibrated so that high levels of trust correspond to high capability to achieve the driver's goal, and low levels of trust correspond to situations where the technology is generally incapable of achieving the driver's goal. Figure 4.4 points to two general approaches in achieving appropriate trust: improve the capability of the automation—improve trustworthiness—to meet people's expectations or make the technology more trustable. Increasing trustworthiness to match people's trust with automation that requires drivers to remain able to take back control—SAE Level 2 or 3—can fail because the better performing automation might engender greater and equally inappropriate trust, leading drivers to be even less likely to take control when needed. This automation conundrum is well documented in other domains (Endsley, 2016; Sebok & Wickens, 2017; see also, this Handbook, Chapter 21). A more promising approach is to enhance the calibration and resolution of trust by making automation more trustable, which is the focus of the balance of this section.

A comprehensive literature review suggested that trustworthiness can be described in terms of the purpose, process, and performance characteristics of the automation and that trust can be calibrated by communicating this information to people (Lee & See, 2004). *Purpose* refers to the goals and intended application of the automation—why the automation was developed. *Process* refers to the mechanisms and algorithms the automation uses to achieve this purpose—how the automation works. *Performance* refers to how consistently and precisely the automation achieves its goals (Lee & See, 2004). These characteristics describe different types of information that guide the development of trust. These dimensions identify information requirements for the HMI of AVs that can promote more appropriate trust. However, these characteristics are not invariant but change with the situation.

4.4.2 Trustable Automation: Create Simple Use Situations and Automation Structure

The *use situation* or operational domain can strongly influence the behavior of the automation and defines where and when the automation can be expected to achieve the users' goals. The situation is particularly critical for vehicle automation because a situation may be ill-suited for the automation and likely require a person to intervene—a curving road with poor lane markings—but the moment-to-moment performance of the automation can induce over-trust. For example, the affective processes that guide trust development make it likely that the concrete, immediate feedback, such as the precise lane-keeping of a lane-keeping aid, will outweigh the abstract and general information about how the automation works and the need for the driver to monitor the automation continuously. Furthermore, calibrating trust in the context of precise lane-keeping depends on drivers relating the capabilities of the automation to the situation by extrapolating past experiences, evaluating ongoing experience, and projecting future outcomes (Emirbayer & Mische, 1998); and, in this situation, the salience of the ongoing experience tends to dominate, leading to poorly calibrated trust.

The complexity of the roadway environment and the evolving capacity of vehicle automation may make it particularly difficult for drivers to calibrate their trust. Although drivers might see driving as a routine activity the nearly infinite combination of situations makes it extremely complex for automation. As a consequence, vehicle automation will need to evolve and "learn to drive" over many years. Vehicle automation might update on a weekly or monthly basis, introducing new capabilities and vulnerabilities (Bansal et al., 2019). These changes and the highly variable roadway situations create a *wicked* learning environment that makes it difficult to develop expertise (Hogarth, Lejarraga, & Soyer, 2015). In contrast *kind* environments have stable relationships where people can learn to recognize situations and the appropriate response (Kahneman & Klein, 2009).

Automation can be crafted to be kinder and trustable by considering the *structure* of the automation. One approach is to introduce new features and upgrades in a manner that corresponds to the task chunking that defines driving activities and that logically extends the ODD. For example, an update that extends a lane-keeping system so that it can automatically change lanes should consider that drivers are likely to chunk such a task to include checking the blind spot and so would expect the automation to include that subtask with the update.

4.4.3 Trustable Automation: Display Surface and Depth Indications of Capability

An individual can infer the three characteristics of automation—purpose, process, and performance—by observing the automation over time. These characteristics describe the automation in terms of levels of attributional abstraction that help guide trust towards a level appropriate for the situation (Lee & See, 2004). However, people do not experience these characteristics directly, but only through displays and controls that provide an imperfect indication of these underlying characteristics. Specific guidelines for conveying purpose, process, and performance include (Lee & Seppelt, 2012; Seppelt & Lee, 2019; Weast et al., 2016):

Purpose

- Define the operational domain in concrete terms and provide a continuous indicator of whether the vehicle is in the operational domain or not.
- Indicate impending changes in the driving situation relative to the operational domain.
- Monitor the driver's behavior and provide feedback when the driver's behavior, such as attention to the road, deviates from that required by the purpose of the automation.
- Match the goals of the automation to those of the person (e.g., speed or energy efficiency).

Process

- Indicate what the automation "sees" and does not "see."
- Communicate the state of the vehicle and explain its behavior. The vehicle should indicate what events led the vehicle to suddenly brake or change lanes.
- Allow people to request more information to explain behavior and to reduce the flow of information.
- Automation that mimics the process of human driving will produce fewer surprises and be more trustable.

Performance

- Show proximity to control capacity to indicate when performance is likely to suffer, such as in steering through a curve.
- Consider haptic and auditory cues to highlight situations where performance is degraded.

The purpose, process, and performance of automation can be thought of as depth features that define the capability of the automation, which are imperfectly revealed by the surface features of the interface. Sometimes the surface features of the interface themselves can strongly influence trust even though they have no direct connection to the underlying capability of automation. For example, the scent of lavender leads to greater interpersonal trust by inducing a calm, inclusive state (Sellaro, van Dijk, Paccani, Hommel, & Colzato, 2014), and pastel colors enhance trust in cyberbanking interfaces (Kim & Moon, 1998).

Sometimes the surface features of the automation conflict with the depth features that define its true capabilities. One example of this is the label used for various vehicle automation features. These labels can imply that the automation has greater capabilities than it has, but often the labels are ambiguous and confusing (Abraham, Seppelt, Mehler, & Reimer, 2017; Nees, 2018). Although the labeling and description of automation in the owner's manual is an important basis of trust, real-time feedback from the HMI in the vehicle might be more influential (Li, Holthausen, Stuck, & Walker, 2019; this Handbook, Chapter 15). Both visual and auditory displays of performance and process of the automation act as an externalized mental model that helps drivers see its limits (Seppelt & Lee, 2007; 2019).

Moving beyond the HMI, the behavior of automation can make the automation more trustable and promote appropriate trust. The variability of the lane position of the vehicle can be a salient cue that might convey the depth features of automation. Because degraded control behavior of the vehicle engages drivers (Alsaid, Lee, & Price, 2019), less precise lane-keeping can prompt drivers to look to the road (Price, Lee, Dinparastdjadid, Toyoda, & Domeyer, 2019). Precise lane-keeping in situations where the automation is less capable could promote over-trust. Similarly, acceleration cues of ACC helped redirect driver's attention to the road and anticipate conflicts with the vehicle ahead (Morando, Victor, & Dozza, 2016). Similarly, vehicles can announce their intent to change lanes through the roll of the vehicle's body (Cramer & Klohr, 2019). In general, the salient surface features, such as vehicle motion, should be mapped to important depth features of the automation to convey how its capability changes with the evolving situation.

The trust-relevant characteristics of the automation's purpose, process, and performance are often inferred from direct experience with the system, but this might not be possible with the introduction of dramatically new technology, such as self-driving cars. Many people will not have any direct experience on which to base their initial trust of these systems. In the absence of direct experience, people base their trust on relational and societal bases (Lee & Kolodge, 2019). Relational bases include experience with a brand (e.g., GM or Ford) or a type of technology (e.g., computers). Societal bases include the policy and regulatory structures that ensure vehicle safety. Like labels that can lead to inappropriate trust, so can the relational and societal bases of trust (also see this Handbook, Chapter 5).

4.4.4 Trustable Automation: Enable Directable Automation and Trust Repair

Revealing the depth features of automation so that they are apparent to the person is sometimes referred to as transparent or observable automation (Endsley, 2016; Klein et al., 2004). Often neglected in the quest to release people of the burden of control is the need to make automation directable by giving people the ability to adjust and guide the automation. As an example, automation that supports lane-keeping might effectively center the vehicle but might make it difficult for the driver to adjust the vehicle's position to accommodate other traffic or potholes. An alternate approach could be for the automation to be more directable and accept steering input from the driver and provide feedback through the steering wheel that indicates when the driver approaches safety boundaries, such as the edge of the road (Abbink et al., 2017; Abbink, Mulder, & Boer, 2012). Such feedback could also indicate to the driver when sensors are providing less precise information. As in the previous discussion of agency, control changes the way we perceive the world and letting the driver direct the automation provides support for appropriate trust that is not easily achieved by simply displaying trust-related information (Wen & Haggard, 2018). More specifically, like active touch and fluent perceptual-motor interaction, perception through direction of automation will likely be more effective in helping drivers detect limitations than perception through observation of automation (Flach, Bennett, Woods, & Jagacinski, 2015; Gibson, 1962). Directable automation may provide a strategy to circumvent one of the ironies of automation: better automation leads to less frequent engagements and the occasional engagements tend to be more challenging (Bainbridge, 1983). Automation that keeps the driver engaged and invites interaction may prepare people for occasional instances where they must take control.

Making the automation directable engages the person in developing trust; trust repair engages the automation in re-developing trust. The process of recovering peoples' trust following an automation failure can be an active process of trust repair (de Visser, Pak, & Shaw, 2018; Kim & Cooper, 2009; Tomlinson & Mayer, 2009). Trust repair describes the process of interacting with the person to recover trust rather than simply relying on the baseline representation of the automation's purpose, process, and performance. Here, additional information is presented to demonstrate that the system is trustworthy. One way to frame this information is in terms of extrapolating past experiences, evaluating the present experience, and projecting future outcomes—explain, show, and promise (Emirbayer & Mische, 1998). This can include an explanation of past failings. It can also include showing information in the present that shows why the automation fails, and it could project to the future with a promise for why the automation will not fail (Kohn, Quinn, Pak, de Visser, & Shaw, 2018). For example, if the automation requests the driver to take back control, it might build trust by describing what about the situation—unusually heavy rain—led to the need for the driver to intervene and why this would be unlikely to occur in the future. More generally, the elements of trust recovery—explain, show, and promise—can convey the locus of causality (e.g., an external event, such as heavy rain versus an internal event, such as a software bug), degree of control over the situation (e.g., the degree to which the rain was unavoidable), and stability of the cause (e.g., the unpredictable nature of intense rain) (Tomlinson & Mayer, 2009).

The aim of trust repair, like the other ways of engineering the relationship with automation, is to promote highly resolved and well-calibrated trust, not to increase trust. Following this logic, thought should be given to the complement of trust repair—trust tempering. Trust tempering would actively monitor situations where the system performed well, but failure was likely, and then explain why such success cannot be counted on in the future can moderate trust and help people avoid relying on automation in such situations in the future.

4.4.5 Trustworthy Automation and Goal Alignment

Technically adept vehicles that keep people safe is a necessary, but not a sufficient, condition to ensure people trust and accept automation, particularly when people cede control. Even with ACC, allowing drivers to choose the driving style that aligns with their goals (e.g., fast or economical) can enhance trust (Verberne, Ham, & Midden, 2012). Contrary to most research on trust in automation, the automation in the transportation ecology shown in Figure 4.1 does not always work to achieve the goals of each individual. One example is how connected vehicles might be routed to minimize congestion. The routing might produce shorter trips on average but might delay some drivers. In this case, the goal might be to reduce the trip duration for all vehicles, but this does not necessarily align with the goals of the individual rider. This is particularly true of those who are not directly benefiting from vehicle automation, such as pedestrians and other road users who must negotiate AVs for right-of-way at an intersection (Domeyer, Dinparastdjadid, Lee, & Douglas, 2019). Such incidental users, who are affected by automation, but not served by the automation, might reject automation simply because its goals fail to align with their goals (Inbar & Tractinsky, 2009). In such situations trust and social norms can help avoid the tendency of people to pursue the rational immediate benefit for themselves at the cost of collective benefit—sometimes termed the tragedy of the commons (Hayashi, Ostrom, Walker, & Yamagishi, 1999; Ostrom, 1998) (also see Chapter 19 for a discussion of joint optimization in the transportation system). In these situations, face-to-face interactions can enhance cooperation and mitigate the tragedy of the commons. Unfortunately, increasingly autonomous vehicles might reduce such face-to-face interactions and undermine the tolerance of other road users for autonomous vehicles (Domeyer, Lee, Toyoda, 2020). More generally, the design of the network structure and the associated spatial layout can predict cooperative behavior: small neighborhoods promote more cooperation. Specifically, the benefit/cost ratio for cooperative behavior must exceed 1+ (group size/number of groups) for cooperation to flourish (Miller, 2013).

4.5 TRUST AND ACCEPTANCE OF VEHICLE TECHNOLOGY

The Technology Acceptance Model (TAM) has been used in other domains to determine how features of technology influence acceptance and use of that technology. According to that model, technology acceptance often depends on two perceptions of technology: perceived usefulness and perceived ease of use (Davis, 1989). Perceived usefulness describes how well the functionality meets the person's needs and perceived ease of use describes how accessible that functionality is. Typically, perceived usefulness outweighs perceived ease of use in predicting acceptance, and this might be particularly true in the driving domain. In driving, perceived usefulness might be particularly influential because, an AV could turn the 52 minutes of the average daily commute into time the rider could spend working or relaxing (Hu & Young, 1999). People might consider the ability to work while commuting to be very useful and that utility might dominate trust in predicting acceptance of vehicle technology (Ghazizadeh, Lee, & Boyle, 2012). One study that addressed this issue found trust and perceived usefulness influenced the intention to use an autonomous vehicle to a similar extent and that perceived ease of use had a much smaller influence (Choi & Ji, 2015).

Unlike the typical application of TAM to information technology in the workplace, driving involves considerable risk. Although autonomous vehicles promise to greatly reduce the risk of driving, it is not the actual risk and safety that govern trust and technology acceptance but the perceived risk. With perceived risk, people focus on stories and feelings, not statistics and analysis, and so they might neglect the many crashes autonomous vehicles avoided and focus on the few caused by automation (Slovic, Finucane, Peters, & MacGregor, 2004). Consequently, people may perceive autonomous vehicles as much riskier than they actually are. Perceived risk can also deviate from actual risk in situations where the technology is not readily observable, mishaps produce deadly consequences, and people feel they have no control. Such situations produce dread risk (Slovic, 1987; Slovic et al., 2004). An analysis of open-ended items from a large survey found people responding to vehicle automation in terms of dread risk (Lee & Kolodge, 2019). If vehicle automation produces feelings of dread risk, then AVs might need to be 1,000 times safer than conventional vehicles to be perceived as having the same risk (Lee, 2019; Slovic, 1987).

Trust plays a critical role in mediating risk perception (Slovic, 1993). In the context of AVs, trust partially mediated how potential environmental benefits influence risk acceptance, and fully mediated the effect of these benefits on people's intention to use AVs (Liu, Ma, & Zuo, 2019). In other words, people are willing to accept more risk for riding in an environmentally friendly vehicle if they trust automation. Generally, trust is slow to develop and quick to lose (Muir, 1987; Slovic, 1993). This may be even more pronounced with AVs where a failure might lead people to see risk in terms of dread risk. Such a transition might lead to trust collapse and a slow recovery as people monitor the behavior of automation with the expectation of further violations of their trust (Earle, Siegrist, & Gutscher, 2010; Slovic, 1993).

Although TAM suggests that the perceived usefulness of AVs will be a powerful force in their acceptance, the nature of driving makes it likely that risk will likely play an important role. This role might be accentuated by the potential for people to perceive risk associated with vehicle automation as dread risk. Importantly, trust seems to have a strong influence on both perceived usefulness and risk, making it critical to craft trustworthy and trustable automation.

4.6 ETHICAL CONSIDERATIONS AND THE TELEOLOGY OF TECHNOLOGY

The focus of this chapter has been on trust and development of appropriate trust—creating trustable technology. With increasingly autonomous technology, people will ride in vehicles or share the road with the vehicles while having little control over the technology. As such, we need to shift to creating trustworthy automation. Trustworthy automation is automation that should be trusted because it reliably achieves people's goals. This requires a focus not just on ensuring technology can maintain the position of the vehicle in the lane and avoid obstacles, but that it behaves politely with other road users and how it provides service and equitably uses public resources. Vehicle automation can provide mobility and greater agency to those who are poorly served by today's transportation options, such as those that are older, economically disadvantaged, or have vision and mobility limitations (see also, this Handbook, Chapters 10, 17). However, recent history has shown that the algorithms that power social media and public policy tend to exacerbate existing biases and inequities (O'Neil, 2016). To counter this tendency, there is an increasing need for policies to ensure the trustworthiness and ethical implementation of technology (Etzioni & Etzioni, 2016).

Some frame the ethical challenge of AVs in terms of the Trolley Problem, where the algorithm designers face decisions about which people an algorithm causes to die and which it saves in particular crash situations (Jean-Francois, Azim, & Iyad, 2016; Shariff, Bonnefon, & Rahwan, 2017). Such situations and the ethical dilemmas posed are rarely realistic challenges for designers (Bauman, McGraw, Bartels, & Warren, 2014; De Freitas, Anthony, & Alvarez, 2019). A more realistic dilemma is the macro-level "Trolley Problem" that companies and regulatory agencies face: should they pull the lever and enable self-driving cars on the road knowing that the technology is imperfect and their action will be responsible for thousands of people dying or should they wait and let thousands more die in conventional vehicles as is happening today. At a more pragmatic level, developers and policymakers must decide whether it is ethical for fully automated, self-driving vehicles to follow the letter of law. Following the speed limit rather than the speed of the traffic stream might increase the risk of collision; however, over time such behavior might lead others to drive more slowly and follow the speed limit, which might ultimately make driving safer. Chapter 14 describes some of the challenges faced by governmental responses to AVs.

As technology becomes more capable, we confront the threat that increasingly autonomous technology will undermine human agency (Hancock, Nourbakhsh, & Stewart, 2019; Lee, 2019). At the same time, autonomous systems can free us from mundane tasks, such as navigating a congested highway and give us agency to pursue more meaningful activities. The diverse conceptualization of trust presented in this chapter may provide the theoretical perspective needed to identify what level of control people need to achieve their goals and feel enhanced agency rather than diminished agency.

4.7 CONCLUSION

Trust complements the concepts of mental models and situation awareness in explaining how and when people rely on automation. Trust is a multi-faceted term that operates at timescales of seconds to years to describe whether people rely on, accept, and tolerate vehicle technology. Trust mediates micro interactions concerning how people rely on automation to engage in non-driving tasks to macro interactions concerning how the public accepts new forms of transport. Public acceptance may depend on the trust of incidental users, such as pedestrians who must negotiate with AVs at intersections, and drivers who must share the road with AVs. With such incidental users, designers need to consider how to reconcile the goals of the riders of the AVs and the goals of the other road users who must tolerate the AVs if they are to succeed (Domeyer, Lee, & Toyoda, 2020). Unlike traditional users for which the automation was primarily designed, the trust and tolerance of incidental users might play a critical role in the success of increasingly AVs as they use public roadways and resources in ways that conventional vehicles might not. Chapter 5 addresses these issues of public acceptance in more detail.

Harmonizing trust relationships in the transportation network represents a complex and multi-faceted design challenge (this Handbook, Chapter 19). The specific design considerations depend on the particular trust relationships (e.g., driver interaction SAE Level 2 automation or pedestrians interacting with driverless vehicles); however, some specific advice to promote appropriate trust can be offered:

Create Simple Use Situations and Automation Structure
- Create ODDs that are clear, concrete, and simple.
- Minimize the complexity of automation functions and match them to the natural divisions of driving activities.
- Consider human-like driving behavior to make vehicle behavior more predictable.

Display Surface and Depth Indications of Capability
- Create displays that reveal the purpose, process, and performance of automation.
- Show what the automation sees and how it plans to respond.
- Rather than mask control challenges, reveal them through the vehicle control behavior.

Enable Directable Automation and Trust Repair
- Even “fully automated” vehicles should provide people with appropriate control and feedback.
- Provide explanations when mishaps occur.
- Enable people to ask “why” the automation behaved as it did.

While each of these techniques promote appropriate trust, a critical consideration is the context of the transportation ecology; what promotes trust for some might have unintended consequences of undermining trust of others. A network perspective is essential in crafting a trustworthy and cooperative AV system (Miller, 2013).

ACKNOWLEDGEMENTS

This work is partially supported by a grant from Toyota CSRC and the National Science Foundation (NSF 18–548 FW-HTF). The chapter was greatly improved with the comments of the editors, Erin Chiou, Josh Domeyer, and Mengyao Li, and the other members of the Cognitive Systems Laboratory at the University of Wisconsin—Madison.

REFERENCES

Abbink, D. A., Carlson, T., Mulder, M., de Winter, J., Aminravan, F., Gibo, T., & Boer, E. (2017). A topology of shared control systems–Finding common ground in diversity. *IEEE Transactions on Human-Machine Systems*, 48(5), 509–525.

Abbink, D. A., Mulder, M., & Boer, E. R. (2012). Haptic shared control: Smoothly shifting control authority? *Cognition, Technology and Work*, *14*(1), 19–28. doi:10.1007/s10111-011-0192–5

Abraham, H., Seppelt, B., Mehler, B., & Reimer, B. (2017). What's in a name: Vehicle technology branding and consumer expectations for automation. *AutomotiveUI 2017–9th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, Proceedings*, 226–234. doi:10.1145/3122986.3123018

Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature*, *393*, 470–474.

Alsaid, A., Lee, J. D., & Price, M. A. (2019). Moving into the loop: An investigation of drivers' steering behavior in highly automated vehicles. *Human Factors*. doi:10.1177/0018720819850283

Bagheri, N. & Jamieson, G. A. (2004). The impact of context-related reliability on automation failure detection and scanning behaviour. *2004 IEEE International Conference on Systems, Man and Cybernetics*, *1*, 212–217.

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), 775–779. doi:10.1016/0005–1098(83)90046-8

Bansal, G., Nushi, B., Kamar, E., Weld, D., Lasecki, W., & Horvitz, E. (2019). A case for backward compatibility for Human-AI teams. *arXiv:1906.01148*.

Bauman, C. W., McGraw, P. A., Bartels, D. M., & Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass*, *8*(9), 536–554. doi:10.1111/spc3.12131

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*(5304), 1293–1295.

Caspar, E. A., De Beir, A., Magalhaes De Saldanha Da Gama, P. A., Yernaux, F., Cleeremans, A., & Vanderborght, B. (2015). New frontiers in the rubber hand experiment: When a robotic hand becomes one's own. *Behavior Research Methods*, *47*(3), 744–755. doi:10.3758/s13428-014-0498–3

Choi, J. K. & Ji, Y. G. (2015). Investigating the importance of trust on adopting an autonomous vehicle. *International Journal of Human-Computer Interaction*, *31*(10), 692–702. doi:10.1080/10447318.2015.1070549

Cramer, S. & Klohr, J. (2019). Announcing automated lane changes: Active vehicle roll motions as feedback for the driver. *International Journal of Human-Computer Interaction*, *35*(11), 980–995. doi:10.1080/10447318.2018.1561790

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340.

De Freitas, J., Anthony, S. E., & Alvarez, G. A. (2019). Doubting driverless dilemmas. *PsychArXiv*, 1–5. doi:10.7498/aps.62.064705

De Kosnik, A. (2014). Disrupting technological privilege: The 2013–2014 San Francisco Google bus protests. *Performance Research*, *19*(6), 99–107. doi:10.1080/13528165.2014.985117

de Visser, E. J., Monfort, S. S., Goodyear, K., Lu, L., O'Hara, M., Lee, M. R., … Krueger, F. (2017). A little anthropomorphism goes a long way. *Human Factors*, *59*(1), 116–133. doi:10.1177/0018720816687205

de Visser, E. J., Pak, R., & Shaw, T. H. (2018). From 'automation' to 'autonomy': The importance of trust repair in human–machine interaction. *Ergonomics*, *0139*, 1–19. doi:10.1080/00140139.2018.1457725

Domeyer, J., Dinparastdjadid, A., Lee, J. D., & Douglas, G. (2019). Proxemics and kinesics in automated vehicle-pedestrian communication: Representing ethnographic observations. *Transportation Research Record*. doi:10.1177/0361198119848413

Domeyer, J., Lee, J. D., & Toyoda, H. (2020). Vehicle automation-other road user communication and coordination: Theory and mechanisms. *IEEE Access,* 8, 19860–19872.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition*, *6*(5), 509–540.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., Dawe, L. A., & Anderson, B. W. (2001). Predicting misuse and disuse of combat identification systems. *Military Psychology*, *13*(3), 147–164.

Earle, T. C., Siegrist, M., & Gutscher, H. (2010). Trust, risk perception and the TCC model of cooperation. In M. Siegrist, T. C. Earle, & H. Gutscher (Eds.), *Trust in Risk Management: Uncertainty and Scepticism in the Public Mind* (pp. 1–50). London: Earchscan.

Elsner, C., Falck-Ytter, T., & Gredebäck, G. (2012). Humans anticipate the goal of other people's point-light actions. *Frontiers in Psychology, 3*, 120. doi: 10.3389/fpsyg.2012.00120

Emirbayer, M. & Mische, A. (1998). What is agency? *American Journal of Sociology*, *103*(4), 962–1023.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*(1), 32–64. doi:10.1518/001872095779049543

Endsley, M. R. (2013). Situation awareness. In J. D. Lee & A. Kirlik (Eds.), The Oxford Handbook of Cognitive Engineering (pp. 88–108). New York: Oxford University Press.

Endsley, M. R. (2016). From here to autonomy. *Human Factors*, *59*(1), 5–27. doi:10.1177/0018720816681350

Engström, J., Bärgman, J., Nilsson, D., Seppelt, B., Markkula, G., Piccinini, G. B., & Victor, T. (2017). Great expectations: A predictive processing account of automobile driving. *Theoretical Issues in Ergonomics Science*, *19*(2), 154–194. doi:10.1080/1463922X.2017.1306148

Ert, E., Fleischer, A., & Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism Management*, *55*, 62–73. doi:10.1016/j.tourman.2016.01.013

Etzioni, A. & Etzioni, O. (2016). AI assisted ethics. *Ethics and Information Technology*, *18*(2), 149–156. doi:10.1007/s10676-016-9400–6

Fehr, E. & Camerer, C. F. (2007). Social neuroeconomics: The neural circuitry of social preferences. *Trends in Cognitive Sciences*, *11*(10), 419–427. doi:10.1016/j.tics.2007.09.002

Flach, J. M., Bennett, K. B., Woods, D. D., & Jagacinski, R. J. (2015). Interface design: A control theoretic context for a triadic meaning processing approach. *The Cambridge Handbook of Applied Perception Research*, 647–668. doi:10.1017/CBO9780511973017.040

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly*, *27*(1), 51–90.

Ghazizadeh, M., Lee, J. D., & Boyle, L. N. (2012). Extending the technology acceptance model to assess automation. *Cognition, Technology & Work*, *14*(1), 39–49. doi:10.1007/s10111-011-0194-3

Gibson, J. J. (1962). Observations on active touch. *Psychological Review*, *69*(6), 477–491.

Guznov, S., Lyons, J., Nelson, A., & Woolley, M. (2016). The effects of automation error types on operators' trust and reliance. In S. Lackey & R. Shumaker (Eds.), *Lecture Notes in Computer Science* (Vol. 9740, pp. 116–124). Berlin: Springer. doi:10.1007/978-3-319–39907-2_11

Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience*, *18*(4), 197–208. doi:10.1038/nrn.2017.14

Haggard, P. & Tsakiris, M. (2009). The experience of agency: The experience of agency feelings, judgments, and responsibility. *Current Directions in Psychological Science*, *18*(4), 242–246.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, *53*(5), 517–527. doi:10.1177/0018720811417254

Hancock, P. A., Nourbakhsh, I., & Stewart, J. (2019). On the future of transportation in an era of automated and autonomous vehicles. *Proceedings of the National Academy of Sciences*, *116*(16), 7684–7691. doi:10.1073/pnas.1805770115

Hayashi, N., Ostrom, E., Walker, J., & Yamagishi, T. (1999). Reciprocity, trust, and the sense of control: A cross-societal study. *Rationality and Society*, *11*(1), 27–46.

Hoff, K. A. & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434. doi:10.1177/0018720814547570

Hogarth, R. M., Lejarraga, T., & Soyer, E. (2015). The two settings of kind and wicked learning environments. *Current Directions in Psychological Science*, *24*(5), 379–385. doi:10.1177/0963721415591878

Hu, P. & Young, J. (1999). *Summary of Travel Trends: 1995 Nationwide Personal Transportation Survey*. Washington, DC: Federal Highway Administration.

Inbar, O. & Tractinsky, N. (2009). The incidental user. *Interactions, 16*(4), 56–59.

Jean-Francois, B., Azim, S., & Iyad, R. (2016). The Social dilemma of autonomous vehicles. *Science*, *352*(6293), 1573.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Macmillan.

Kahneman, D. & Klein, G. A. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526. doi:10.1037/a0016755

Kim, J. & Moon, J. Y. (1998). Designing towards emotional usability in customer interfaces—Trustworthiness of cyber-banking system interfaces. *Interacting with Computers*, *10*(1), 1–29.

Kim, P. H. & Cooper, C. D. (2009). The repair of trust: A dynamic bilateral perspective and multilevel conceptualization. *Academy of Management Review*, *34*(3), 401–422.

Klein, G. A., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., Feltovich, P. J., Hoffman, R. R., … Ford, K. M. (2004). Ten challenges for making automation a "Team Player" in joint human-agent activity. *IEEE Intelligent Systems*, *19*(6), 91–95.

Knoblich, G. & Sebanz, N. (2008). Evolving intentions for social interaction: From entrainment to joint action. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *363*(1499), 2021–2031. doi:10.1098/rstb.2008.0006

Kohn, S. C., Quinn, D., Pak, R., de Visser, E. J., & Shaw, T. H. (2018). Trust repair strategies with self-driving vehicles: An exploratory study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 1108–1112. doi:10.1177/1541931218621254

Kosfeld, M., Heinrichs, M., Zak, P. J., Fishbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature*, *435*, 673–676.

Langton, S. R. H. & Bruce, V. (1999). Reflexive visual orienting in response to social attention of others. *Visual Cognition*, *6*(5), 541–567.

Lasseter, J. (1987). Principles of traditional animation applied to 3D computer animation. *ACM SIGGRAPH Computer Graphics*, *21*(4), 35–44. doi:10.1145/37402.37407

Lee, J. D. (2006). Affect, attention, and automation. In A. Kramer, D. Wiegmann, & A. Kirlik (Eds.), *Attention: From Theory to Practice* (pp. 73–89). New York: Oxford University Press.

Lee, J. D. (2018). Perspectives on automotive automation and autonomy. *Journal of Cognitive Engineering and Decision Making*, *12*(1), 53–57. doi:10.1177/1555343417726476

Lee, J. D. (2019). Trust and the teleology of technology. *Ergonomics*, *62*(4), 500–501. doi:10.1080/00140139.2019.1563332

Lee, J. D. & Kolodge, K. (2019). Exploring trust in self-driving vehicles with text analysis. *Human Factors*. doi:10.1177/0018720819872672

Lee, J. D. & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10), 1243–1270. doi:10.1080/00140139208967392

Lee, J. D. & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80.

Lee, J. D. & Seppelt, B. D. (2012). Human factors and ergonomics in automation design. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 1615–1642). Hoboken, NJ: Wiley. doi:10.1002/9781118131350.ch59

Li, M., Holthausen, B. E., Stuck, R. E., & Walker, B. N. (2019). No risk no trust: Investigating perceived risk in highly automated driving. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 177–185). New York: ACM.

Liu, P., Ma, Y., & Zuo, Y. (2019). Self-driving vehicles: Are people willing to trade risks for environmental benefits? *Transportation Research Part A: Policy and Practice*, *125*(March), 139–149. doi:10.1016/j.tra.2019.05.014

Madhavan, P. & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. *Theoretical Issues in Ergonomics Science*, *8*(4), 277–301. doi:10.1080/14639220500337708

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709–734.

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, *13*(3), 334–359.

Meyer, J. & Lee, J. D. (2013). Trust, reliance, and compliance. In A. Kirlik & J. D. Lee (Eds.), *The Oxford Handbook of Cognitive Engineering* (pp. 109–124). New York: Oxford University Press.

Mikolajczak, M., Gross, J. J., Lane, A., Corneille, O., de Timary, P., & Luminet, O. (2010). Oxytocin makes people trusting, not gullible. *Psychological Science*, 21(8), 1072–1074. doi:10.1177/0956797610377343

Miller, H. J. (2013). Beyond sharing: Cultivating cooperative transportation systems through geographic information science. *Journal of Transport Geography*, *31*, 296–308.

Montague, E., Xu, J., & Chiou, E. (2014). Shared experiences of technology and trust: An experimental study of physiological compliance between active and passive users in technology-mediated collaborative encounters. *IEEE Transactions on Human-Machine Systems*, *44*(5), 614–624. doi:10.1109/THMS.2014.2325859

Morando, A., Victor, T., & Dozza, M. (2016). Drivers anticipate lead-vehicle conflicts during automated longitudinal control: Sensory cues capture driver attention and promote appropriate and timely responses. *Accident Analysis & Prevention*, *97*, 206–219. doi:10.1016/j.aap.2016.08.025

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, *27*, 527–539.

Muir, B. M. & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, *39*(3), 429–460.

Mulder, M., Abbink, D. A., & Boer, E. R. (2012). Sharing control with haptics: Seamless driver support from manual to automatic control. *Human Factors*, *54*(2), 786–798.

Nass, C., Jonsson, I. M., Harris, H., Reaves, B., Endo, J., Brave, S., & Takayama, L. (2005). Improving automotive safety by pairing driver emotion and care voice emotion. *Conference on Human Factors in Computing Systems* (pp. 1973–1976). New York: ACM.

Nass, C. & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103. doi:10.1111/0022–4537.00153

Nave, G., Camerer, C., & McCullough, M. (2015). Does oxytocin increase trust in humans? A critical review of research. *Perspectives on Psychological Science*, *10*(6), 772–789. doi:10.1177/1745691615600138

Nees, M. A. (2018). Drivers' perceptions of functionality implied by terms used to describe automation in vehicles. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 1893–1897. doi:10.1177/1541931218621430

Neisser, U. (1976). *Cognition and Reality: Principles and Implications of Cognitive Psychology*. San Francisco, CA: W. H. Freeman and Company.

Norman, D. A. (1983). Some observations on mental models. In D. Gentner & A. L. Stevens (Eds.), *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum.

Norman, D. A. (2004). *Emotional Design*. New York: Basic Books.

NTSB. (2016). *Preliminary Report: HWY16FH018*. Washington, DC: National Transportation Safety Bureau.

O'Neil, K. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Ostrom, E. (1998). A behavioral approach to the rational choice theory of collective action: Presidential address, American Political Science Association, 1997. *American Political Science Review*, 92(1), 1–22.

Parasuraman, R. & Riley, V. A. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*(2), 230–253.

Payyanadan, R. P. & Lee, J. D. (2018). Understanding the ridesharing needs of older adults. *Travel Behaviour and Society*, *13*(June), 155–164. doi:10.1016/j.tbs.2018.08.002

Picard, F. & Friston, K. (2014). Predictions, perception, and a sense of self. *Neurology*, *83*(12), 1112–1118. doi:10.1212/WNL.0000000000000798

Price, M., Lee, J. D., Dinparastdjadid, A., Toyoda, H., & Domeyer, J. (2019). Effect of automation and vehicle control algorithms on eye behavior in highly automated vehicles. *International Journal of Automotive Engineering*, *10*(1), 73–79.

Reeves, B. & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places*. New York: Cambridge University Press.

Romero, S. (2018). *Wielding Rocks and Knives, Arizonans Attack Self-Driving Cars*. Retrieved from www.nytimes.com/2018/12/31/us/waymo-self-driving-cars-arizona-attacks.html

Rotter, J. B. (1980). Interpersonal trust, trustworthiness, and gullibility. *American Psychologist*, *35*(1), 1–7.

Rouse, W. B. & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin and Review*, *100*(3), 349–363.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). No so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393–404.

SAE. (2016). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (J3016). Global Ground Vehicle Standards*. Warrendale, PA: Society of Automotive Engineers. doi:10.4271/J3016_201609

Sahaï, A., Pacherie, E., Grynszpan, O., & Berberian, B. (2017). Co-representation of human-generated actions vs. machine-generated actions: Impact on our sense of we-agency? IEEE International Symposium on Robot and Human Interactive Communication, 341–345. Lisbon, Portugal: IEEE.

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors*, *58*(3), 377–400. doi:10.1177/0018720816634228

Schmidt, R. A. (1989). Unintended acceleration: A review of human factors contributions. *Human Factors*, *31*(3), 345–364.

Schmidt, R. A. & Young, D. E. (2012). Cars gone wild: The major contributor to unintended acceleration in automobiles is pedal error. *Frontiers in Psychology*, *1*, 1–4.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76.
Sebok, A. & Wickens, C. D. (2017). Implementing lumberjacks and black swans into model-based tools to support human-automation interaction. *Human Factors*, *59*(2), 189–203. doi:10.1177/0018720816665201
Seemann, A. (2009). Joint agency: Intersubjectivity, sense of control, and the feeling of trust. *Inquiry*, *52*(5), 500–515. doi:10.1080/00201740903302634
Sellaro, R., van Dijk, W. W., Paccani, C. R., Hommel, B., & Colzato, L. S. (2014). A question of scent: Lavender aroma promotes interpersonal trust. *Frontiers in Psychology*, *5*(OCT), 1–5. doi:10.3389/fpsyg.2014.01486
Seppelt, B. D. & Lee, J. D. (2007). Making adaptive cruise control (ACC) limits visible. *International Journal of Human-Computer Studies*, *65*(3), 192–205.
Seppelt, B. D. & Lee, J. D. (2019). Keeping the driver in the loop: Dynamic feedback to support appropriate use of imperfect vehicle control automation. *International Journal of Human-Computer Studies*, *125*, 66–80. doi:10.1016/J.IJHCS.2018.12.009
Shariff, A., Bonnefon, J. F., & Rahwan, I. (2017). Psychological roadblocks to the adoption of self-driving vehicles. *Nature Human Behaviour*, *1*(10), 694–696. doi:10.1038/s41562-017-0202–6
Sheridan, T. B. & Ferrell, W. R. (1974). *Man-Machine Systems: Information, Control, and Decision Models of Human Performance*. Cambridge, MA: MIT Press.
Sidarus, N., Vuorre, M., Metcalfe, J., & Haggard, P. (2017). Investigating the prospective sense of agency: Effects of processing fluency, stimulus ambiguity, and response conflict. *Frontiers in Psychology*, *8*, 1–15. doi:10.3389/fpsyg.2017.00545
Slovic, P. (1987). Perception of risk. *Science*, *236*(4799), 280–285.
Slovic, P. (1993). Perceived risk, trust, and democracy. *Risk Analysis*, *13*(6), 675–682.
Slovic, P. (1999). Trust, emotion, sex, politics, and science: Surveying the risk-assessment battlefield. *Risk Analysis*, *19*(4), 689–701.
Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2004). Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis*, *24*(2), 311–322.
Tomlinson, E. C. & Mayer, R. C. (2009). The role of causal attribution dimensions in trust repair. *Academy of Management Review*, *34*(1), 85–104. doi:10.5465/AMR.2009.35713291
van der Wel, R. P. R. D. (2015). Me and we: Metacognition and performance evaluation of joint actions. *Cognition*, *140*, 49–59. doi:10.1016/j.cognition.2015.03.011
Verberne, F. M. F., Ham, J., & Midden, C. J. H. (2012). Trust in smart systems: Sharing driving goals and giving information to increase trustworthiness and acceptability of smart systems in cars. *Human Factors*, *54*(5), 799–810. doi:10.1177/0018720812443825
Victor, T. W., Tivesten, E., Gustavsson, P., Johansson, J., Sangberg, F., & Ljung Aust, M. (2018). Automation expectation mismatch: Incorrect prediction despite eyes on threat and hands on wheel. *Human Factors*, *60*(8), 1095–1116. doi:10.1177/0018720818788164
Weast, J., Yurdana, M., & Jordan, A. (2016). *A Matter of Trust: How Smart Design Can Accelerate Automated Vehicle Adoption. Intel White Paper.* Santa Clara, CA. doi:10.1016/S1353–4858(19)30060–1
Wegner, D. M., Sparrow, B., & Winerman, L. (2004). Vicarious agency: Experiencing control over the movements of others. *Journal of Personality and Social Psychology*, *86*(6), 838–848. doi:10.1037/0022–3514.86.6.838
Wen, W. & Haggard, P. (2018). Control changes the way we look at the world. *Journal of Cognitive Neuroscience*, *30*(4), 603–619. doi:10.1162/jocn
Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience*, *5*(3), 277–283. doi:10.1038/nn816

Woods, D. D. (2015). Four concepts for resilience and the implications for the future of resilience engineering. *Reliability Engineering and System Safety*, *141*, 5–9.
Wooldridge, M. & Jennings, N. R. (1995). Intelligent agents: Theory and practice. *Knowledge Engineering Review*, *10*(2), 115–152.
Wu, K., Zhao, Y., Zhu, Q., Tan, X., & Zheng, H. (2011). A meta-analysis of the impact of trust on technology acceptance model: Investigation of moderating influence of subject and context type. *International Journal of Information Management*, *31*(6), 572–581.
Zuboff, S. (1988). *In the Age of Smart Machines: The Future of Work, Technology and Power*. New York: Basic Books.