Autonomous Individuals in Autonomous Vehicles: The Multiple Autonomies of Self-Driving Cars

ERIK STAYTON
Nissan Research Center – Silicon Valley & Massachusetts Institute of Technology

MELISSA CEFKIN
Nissan Research Center – Silicon Valley

JINGYI ZHANG
Nissan Research Center – Silicon Valley

We take the polysemy at the heart of autonomy as our focus, and explore how changing notions of autonomy are experienced and expressed by users of self-driving cars. Drawing from work-practice studies and sociomaterial approaches to understanding technologies, we discuss how driving as a task is destabilized and reconfigured by the introduction of increasingly automated systems for vehicle control. We report on the findings of a hybrid ethnographic experiment performed at Nissan Research Center – Silicon Valley, in which we video recorded interactions of 14 participants inside a simulated autonomous vehicle, and conducted semi-structured post-interviews. We look at the responses of our participants in light of three different themes of autonomy, which emerged through the analysis of the study data in the context of a broader program of ethnographically informed research: autonomy as freedom from the task of driving; autonomy as independence and individual labor; and machinic autonomy’s ironic opposite, an increasing interdependence with human-machine systems that raises new issues of trust and control. We argue that AV development will have to address the social dimensions of roadway experience, and that this will require a multi-perspective approach (speculative work alongside other empirical examinations) to the specific ways human autonomy and sociality is aided, altered, or undercut by these systems.

“Finally, when everything else has failed, the resource of fiction can bring—through the use of counterfactual history, thought experiments, and ‘scientification’—the solid objects of today into the fluid states where their connections with humans may make sense. Here again, sociologists have a lot to learn from artists.” (Bruno Latour, Reassembling the Social, p. 82)

INTRODUCTION

“Could you turn on the autonomy for me?” A few beats of silent confusion follow. “See the little button on the right side of the steering wheel that says ‘CRUISE ON/OFF’? Press it for me.” After a moment’s pause, the participant has found the button and things start to move. The system management displays tucked behind the simulator show the little vehicle icon beginning to glide along schematically represented streets. This being a simulator, the car itself doesn’t go anywhere, but around it, on 360 degrees of screens, the virtual terrain begins to shift, a little sickeningly. “Ok, press it again. Good. When I ask you to start the experiment, just press that button again.” The simulator car’s cruise control button, made unnecessary when the vehicle was immobilized—engine removed and wheels propped up—has been co-opted to instead switch the software automation systems. This button, in the lab’s parlance, turns the autonomy on.
But what is really being turned on and off, being enabled and disabled, in this interaction? *Autonomy* is a multifaceted and complex notion. It can evoke the autonomy of the liberal individual, the autonomy of the nation-state, and the autonomy of the self-operating machine. The language of autonomy is in tension between technical and colloquial use. Speaking of the *autonomous vehicle*, then, often elides the question of “what kind of autonomy?” Or “whose autonomy?” Autonomy from or relative to what? There is a seeming self-evidence in the notion that the machine is autonomous because it is somehow operating outside of human control. But we suggest that when people talk about the “autonomous car” they do not simply mean “the car that is autonomous,” but also “the car that makes *me* more autonomous.” So what needs to be asked may be less about the technical capacities of the system, and more about its human meanings. What kinds of new interactions are being produced? What do users give up to gain the convenience, or “autonomy,” they believe they want? And do they really want it when they have it?

This dynamic highlights the strange polysemy at the heart of autonomy: one may be freed from certain tasks but also further embedded in sociotechnical systems that are beyond individual control. Here we explore this dynamic as a speculation on a future with increasingly automated vehicles in our midst. As developers of automated vehicle systems, we are implicated as part of the source of users’ struggles; we too are trying to come to terms with what kind of a world we are involved in producing, even while we attempt to direct it for the better. This study is an attempt to reckon with the direction of that future. First we further develop the background and approach to our investigation, which takes the form of a hybrid “ethnographic experiment.” Then we examine three key themes that emerged in our research: the contingency of autonomy, the awkwardness of monitoring and being monitored, and the difficulty of trusting in humans and machines. We argue that the price of achieving one sort of autonomy is perhaps the sacrifice of another; and that users recognize this, as they struggle to come to terms with the ways their existing ideas about trust, and practices of interacting with vehicles, must shift in relation to machine autonomy. Finally, we reflect on our use of a speculative approach to elicitation in our attempt to design new relationships between human and machine autonomies.

**BACKGROUND AND APPROACH**

Our theoretical and methodological perspectives draw from literature in work-practice ethnographies, actor-networks, sociomateriality, and grounded theory: in this tradition, we take seriously the situatedness of human action. Ethnographies of all manner of practice have long exposed the contextual nature of meaningful human action. What we do is dependent on our environment. As Jean Lave describes in *Cognition in Practice*, the shopper may not know quite what she is buying until she sees it on the shelf, and is confronted by the options before her (Lave 1988). Or as Hutchins shows, pilots communicate and form plans, not as individual brains with separate mental capacities, but as a “cockpit system” with “cognitive properties” defined by social and material factors: people, radios, gauges, pips, and paper cards (Hutchins 1995b). Technical approaches to driving split driving into multiple kinds of tasks—e.g. Strategic, Tactical, and Operational components, that break down the act of driving into trip planning and route selection, maneuvering, and split-second responses (Michon 1985). In contrast, we attend to driving as a cultural and sociomaterial practice. In other words, driving is a practice that happens in relation to others and the
world, emerges from the interactions of social actors and material objects, and which makes meaning as it serves practical needs. Drivers do not just perform tasks. They have bodies and cultures. A focus on the embodiment of work likewise exposes, in what might have seemed empty from an information-processing vision, hidden plenitudes; an ancillary activity such as accounting (Suchman 2011), or in our case performing responsibility in mobility, may become a key source of social meaning.

For example, Lutz and Fernandez suggest that automobledom has become implicated in the “myth that good parenting” in the modern cultural mode “means ferrying one’s children in the car” (Lutz and Fernandez 2010, 26). Such ferrying is not simply operational, getting one’s passengers from A to B, but is about caring, providing for, and performing the role of guardian. Thus we should not expect that replacing the parental driver with an autonomous robotic chauffeur should leave participants’ affective relationships unchanged. Even the Vatican, hardly the first place one thinks of as a bastion of revolutionary sociology, has identified driving as a social act: their guidelines for Pastoral Care of the Road state that driving is “basically a way of relating with and getting closer to other people, and of integrating within a community of people” (Lutz and Fernandez 2010, 158). The social extends beyond the technological frame of driving as mechanical control.

Science and technology studies work has shown that supposedly autonomous systems are rarely so in practice; “full” autonomy is a mirage, and even systems that might seem quite outside of human control, like Mars rovers, are part of complex systems of human oversight and joint action (Clancey 2014; Hutchins 1995a; Mindell 2011; Mindell 2015). Nissan has taken the approach of embracing joint human machine control. One manifestation of this is the Seamless Autonomous Mobility (SAM) concept, in which remote human vehicle managers can step in to instruct the automated vehicle (AV) in problem situations. This “teleoperations,” or human supervisory control, approach (Sheridan 1992; Woods and Hollnagel 2006) keeps humans in the loop to handle edge cases and novel situations not yet learned by the system. It also opens up all manner of new human-machine interaction considerations. The literature in human supervisory control is likewise clear about the fact that automation does not merely eliminate, but changes the tasks performed. But sociomateriality extends this reductive, task-based thinking. Humans in cars do not merely move wheels and pedals in functional ways. They negotiate and wayfind (Brown and Laurier 2005; Keisanen 2012; Laurier, Brown and Lorimer 2012). They express their autonomy as mobile subjects (Bishara 2015). Humans in automated cars will share many of these practices. And these practices matter for how vehicles will be thought about and used.

The characteristic elision of sociomaterial complexity that underlies “autonomy” is not unique to automated vehicles. It appears across modern design and engineering practice. Any organization that tries to make a product or service better, easier, faster, or more efficient for the user inevitably faces the question of who their user really is (Cohen 2005), and what their disruptive innovation really does. For whom, and from what perspective, do things become easier? Or more difficult? Ethnographic studies of collaborative work practice (Cefkin 2014; Cefkin, Thomas and Blomberg 2007; Suchman 1998), and sociomaterial approaches to technology (Orlikowski and Scott 2008; Scott and Wagner 2003; Suchman 2007), have exposed the complexities of these kinds of questions. Technical interventions reconfigure existing ways of doing things that have developed through intermeshing of human needs and technical affordances. The social and material develop together, and
change each other; but sharp breaks in the material properties of work systems force corresponding restructurings of social processes.

For example, Suchman and Jordan (1989) argue that information processing tasks in the workplace are often automated without attention to real complexities, focusing instead on the small task components that are amenable to ICT-based approaches. The resulting tools, awkward and often ill-fitting prosthetics for labor, require new adaptations by remaining workers. This pattern of “appropriation” (Suchman and Jordan 1989) applies equally to the automation of driving. Since the task of driving is more than rule following—staying in the lane, obeying lights and signs—to drive is not only to navigate through physical space, but through a social space of symbols and cultural signals (Bishara 2015; see also Goffman 1963). When one extracts the mechanical components of driving and replaces them with a new sociotechnical system of automation or “heteromation” (Ekbia and Nardi 2014), one gets the sense that automation could proceed from partial to complete in a piecewise fashion. But this is an illusion: the task of driving, and its social meanings, would not remain fixed in this transition. Practices are moving targets.

Appropriations in design are always partial. What tasks can be productively automated, and how, is a constant problem for the development of automated systems—and a key issue for us as autonomous vehicle designers. Many questions emerged for us in thinking about the human side of supervised autonomous control: How would human passengers respond to oversight or intervention by remote human beings? How long would they wait at an obstruction for a vehicle manager to bail them out? And how comfortable would they be about that interaction? How would they perceive their new relationship to the vehicle system? Adding autonomy to vehicles is a moment in which we must ask how the rest of driving practice, cultural and psychological, will respond. But we face the difficulty of how this can be investigated empirically.

These changes are still speculative ones, as the systems that stand to precipitate this restructuring are still in development. Building on critiques of the doctrine of studying the “out there” and in the spirit of anticipatory or speculative ethnography (Halse and Clarke 2008; Lindley, Sharma, and Potts 2015; Nafus and Anderson 2006; Venkataramani and Avery 2012), we have had to make our own microworld in which to observe these phenomena.

This paper draws especially from data gathered during a simulator experiment performed by social science researchers at Nissan Research Center – Silicon Valley. On first glance, our materials are not particularly ethnographic. Participants experienced a series of interactions as if they were in an autonomous vehicle that was driving them to a meeting on NASA Ames campus. The simulator used had 360-degrees of screens around a real vehicle at its center. Each participant experienced two short drives in which events in the simulated world required the vehicle to come to a halt. We video recorded their responses to these situations, and performed post-interviews. We gathered approximately 7 hours of relevant video data, and 7 hours of interview data, from a total of 14 people. However, we did not approach this data from a functionalist, experimental perspective (for example, one interested in measuring reaction times, or quantifying the user’s gaze). Instead, we examined the data anthropologically, looking at users’ interactions with the system as material that expressed their perspectives on the system, their beliefs about it, their comfort or discomfort with it, and their needs, wants, desires, and systems of meaning and interpretation. This experiment was one of several elements of a broader program of research into the social
implications of autonomous vehicles (Vinkhuyzen and Cefkin 2016), which also included field observations, interviews, and other ethnographically informed approaches.

In this particular study we observed participants as they encountered two kinds of obstacles in their autonomous vehicle, a construction zone and an accident. In real life navigating such instances requires drivers to assess the appropriate maneuver—to wait (and for how long) or to go around (when it is appropriate to do so)—and to make a potentially illegal move that is nonetheless consistent with the expected rules of the road in this instance: crossing a no-crossing line (in the United States, a double yellow line) and passing on the wrong side of the road. The AV would require a new path to pass the scene, and it was here that a remote supervisor was available to assist. Using the on-board sensors, the remote supervisor could assess the situation and send the AV new instructions. Our question was whether participants would take over for themselves—they were free to take over manual control at any time, though they also had secondary tasks to perform on their devices—or let the remote supervisor do so. We also wished to identify when additional information or status from the remote supervisor would be sought by the participant.

After participants experienced the two drives, we performed semi-structured interviews with an eye toward eliciting why participants chose to preempt or wait for the automation system at various points. And we sought to identify what aspects made them comfortable or uncomfortable, how they made sense of these issues, and how they would feel about using a similar system in the real world. This hybrid mode of investigation, building from design anthropology, is a way for us to overcome the difficulties in studying speculative objects. Technologies that do not yet exist must be imagined or brought into being as they are investigated. We undertook this study with an ethnographic sensibility, intending to examine the patterns of life that would emerge in the day-to-day interaction with the technology.

This investigation exposed a variety of fascinating responses to the experience of being conveyed around by an automated vehicle in a simulated world. Autonomy, as our participants describe, is a partial and contextual thing, which must be negotiated between humans and machines. It also implies a freedom from restraint that conflicts with, and must be rethought in light of, remote human monitoring. And it demands a level of trust in human-machine systems that brings with it concerns about privacy and surveillance. These multiple autonomies (from labor, from others, from oversight) are the stage for coming conflicts about the value and purpose of mechanical automation, on and off the road.

WHEN IS MACHINE AUTONOMY DESIRED?

The automobile, as perhaps the ultimate tool for individual mobility, is intended to be convenient. Buses and trains run on schedules. They require waiting at stations, and transferring from one to another means even more waiting (or, even worse, missing connections entirely). A person who hops on the train cannot simply go where she wants—the train traveler, as in The Practice of Everyday Life, is regulated and immobilized by the chiasm of the window and the rail, which makes change visible but prevents the subversion of motion (de Certeau 1984). The car, by contrast, is the choice of the liberated individual who wants to move on demand: where she wants, when she wants. Automobiledom promises “independence from reliance on the schedules and desires of others” (Lutz and Fernandez 2010). Our participants revealed to us that vehicle autonomy is indeed desired when it adds to human autonomy, and when it frees people from tasks they dislike, but not necessarily
when it limits their perceived freedom. Machine autonomy is contextually, not universally, good.

This whole notion of car-based freedom is, as Fernandez and Lutz point out in *Carjacked*, a pleasing and socially costly illusion. The automobile as a tool of individual mobility has been historically inseparable from a new kind of experiential imprisonment. Car travel is in its own ways profoundly constrained and inconvenient. The traveler finds that the roads are never clear just for her. Other people get in the way. Highway hypnosis, road rage, headaches; accidents, traffic jams, finding parking; wide avenues and suburban sprawl; breakdowns, maintenance, repair; even smog and pollution: these are the costs of the automobile. So the autonomy of automobility brings with it the convenience of going where one wants to go, but also the inconveniences of traffic, risk, and mental and physical labor. And these are among the problems the automated car seems poised, perhaps, to solve. The car, as a latent space of inefficiency and un-productivity, is perhaps ready to be “reclaimed” for sleeping, reading, eating, or most ironically for many of us, “productive” labor. (We would challenge the notion that time spent in the car, thinking, seeing, listening, and experiencing, is truly waste, but no matter.)

The participants in our simulator received a taste of this life of mobile leisure: whisked around a virtual map of NASA Ames, from one imagined “meeting” to another, they were free (and encouraged) to be on their laptops or phones as long as they were comfortable that the vehicle was operating safely. And most at least seemed to be. They glanced up a lot, especially at first, and a couple spent enough time looking out front that they did not finish the preparatory tasks we set them (fictional preparations to make for their meetings). But most were eventually engrossed in their devices. This level of focus sometimes produced amusing results. Nate, a 22 year old intern, suffered a simulator glitch that teleported him inside a truck—his simulated AV instantly jumped 25 feet down the road due to a human error in our configuration of the test. As the screens around him went entirely white, he looked up, shocked and confused, unsure of whether the vehicle had crashed into something while his attention was elsewhere.

In general, passengers possessed a marked ambivalence toward machine autonomy. It was convenient, to be sure. Though participants’ responses were clearly colored by knowing they were safely ensconced in a simulator, they reported enjoying the freedom to surf the web, write emails, and even to take in the simulated scenery without concern for crashing. But different participants displayed different levels of comfort with the operation of the system in the test, and imagined different responses to it on real roads. What is most surprising is that these responses were not binary, yes or no, “I would use it” or “I would not.” The context of use mattered significantly. Our post interviews exposed that participants perceived commuting to work or going to a meeting as qualitatively differently acts of driving than driving with one’s children or on weekends. These are different sociomaterial practices, and put the driver (or erstwhile driver) in a different relationship to safety, risk, and responsibility by virtue of their social relation to others in the vehicle, and their reasons for travel. Multiple passengers suggested that they would be more willing to entrust their own safety to the system than that of friends, coworkers, or family members. Responsibility to others in the car would be performed, our participants’ responses suggest, by taking over. If his partner was in the car, one said, he would turn the automation off. Exposure to quantitative measures of risk and safety—reduced accident rates—and more experience with the vehicle might alter these responses over time. But these responses show
that the quantitative measures of risk that dominate the discussion of AV development and AV ethics are disjoint from the actual experience of responsibility. Being responsible means more than being numerically safe. It means being accountable, acting, being in command.

By virtue of our working in a car company in Silicon Valley, many of our coworkers are white-collar “gearheads” (one author included). And so our population of internal testers skews toward this demographic. They are information workers with long commutes, for whom an automated car really could be an office on wheels. And yet many of them love to drive. As such, they might seem to embody a contradiction as they work to automate away something they love to do. Indeed, many of our passengers suggested they would override the autonomy in real life, or might turn it off in particular circumstances, relying on their own skills instead of programmatic ones. But even the car enthusiasts among them expressed contextual preferences rather than flatly opposing the use of vehicle autonomy. Not everyone who is excited about driving and motorsports is interested in always controlling their vehicle. Emily, an administrative assistant, declared that she looks forward to being able to be on the phone in her car, despite also being an avid motorcyclist. When asked if she wanted a self-driving motorcycle, she denied this vehemently. She replied: “I want to drive when it’s fun to drive and I’m in the mood” and not have to drive when tired, in traffic, or when the drive is otherwise “uninteresting.” Questions about comfort with autonomy have no blanket answer; participants generally differentiated situations in which they would comfortably use autonomy and situations in which they would not. So whether or not to use vehicle autonomy is a choice that is made and remade, not a single binary decision.

This feedback suggests that any solutions for teleoperated remote control of a vehicle must also be sensitive to contextual preferences. Its efficacy may vary depending on the passengers present, the purpose of the trip, and the conditions on the road. Humans inside the vehicle may wish to interact with the autonomy and supervisory systems in different ways. Where these lines are drawn may be deeply personal, and we have no general answer—though Emily’s response distinguished the city from the mountains, and traffic jams from the open road, others might cut up their world through a different sort of analytic. Driving is a social act in the quotidian sense of interacting with others in shared spaces, but as Bishara points out, driving also produces special kinds of socialities within the vehicle and between those in the vehicle and their environment (Bishara 2015). The road may be subverted, experimented with, made into a field for the construction of a driverly identity; particular roads or locations may be haunted by past events—accidents or breakdowns—and thereby require special attentiveness (Verrips and Meyer 2001). And car ownership and use itself may be a medium for social ties of responsibility to others (Myers 2017). Driving is a “technique of the body” (Bishara 2015, 36), and autonomy destabilizes its practices. Machine autonomy is not a natural good for people in cars, always, all the time. It is another thing that people may wish to turn on and off, something that must be made sensitive to the needs and desires of passengers on a particular trip.

THE AWKWARDNESS OF HUMAN MONITORING

The American imaginary of the automobile puts the lone individual on the road facing off singlehandedly against the wilderness. One need look no further than automobile marketing to see the preeminence of this idea. Across deserts, through green forests and urban jungles,
up and down mountainsides, our objects of automotive desire are flaunted before us as things untethered from the strictures of daily life. Though this image is always beyond our reach as the product of a carefully produced mediated fantasy—as the tiny white text on these advertisements often says: Professional driver. Closed course. Do not attempt—it still manages to compel. But the autonomous car, whatever its name, will never be “fully” autonomous. The automated car is a networked device, dependent on interactions with global information networks for everything from maps to traffic data to vehicle-to-vehicle or vehicle-to-infrastructure communications, so it is likely these vehicles will never be able to be unplugged (Stayton 2015). They will be, like our phones, connected devices; and, like aircraft, trucks, buses, and other fleets of vehicles, they will be remotely monitored and managed. Passengers may well become accustomed to this kind of connected experience, but the responses of our participants suggested this will be no easy or simple transition. Being monitored by a remote supervisors involves a distinct kind of driving experience.

Nissan’s SAM concept in particular puts remote human managers in charge of helping AVs through difficult situations. And the experience of this kind of remote management is fundamentally new to the average driver. Assistance services like OnStar exist, and already provide a significant amount of information to the personnel who manage the vehicles, but they do not yet direct the path of the car. Remote starter interrupt devices—installed for example by “Buy Here, Pay Here” used car dealerships to disable the cars of borrowers who get behind on their payments—get closer to the phenomenology of the remotely managed car. But these can only stop vehicles rather than making them move (Hill 2014). For all our participants, their simulated drive was the first in which they had been told that a human vehicle manager, located remotely, would be monitoring their vehicle’s progress and intervening if the vehicle came to a stop at an obstacle that the autonomy could not handle on its own.

Participants were told at the start of the experiment that there was a human teleoperator who would be monitoring and could provide assistance. In addition, passengers were always notified of the human teleoperator’s engagement. A display in the dashboard provided the car’s status: “Waiting for Supervisor,” “Supervisor Engaged,” “Following Supervised Path,” as the vehicle waited for assistance, registered its connection to a remote manager, and then carried out that manager’s instructions. This low-impact approach meant that for many participants, the supervisor faded into the background even to the point of invisibility. When they had to wait, they were waiting to see if “the car” could “figure things out.” Several, including Charlize, a 25 year old analyst working in human resources, reported that they did not think about the involvement of the remote human until they had been stopped for some time—agitated, looking ahead at the construction zone in front of her, she touched the wheel to take control just as the supervisor’s instructions made it to the car, some time after her vehicle had reached the scene; “Okay then” she muttered under her breath, her tone conveying surprise mixed with some annoyance. A rare few, like Emily, never thought about the supervisor at all.

But when they did think about her—when pausing caused participants to reflect on the nature of their relationship to that remotely located human they had potentially never met who was about to take control of their vehicle—responses turned to interesting directions. Nanak, a summer intern working on vehicle simulation, said he did not try to contact the human operator because he did not want to “bother” her. He explained that he could clearly see and handle this situation himself. So why would he involve a skilled operator whose
services might be needed elsewhere? The assumption that the operator was busy dealing with more complicated tasks than his led him to try to handle the situations alone: had something very complicated, difficult, or confusing come up, he suggested, he might have preferred to trust this professional to handle it. The teleoperator in Nanak's vision was an expert resource for extraordinarily difficult or challenging situations, not simply an effector for routine maneuvers that are still beyond the capabilities of the autonomy alone. Passengers using their vehicles day-to-day would certainly have greater opportunity to become accustomed to teleoperation, and evaluate when it is helpful to have a remote vehicle manager involved in operations, but this passenger's comments suggest that the use-cases for human supervision are open to individual interpretation. And the presence of that supervisor brought a new social politics into the equation of driving: that of the value or sanctity of the individual's labor. Another participant suggested that the mere presence of a human supervisor somewhere in the system acted to prevent his own overriding of the vehicle. He reported an awkwardness around “taking away their job,” a feeling that would not have been present had the system been a fully computerized one. The remote human, unlike the machine, still has a certain dignity, and one may feel the need to respect her time, skills, and execution of her tasks. Being in a supervised vehicle presents complicated questions about the social mores of intervening with the work of people located elsewhere, mores that are not yet set and therefore likely all the more awkward to negotiate for the first time.

But this awkwardness was apparently a mixed experience. On the one hand, this participant reported reluctance to interfere with the human-machine system of autonomy: “I don’t know if I can” take over, he said in our interview, recapitulating his previous thought process, because “somebody else is in charge of my car.” His affect, delivering these lines, evoked concern. He seemed troubled. But he also experienced what he described as “a little relaxation that happens” on seeing someone else in control of a situation. He phrased this relaxation as a general principle, a lay theory about concern and responsibility: obviously, someone else being responsible would make you feel at ease. But this lay theory did not hold universally. The status of the supervisor as a component in the system—what that supervisor was presumed to be there to do, and how much information he or she was presumed to have—seemed to have much to do with participants’ varied concerns about their interactions. Joshua, a summer intern working on connected vehicle systems, trusted the operator more than he trusted himself: the sensors would be better than his eyes. He explained that he assumed that operator would have sensor feeds from multiple cars, and would therefore know more about the situation than one human’s first-person view could ever show. This utopian human-machine system made him more comfortable than he would have been in a cab: it, unlike a cab driver, was “programmed” to keep him safe. His increased comfort, however, does not negate the potentially awkward aspects of now being under the authority of some remote and unknown person. And Joshua’s comments cut against the grain of statements by many other passengers who wondered how a remote supervisor could ever react as competently as they could, with their own first-hand knowledge. For these passengers, contending with this remote human agency was uncomfortable and destabilizing, a new practice of negotiating conflicting desires (to take over) and responsibilities (am I allowed to take over?). These different views, and their different affects, suggest different assumptions about the technical capabilities of both vehicle and vehicle management center: Joshua was working on a project to collect vehicle
data from On-Board Diagnostics (OBDII) ports, and centralize it on a cloud data platform. His assumptions about connected vehicles and their capabilities are perhaps more reflective of his own work than the simulated drives he experienced. This exposes an important point about mental models of supervised operations: what passengers believe will be formative for their interactions with the system.

Interactions with remote vehicle supervision systems require passengers to remake assumptions about the individuated driver cocooned away from the rest of the world. Participants were called to reckon with their new interrelationship with another human being capable of controlling their vehicle. And this relationship could be an awkward one; for some it brought to attention the expertise and status of the remote operator: What sort of tasks ought she be called to attend to? Is it rude to preempt her labor? For some, supervisors seemed remote, in knowledge as well as location; for others they were more present and capable than someone actually on the scene. But all these questions of authority, comfort, trust, and jurisdiction are embedded not only in the issue of capability, but that of responsibility. The interjection of autonomy and a remote supervisor into the car changes the sociomaterial practices of driving responsibly. Driving does not remain the same when the driver’s individual agency—albeit mediated and constrained by law and custom—is no longer wholly in charge. Old assumptions no longer hold. Who should do what, and whether new parts of the system have responsibility to us (or whether we have responsibilities to them) must be determined anew. And drivers express their experiences of figuring this out with an affect of concern and discomfort.

TRUSTING IN HUMANS AND MACHINES

The central irony of the development of automated systems is that, at least in some ways, the more automated the system is, the more interconnected it must be with vast networks of humans and machines outside the individual vehicle, which must be trusted to operate appropriately. The individual human in a truly manual vehicle can navigate the world. They cannot do this entirely autonomously—bound by social systems, by law and custom, by prior knowledge of the environment, by past experience and sensorimotor capability—but they can at least convincingly mime that autonomy. The autonomous vehicle must be bound and controlled by code, and so can never be so free. This means that passengers within are forced to contend with new networks of control: human supervisors can be directly compared to the computer systems delegated to perform the watching-over on a moment-by-moment basis. We asked our participants to trust this system, to leave it on as much as they felt comfortable. Though they were able at all times to take over and drive manually, none did unless the car was headed for one of the obstacles we had set up to provide reasons for human intervention. When participants encountered these situations, responses varied widely. Some took preemptive control to bring the vehicle to a stop and then turned autonomy back on; others took over only after the vehicle had been stopped for some time; and still others left the vehicle to its own devices throughout the entire situation. But leaving autonomy engaged was not a sign of complete trust. Both taking control, and monitoring the progress of the system while leaving it in control, are ways to moderate a distrust in its capabilities. And issues of trust were not limited to mechanical, operational parameters. What the system knows may be just as important as what it does. This trust has gradations, and treats humans and machines in different ways.
Participants had diverse feelings about placing trust in a human operator. Mark, an intern with the vehicle autonomy team, felt the remote supervisor, “in [his] book, could do no wrong.” Obviously a professional, this supervisor would be able to handle issues without difficulty. Doatea, who spent several years working in India where she was chauffeured around every day, recognized no meaningful difference between a driver in the car, and a supervisor outside of it. But many participants seemed less willing to trust a human than a machine. As Emily put it, she would rather trust software “that’s been created to make this work,” by “hundreds of engineers spending hundreds, or thousands of hours,” than trust a human of unknown skill and professionalism. Jean Lo, an intern with Renault and co-worker of Nanak’s, described that with a computer, “you trust software, security, encryption,” but how can you be sure you can trust the remote human? This sort of thinking was a common refrain, though most who felt this way came to see the situation more positively when they were informed that the supervisor was not “joysticking” the vehicle (taking direct control of the wheel and pedals from afar), but was instead just plotting a path for the autonomy to follow.

The standards for performance between human and machine could go either way: James would allow some “delta to [his] expectation” before intervening with a taxi driver, but supposed his tolerance was less here; Nate, when he found out a human error had caused him to go over a curb, was less forgiving than he would have been to the machine: “You could have done a better job! . . . You are a human, so that’s different.” But in any case, the automated safeguards that were operating at all times seemingly provided a reasonable basis for trust, as they meant the unknown vehicle manager could not, presumably, cause the passengers harm. Many passengers still had trouble conceptualizing why a human was important to the system at all, but felt safer knowing the autonomy still handled moment-by-moment decision making. This privileging of machinic reliability over human caprice is widely recognized, in various ways, across studies of information technologies. The 2016 revelations about Facebook’s trending topics (Nunez 2016), for example, dramatize the collision of algorithms’ putative mechanical objectivity and humans’ putative biases: it was shocking and controversial that these trends, supposedly representing major discussion topics on Facebook, were curated by human analysts rather than being generated by a presumably “neutral” computer model. This tendency toward Latourian disciplining and delegation—“never rely on undisciplined men, but always on safe delegated nonhumans” (Latour 1988, 305)—assumes that nonhumans can be made safe and dependable, moreso in the absence of human inputs. It is a belief system, not a statement about reality, which has more to do with cultural preconceptions about the properties of the organic (creative, capricious) and the machinic (predictable, dependable) than it does about their actual operation. Human involvement appears as risk in part because we are not accustomed to thinking about its ubiquity. The value of joint human-machine systems is difficult to parse from a perspective that trusts the reified technical object, and does not attend to the continued human effort that is always required, in some form, to get such systems to behave properly. When joint-ness is seen as a weakness, rather than a strength, trust in the system decreases with human involvement.

That is not to say that our participants were ignorant. Far from it. The virtues of humans and machines are up for debate, and we did not select for expertise in human-machine systems engineering. However, our participants were measured in their preferences, and subtle in their critiques. Charlize explained to us: “I think the remote supervisor does make
me more confident in it [the car], but what if they aren’t paying attention?” In this view, the unreliability of algorithmic responses can be compensated for by the joint involvement of people. But those people remain at least potentially untrustworthy. Many participants expressed such lay theories about trust and its distribution. A few made interesting suggestions about how to heighten a feeling of trust: Jesse, a visitor from another Nissan lab, wanted to know “the name of [his] guardian angel, even if [the system] lied to [him].” The simple touch of seeing a human name would have made him feel better, more connected to the person partially in control of his fate. Another participant, Marianne, stated a preference for an “Uber-like” star rating of the vehicle supervisor, so the passenger could have a sense for the skill and training that the supervisor possessed. It is unclear how a passenger would respond to an unrated or low-rated operator (one assumes not well). But these comments, taken together, seem to suggest that human-ness is not a one-way street: it does not monotonically decrease trust.

Putting a more human face on the supervisor might help some people get comfortable with a remote human role in vehicle operations. The skill, professionalism, and training of remote supervisors, and how the vehicle users are made aware of these qualities, may be critical to the acceptance of teleoperation as part of a new practice of driving. The association of remote teleoperators with call-center customer support representatives, explicitly made by several passengers during our interviews, invoked serious doubts about the capabilities and motivations of the human components in the vehicle management system. Emphasizing an appropriately professional work culture among teleoperators might go a long way to addressing these types of concerns. These findings should not be that much of a surprise, in the context of our prior discussion, as they lead back to the issue of responsibility, and the social contract between driver and passengers. The unknown and remote person cannot be trusted because their relationship as a responsible party is not clear. They cannot be held to account. Making that supervisor somehow known starts to engage them again with preexisting expectations for delivering care—the apex of that responsibility being that of one family member to another.

While the operational integrity of vehicle managers might be more positively framed by association with air-traffic control than with customer support, issues of trust that fall outside of the operational, into realms of security and privacy, may be more difficult to solve. While some passengers felt that a human operator made the system “friendlier,” this was generally interpreted as coming at an inherent cost to privacy. Recognition that other services like OnStar already involve vehicle tracking made the teleoperator more palatable to Jean Loup. But he still wanted to be able to turn supervision off in order to drive unmonitored. Ling, a design intern working on vehicle interfaces, expressed in his interview that if there were a human supervisor with knowledge of his location, he would feel as if someone were “stalking” him. His use of this particular term conveys a personalized dimension to this kind of monitoring. Surveillance is an impersonal thing, doubly so the notion of “mass surveillance.” But stalking is personal, human-to-human, a direct invasion of expectations of privacy. Talking to a virtual agent instead of a person would, Ling suggested, solve this affective problem for him, even if the data collected by the system was the same. But if there were ultimately a human pulling the strings of that agent, some of his privacy concerns would remain. And as Marianne, a service designer, was quick to point out, it would be inappropriate to hide the level of human involvement. People have a right to “know what is real” behind the operation of the system, she urged. While a more human
face to remote vehicle operations seems likely to help some people trust those involved in vehicle control, it may make the privacy risks more obvious to others. And at the heart of all these comments lies a paradox worthy of further study: if the information collected is the same, why did participants’ lay theories consistently lead them to be more concerned about a human than a machine having access to that information? It is not necessarily true that the engagement of humans really makes one’s information more vulnerable than if data is only being mined by automatic scripts! This is a complicated question with many possible answers depending on how data is stored and used. But if these participants’ responses are indicative of a generally held perception, they represent a challenge to the involvement of a human operator in the supervision of vehicles on the road.

The issue of how human the operator should seem is therefore a tricky one. Having the supervisors speak in their own voice could make them more human, but participants were not convinced they would enjoy that kind of experience. Ameila, a developer working on connected vehicles, expressed this as pushing up against the notion of the car as a private space, “my personal space” — which recalls responses to the telephone more generally, in its early days, when it was a site of potential transgression by outsiders into the sanctity of the domestic sphere (Marvin 1990, 64, 85). Many passengers preferred the idea of a Virtual Personal Assistant (VPA), or a human who spoke in the consistent voice of a VPA, to reduce the strangeness of an unknown person taking over their car’s aural space. One suggested that such a computerized voice would allow her to develop a relationship with her vehicle, rather than feeling like others kept intervening, entering the private space of the vehicle cabin. But the idea of a VPA does not work for everyone — notwithstanding that VPAs often work better in theory than they do in practice. Ameila reported that she would still prefer a conversation with a person as opposed to one with a machine that tries to “translate what you say” and “Google the answers, Googles the wrong thing,” etc. She has had bad experiences with Siri and Google Now not being able to understand her voice, and being otherwise unreliable even when they are able to correctly interpret her words. While technical progress may ease some voice interface issues, the operation of a motor vehicle is a sufficiently high-stress area that any communicative difficulties may be exceedingly detrimental to the passenger’s experience. Moderating privacy concerns or feelings of unease by the use of a computerized voice may be a useful technique, but even its supporters agreed it risks treading into ethically worrisome waters if the role of the human being is too obscured. The imposition of unbidden voices on the personal sphere of the cabin is exposed here as a potentially fraught enterprise.

This leads to the third sense of our autonomy, and its opposite, the further imbrication of human action and its dependence on new technical systems. For humans to work, machines must work too. For automated cars, and particularly teleoperation systems, to work, data must flow out to remote locations, to be operated on by unknown combinations of humans and machines. Commands and queries must flow back, and become part of a new sociomaterial space for the vehicle’s passengers. The autonomy of the human driver is complicated and impinged upon by these networks, which make possible the autonomy of the machine. What information a passenger is willing to divulge likely depends on a wide variety of factors — akin to those we have seen previously, related to safety and risk: Who is in the car? Where is the vehicle going? What is the purpose of the trip? The threat of the remote and unaccountable observer is very present for these commentators, though it is curious that this threat seems to be more alive in humans than machines. This sense of
threat goes hand-in-hand with the privileging of computational rules over bureaucratic structures of responsibility: the system has been programmed to “keep me safe,” but the cab driver, whose performance is still monitored, managed, and constrained by social, legal, and bureaucratic systems, has somehow not. Reckoning with these new sources of trust and distrust will be a key part of learning to live with automated cars in the real world.

CONCLUSION: AUTONOMOUS DRIVING AS A SOCIAL PRACTICE

The autonomous car that participants experienced is not \textit{autonomous} in the most obvious sense: naively free from human engagement. It is and must be an arrangement of humans and machines working together, with all the challenges that implies (Bainbridge 1997; Casner, Hutchins and Norman 2015). As machines threaten to exhibit their true autonomy, the freedom and indeed propensity to err, to do things we do not want them to do, they are always at the boundary of struggles between human wills and material obduracy, mediated through systems of control that are neither clearly human or clearly machine: they are sociomaterial. The autonomies that were involved in this delicate dance were not restricted to those of a machine operating on its own. Our participants encountered aspects of machine autonomy that were experientially new to them in this context—assumption of risk, remote management, data collection—that they had to square with their own positions: as individuals responsible to others through their embodied skills; as independent decision-makers free from oversight; and as drivers valuing safety, personal space, data, and privacy both for themselves and others in their vehicles. The latent visions of driving that are explored here are not the same as driving today, nor are they the same as each other. Different participants’ lay theories about trust and responsibility colored their responses to the system that they experienced. The driver who no longer controls the car does not simply sit there with mind, hand, and feet newly freed; these all become occupied by new tasks, new potentials, and new concerns. Can I take over now? Should I? Do I need to brace myself? Is it safe for me to let the system run? What is that system anyway?

These details could not have been seen so clearly without putting participants in a position where the autonomy theater that they experienced was convincing enough to destabilize their notions of their own role. But without some theater, it could not have been seen at all. Our hybrid ethnographic-experiment reinterpreted the traditional tools of a laboratory user-study through an ethnographic lens in order to combine the unique strengths and perspectives of these two fields of endeavor. To do certain kinds of research, we need new vantage points. We need to be able to produce new interactions, knowing full well that what is produced is partly artificial and must be approached with care to the claims that can be made. Such issues are not new to design anthropology (Gunn, Otto and Smith 2013) nor to anthropology as a whole, as it has long examined people’s other worldly and future-oriented hopes and expectations.

We see simulation as a viable means to produce new ethnographic knowledge, though we recognize as others have in various contexts that the knowledge and experience produced by simulation is not going to be quite the same as the “real world” (e.g. Turkle 2009). All speculations are in a sense contrived, but simulation provides one way to get a glimpse into a possible future. Participants in simulated interactions get to experience, even if only briefly, a different set of sociomaterial relations. And these experiences can then be investigated with other methods of elicitation. We resist the idea that findings can be wholly prescriptive, that
they can tell us how to produce new systems whole cloth. But they highlight new questions, new lacunae that require further investigation. And these experiences can open up the participants themselves to new ideas: James, who hadn’t thought that a supervisor would make him feel uncomfortable about taking over; or Joshua, whose assumptions about the supervisor’s knowledge were challenged by his experience, have new ways to think about human autonomy in their own work as developers.

We do not find it sufficient for our purposes to engage in only this kind of research. An anthropological investigation into a speculative future, without sufficient grounding in the present, is at risk of becoming unmoored from any semblance of reality; and applying the fruits of this investigation in a principled way requires careful explorations of its foundations. It is therefore important to us that our experiments in simulation are only part of a multifaceted study of road-use behavior, from focused roadway studies using close readings of video data, to more traditional industry anthropology fieldwork within transit organizations, which in their own ways inform our treatment of the questions here. If the practice of ethnography is of a sort of apprenticeship into existing culture, this speculative ethnography is apprenticeship into new kinds of destabilization. But in our joint roles as social scientists, developers of AV technologies, and designers interested in producing a better future, we sometimes encounter questions to which the world “out there” is incapable of providing all the necessary insight. At least to open questions, if not to close them, simulation as a playground for experiences can provide access to new sociomaterial practices that can then bound and shape development.

In light of this, we end with some of the questions opened by our investigation. Key among the questions for developing socially acceptable autonomous vehicle systems is this: Whose autonomy, or what autonomy, matters? Does a loss of autonomy from supervision always accompany a new freedom from labor? Or how would this be balanced in practice? Do some of these autonomies impose limits on how the technology operates, which might well change the functioning of the resulting systems and their effects on things like accident rates? The answers to these questions are not obvious. Fundamental values are being negotiated here, about what aspects of technology are important. An intervention that favors certain aspects will look very different than one that favors others. And what new sociomaterial practices would emerge out of these varied interventions? We do not yet know. But as designers of new technological systems, we need to keep these changes to practice in the forefront of our minds. Thinking in this way about driving practice opens up the space for different interventions, besides the obvious technical ones of better sensors, better algorithms, better physical infrastructures. Engineers, in our experience, too easily assume that these alone will make AVs possible, pleasurable, and valuable. But we suggest that much of the problem and promise of automation lies outside the technical frame, in the social realm. Driving is a cultural practice. Mobility is not just about getting from A to B, but about when and how and why one moves. The sociomaterial lens applied here is a call for further engagement with the social and cultural dimensions of transportation systems, as these systems inevitably affect essentially everyone in some way, through direct use or through coexistence in shared space. And these people are remade—human autonomies are remade—by our machine interventions. Not causing accidents is not sufficient. We, both as developers and ethnographers of technology, must attend to the ways that practices will change, and the shifts in the personal and cultural significance of meaningful action that will follow.
Erik Stayton is a PhD student in the Program in History, Anthropology, and Science, Technology and Society at MIT. He investigates human interactions with AI systems, and currently studies the values implicated in the design, regulation, and use of automated vehicle systems. He also interns at the Nissan Research Center.

Melissa Cefkin is Principal Researcher and Senior Manager of the Human Centered Systems group at the Nissan Research Center in Silicon Valley. She has had a long career as an anthropologist in industry, including time at the Institute for Research on Learning, Sapient, and IBM Research.

Jingyi Zhang is a human factors researcher at the Nissan Research Center in Silicon Valley. She has spent years working in the fields of Transportation Safety and Human Factors. She holds an M.S. in Industrial Engineering & Operations Research from University of Massachusetts, Amherst.

NOTES

Acknowledgments – The research that produced this paper was performed while Erik Stayton was a summer intern at the Nissan Research Center in Silicon Valley, and is not part of his research at MIT. We offer sincere thanks to all those who read and commented on drafts of this paper, including our anonymous reviewers. We wish to give particular thanks to Graham Jones and Crystal Lee, and our session curators Jamie Sherman and Tiffany Romain, for their detailed edits and suggestions. We also appreciate the time and energy of all our participants and interviewees, including those who tried but were unable to complete the simulator test. Sorry if we made you feel ill!

1. The authors recognize that autonomy is a fraught term, often used very loosely in talking about robotic systems. In prior work, Stayton has preferred to use “automated,” which does not imply the same complete disjunction from human control. But autonomous (or, colloquially, driverless or self-driving) remains a common way to describe highly automated vehicles. Our point in this paper is not to argue for whether or not it is right to refer to these vehicles as autonomous vehicles. Instead, we take a different approach and ask: since people do apply this term, what does it mean to them when they do?

2. This is not just an attempt to make the point that automation is always partial. Nor do we wish to sanctify “driving” as some sort of ineffable practice. Our point is simply that automation is not the movement of particular sections of an activity from one bin (the human one) into another (the machine one). It ends up reshaping, even if subtly, the entire activity, and thereby changes its meanings to those who engage in it and interact with it.

REFERENCES CITED

Bainbridge, Lisanne

Bishara, Amahl

Brown, Barry and Eric Laurier

Casner, S. M., E. L. Hutchins and D. Norman
Lave, Jean

Lindley, Joseph, Dhruv Sharma, and Robert Potts

Lutz, Catherine and Anne Lutz Fernandez

Marvin, Carolyn

Michon, J. A.

Mindell, David

Myers, Fred

Nafus, Dawn and Ken Anderson

Nunzi, Michael

Orlikowski, Wanda and Susan Scott

Scott, Susan V. and E. L. Wagner

Sheridan, Thomas

Stayton, Erik.

Suchman, Lucy.

Suchman, Lucy and Brigitte Jordan.

Suchman, Lucy, Jeannette Blomberg and Julian Orr

Turkle, Sherry

Venkataramani, Arvind and Christopher Avery

Verrips, Jojada and Birgit Meyer

Vinkhuyzen, Erik and Melissa Cefkin

Woods, D. D. and Hollnagel, E.