Through the Handoff Lens: Are Autonomous Vehicles No-Win for Users

Jake Goldenfein, Deirdre Mulligan, Helen Nissenbaum[1]

In December 2018, Waymo, the self-driving vehicle subsidiary of Alphabet launched a commercial passenger transport platform called 'Waymo One'.[2] Limited to a group of participants in Waymo's closed testing program, and only in a small geographic area in Phoenix, Arizona, the launch revealed more about the rhetoric of self-driving vehicles than it illuminated the future of transport. Although Waymo's promotional videos had shown passenger vehicles with no human driver in the front seats, the reality was different.[3] Vaulted as the launch of a truly 'driverless', commercial transport (i.e. ride-hailing) system, the Waymo One service still employed specially trained 'drivers', 'safety supervisors', or 'vehicle operators' that travelled with the vehicles. The drivers' presence was framed more as a customer service than a requirement,[4] but it also raised doubt over the technical *possibility* of fully driverless vehicles, and destabilized the terminology behind 'self-driving', 'driverless', or 'autonomous' vehicles. Although Waymo was long thought the clear industry leader with respect to autonomous vehicles, the future of this technology is hardly clear.

---

[1] This work has been generously supported by The Simons Institute for the Theory of Computing at UC Berkeley, and National Science Foundation Grant: 1650589.

[2] Waymo One press release (https://medium.com/waymo/riding-with-waymo-one-today-9ac8164c5c0e)

[3] Whatever degree of fully driverless testing is occurring is likely a tiny fraction of on-road testing occurring, and may have halted testing fully driverless vehicles entirely. https://arstechnica.com/cars/2018/12/waymos-lame-public-driverless-launch-not-driverless-and-barely-public/

[4] https://www.theverge.com/2018/12/5/18126103/waymo-one-self-driving-taxi-service-ride-safety-alphabet-cost-app

There's no need to rehearse the various benefits autonomous vehicles are claimed to bring.[5] Some of these claims seem sensible and plausible, and some purely rhetorical. Beyond broader social re-arrangements however, different models of autonomous vehicles also entail different configurations for vehicle 'users' or occupants with implications for *human* values. Just as the metaphors we use to describe autonomous vehicles foregrounds and occludes specific values,[6] the terms we use to describe the humans who occupy "driverless" cars carry moral weight. If the car is driverless, then what do we call the human passenger, and what comes with that name? If a car is autonomous, do we need to re-calculate the effects on human autonomy? If vehicles are connected, then how should we process the effects of increased data transmission and surveillance? Further, a functionally equivalent system of automotive vehicle transport delivered through driverless vehicles involves an entire world of reorganization, not simply in terms of spatial rearrangement, or business models, but also in terms of politics and values.

Reckoning with the implications of these reconfigurations requires seeing past terminological obfuscation, beyond the emphasis on discontinuities in the transport experience, and instead

[5] See e.g. UK Department of Transport *The Pathway to Driverless Cars – Summary Report and Action Plan:* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/401562/pathway-driverless-cars-summary.pdf ; Lawrence Burns 'A Vision of our Transport Future' (2013) 497 *Nature* 181; Adriano Alessandrini et al 'Automated Vehicles and the Rethinking of Mobility and Cities' (2015) 5 *Transportation Research Procedia* 145 (https://www.sciencedirect.com/science/article/pii/S2352146515000034) ; Daniel Fagnant and Kara Kockelman 'Preparing a Nation for Autonomous Vehicles: opportunities, barriers and policy recommendations' (2015) 77 *Transportation Research Part A: Policy and Practice* 167 – 181 (https://ac.els-cdn.com/S0965856415000804/1-s2.0-S0965856415000804-main.pdf?_tid=e771638c-23cc-4b3b-93cc-a9fd3ec2ff39&acdnat=1546571852_abd64f4e9599863071703cd8ad73b361).

[6] Meg Leta Jones and Jason Millar, 'Hacking Metaphors in the anticipatory governance of Emerging Technology: The Case of Regulating Robots' in Roger Brownsword, Eloise Scotford, and Karen Yeung (eds) *The Oxford Handbook of Law, Regulation and Technology* (Oxford 2017) ; Wendy Ju's work on metaphors

focusing on values and politics embedded in the various autonomous vehicle trajectories promoted by various parties. While taking in significant research attention to socio-technical reorganization,[7] we focus on the significance of this reorganization and reconfiguration for ethical and political (i.e. societal) values, such as, privacy, autonomy, responsibility, and property. We introduce the *Handoff Model* to complement existing work exploring the technical, legal and policy implications of autonomous vehicles in three ways: 1) rigorously identifying how the function of driving is re-configured to components (human, computational, mechanical, and regulatory) in alternative autonomous driving systems; 2) exploring how these reconfigurations are expressed through the human-machine interface of the vehicles; and 3) interrogating the value propositions captured in these alternative configurations.

To perform this analysis, we have found it useful to create a rough classification of three 'archetypes,' or visions of autonomous vehicle deployments, through which the future of autonomous vehicles is often presented.[8] These archetypes generally comprise a vision or technical, political, commercial and economic requirements presumed for deployment. The first is a model of a fully 'driverless' vehicle that removes the occupant's or user's capacity to control (i.e. *drive*) the car. The defining image is the (now abandoned) Google 'Koala' car that did not feature any human driving controls like a steering wheel within the vehicle. These potential for such vehicles has long been described as transformative as they can travel both occupied and

[7] Eva Fraedrich et al 'Transition Pathways to Fully Automated Driving and its Implications for the Sociotechnical System of Mobility' (2015) 3(11) *European Journal of Futures Research* 3-11 (https://eujournalfuturesresearch.springeropen.com/articles/10.1007/s40309-015-0067-8) ;

[8] A good example is Sven Baiker 'Deployment Scenarios for Vehicles with Higher-Order Automation' in Markus Maurer et al (eds) *Autonomous Driving: Technical, Legal, and Social Aspects* (Springer 2018) where he describes the 'evolutionary' scenario which is continued improvement of ADAS systems, the 'revolutionary' scenario which is the transformation of mobility services with driverless vehicles, and the 'transformative' scenario which includes the creation of integrated public transport style urban solutions. These map relatively clearly onto our description of 'driverless', 'ADAS', and 'connected' models.

unoccupied, meaning they never need to be parked or 'at rest' in the economy.[9] At this stage however, because the cost of such a vehicle is prohibitive, commercial ride-hailing is the only viable business model. That means vehicles are likely owned by large tech-companies like Google and Uber, or perhaps through private investment / equity models that allow individuals to hold shares in a working commercial vehicle.

The second archetype envisions a gradual increase in the degree and competence of automation in privately owned passenger vehicles. This vision puts autonomous vehicles on a spectrum including power steering, cruise control, and automated lane-keeping. Here, automation becomes part of an Advanced Driver Assistance System (ADAS), an approach described as the preferred trajectory of the incumbent automotive industry.[10] While some proponents suggest this approach is safer in the short term, others note it may be more dangerous in the long term for delaying the proliferation of fully driverless vehicles (which are 'inevitably' safer, and will be safer still when we can remove unpredictable and sometime antagonistic human drivers from the road). The ADAS approach retains a human 'driver' (at least some of the time), although reconfigures the locus of control across human and computational elements.

Our third archetype is 'connected cars', sometimes labelled an 'internet of cars'. This model positions vehicles as elements of broader smart-city transport programs. Unlike 'autonomous vehicles', that is, vehicles capable of navigating on the basis of their on-board sensor arrays alone, 'connected vehicles' operate in constant communication with one another as well as with other components of a static infrastructure. This archetype includes a role for the traditional automotive industry, while requiring the involvement of technology and data platforms, very

---

[9] Kevin Spieser et al 'Towards a systematic approach to the design and evaluation of automated mobility-on-demand systems: A Case Study in Singapore' in Gereon Meyer and Sven Beiker (eds) *Road Vehicle Automation* (Springer 2018) (https://link.springer.com/chapter/10.1007/978-3-319-05990-7_20)

[10] Gilles Duranton 'Transitioning to Driverless Cars' (2016) 18 (3) *Cityscape: A Journal of Policy Development and Research* 193 (https://www.huduser.gov/portal/periodicals/cityscpe/vol18num3/ch12.pdf)

likely to be offered by major technology companies in cooperation with cities, states, roads organizations, and other governing bodies. A connected car model may include connected light-posts and roadways, traffic management systems, shared mapping, as well as high levels of connectivity among vehicles. It is also seen as the pathway to more complex driving maneuvers like continuous flow intersections, vehicle 'platooning', and ultra-high-speed travel.

Some authors put these models or archetypes – *full-driverless, driver-assist, and connected-cars* -- on an historical trajectory wherein driver assist eventually succumbs to full automation, and all private ownership is replaced by mobility on demand services.[11] But the reality is more complex, the players are more tangled and integrated, and a path forward is unclear. Indeed, the possibility that there will ever be truly driverless (i.e. *no human* driver) vehicles, capable of operating in all contexts is in serious doubt. Whether in the car or operating remotely, there may always be a human driver or controller somewhere in the operation of the vehicle, if not for technical or safety reasons, then perhaps simply for liability reasons.[12] Although we acknowledge this complexity and a landscape that is constantly shifting, overlapping, contesting, and rearranging, It remains useful to explore the archetypes as distinctive abstractions for the purpose of explaining how each disturbs the existing politics of transport in different ways. For us, the archetypes are a means of exploring deep connections between different architectural designs for achieving ostensibly equivalent functional purposes, on the one hand, and respective political and value proposition for users and society, on the other. To analyze these correspondences, we apply the 'handoff' lens, which reveals key differences in the configurations of system

---

[11] Eva Fraedrich et al 'Transition Pathways to Fully Automated Driving and its Implications for the Sociotechnical System of Mobility' (2015) 3(11) *European Journal of Futures Research* 3-11 (https://eujournalfuturesresearch.springeropen.com/articles/10.1007/s40309-015-0067-8)

[12] See e.g. Dan Hill 'How Should Humans Really be "in the Loop" in cities under autonomous mobility (https://medium.com/butwhatwasthequestion/how-should-humans-really-be-in-the-loop-in-cities-under-autonomous-mobility-d13c04f70bba); and Madeline Claire Elish 'When Your Self Driving Car Crashes, You Could Still be the One who gets sued' (https://qz.com/461905/when-your-self-driving-car-crashes-you-could-still-be-the-one-who-gets-sued/)

components, in turn altering the values embedded in respective systems, not just the capacity to travel. Although the handoff lens is put forth as a means to reveal something about the archetypes that might have been missed without it, we acknowledge that this exercise is as much a test of the concept's usefulness and analytic power. Before proceeding with the analysis, below, we introduce key elements of the *handoff* framework to our readers.

The Handoff Lens

Inspired by claims about computational systems able to take over tasks previously performed by humans, especially, tasks thought to require human intelligence, the concept of *handoff* provides a lens through which to scrutinize them. Prior to the wave of present day attention on automation and artificial intelligence, the delegation of function from human to machine and from machines of one type to machines of a different type, proceeded almost without public awareness. AI has inspired widespread attention and anxiety -- machines that can label ("recognize") images, process ("understand") and produce ("speak") natural language, and even anticipate what we will say and do. In all these cases, where function shifts from one type of actor to another, and people are inclined to say that the second is performing *the same* function as the first, (same function, different actor) we see the need for a more detailed critical analysis. The handoff lens draws attention to assertions such as those, for example, that automated call-answering systems perform the same function as human receptionists, that RFID-based systems collecting road tolls or computational systems distinguishing benign from cancerous skin moles are performing *the same* function as humans or mechanical counterparts. Rather than important questions such as the quality of performance, efficiency, or effects on labor, or other questions about functional performance, the critical eye of handoff instead, directs our attention towards the backdrop of ethical and political values embodied by respective systems – the systems before and after functional handoff. It decomposes the "how" of the function to understand how it is different and what that means for values. It opens our view to not only what might be the same but what may have changed in the reconfiguration of function across component actors. Before proceeding with our analysis of AVS, a brief introduction to the handoff model follows.

The purview of our Handoff analytic model are complex systems comprising diverse functional components. Because the nature of such systems can be varied, incorporating physical mechanisms, embodied computational subsystems, and even humans as the unit of analysis, more precisely is a socio-technical system, which others have theorized and we take to be noncontroversial. The socio-technical is what we mean to cover in the balance of this article, though we mostly revert to the term system to mark a degree of abstraction in our analysis. Abstractly conceived, a system may be defined in terms of its function. This function may be achieved by a system's component parts, themselves conceived as systems, which in turn comprise subsystems, or components. In other words, the model assumes that system and component (or subsystem) are relative terms whose application signals the focus of analysis rather than an ontological statement.[13] In an analogy, someone may think of the human body as a system and the organs as component parts; but for the cardiologist, the heart is the system of interest and the chambers, valves, arteries, etc. as its components, and so on. In another example, the system of a conventional automobile performing the driving function comprises vehicle plus human driver; in turn, each of these components many be analyzed further – the vehicle composed of various subsystems, such as braking, safety, ignition, and so on.

As noted, systems perform functions, and it is the redistribution of these functions that interests us. What that function is, in general terms, is answered by the question, "what does this system

[13] Terminology presented a dilemma. We use the generic term **component** to apply to both human and non-human parts of the **sociotechnical system**. While the term component does not naturally apply to human **actors** for our purposes it is important to be able to refer in like manner to human and non-human components of a system. Actor-Network-Theory [[ref]], which most certainly has influenced us, came up with *actant* as a way out of the dilemma but our preference is to not adopt theoretical jargon, which can be off-putting for general readers. Going forward, we will mostly stick with the term **component** and sometimes will revert to actor, or subsystem. In addition to human actors and physical objects that can be or constitute system components, we allow for the possibility of groups and institutions as components.

do?" Components also perform functions, similarly answering the question, "what does it do," expecting that the answer will address how the component function contributes to the overall function of the system. Loosely speaking, a system's function can be described at different levels of abstraction: up a level, referring to goals, purposes, or even values; down a level, the way a designer or engineer might explain *how* it does what it does. But for our purposes it is necessary to more precisely define and separate out goals, purposes, function, and how (implementation), for the interplay between these levels due to handoffs of functions illuminates the impact on values of these reconfigurations.

This more granular exploration of a function's meaning at multiple levels is useful for different reasons. At the lower level of "how," an analyst needs to explain how the components work to produce overall system function, what drives their respective actions (or motions) and how they interact with one another. To this end, we introduce the idea of components **acting on** or **engaging** other components. Take a traditional automobile (vehicle+driver system) driving on a road. Darkness falls and a human driver (component) pushes a button, which in turn causes the car headlights to come on. Thus, the different components act on each other to produce an outcome—fulfill function, "turn on the headlights." The driver decides, pushes a button; the button causes the headlights to flash on. This trivial case serves to introduce one further concept, namely, the **mode** of acting on or engaging. The idea that there are different modes of acting-on originates from insights garnered from the works of disparate scholars and theorists, who have expounded on the impacts of technology on the lives of individuals and societies. While some, such as Larry Lessig, sought to elide the differences by identifying all as modes of regulation, others have argued that different modes make a difference.[14] For the moment, mainly, we see value in merely highlighting differences without, yet, developing their significance.

[14] See e.g. Roger Brownsword, *Lost in Translation: Legality, Regulatory Margins, and Technological Management*, (2011) 26 BERKELEY TECH. L.J. 1321; *See* Karen Yeung, *The Forms and Limits of Choice Architecture as a Tool of Government*, (2016) 38 LAW & POLICYY 186

One familiar **mode** of acting on another object is physical force. In this simple form, one physically embodied actor may cause an outcome in another.[15] The outcome may be achieved either by forcing action or preventing action. The human actor pushes a button and sets off a causal chain of action resulting in the headlights flashing on. Physical ("material") causation, or—one could say—"brute force" may operate in multiple, different ways, for example, a physical component (or set of objects) may act on another component by constraining its range of action (e.g. a safety overlock) without necessarily causing a particular outcome; or, there could be far more complex causal interdependencies, as when numerous components function together to produce a complex configuration of outcomes on other components, and so on.

A different mode of acting on – one might say, more subtle -- is **affordance**. As defined by the cognitive psychologist J.J. Gibson affordances are relational properties of things in the environment whose meaning or significance is derived from their service to a given agent's needs or capabilities. When saying that something is nourishing, or is a tool, or is secure cover, these observed properties must be understood relation to actors of particular shapes, sizes, abilities, and needs. As adapted and widely popularized by Donald Norman, designers can and should take advantage of affordances in order to create artifacts that are understood by users and elicit desired behaviors from them for successful usage. According to Norman, an object's affordances suggest uses to us by triggering human cognitive and perceptual capacities. Good designers of either material objects, such as doors and switches are able to elicit correct usage, or desired reactions by effectively exploiting human actors' tendencies to respond to cues of various kinds – logical, cultural, semantic, and feedback – in systematic ways. The same ideas extend to digital objects, such as websites and appliance controls for example a social media site that displays its users' information creates the possibility for its repurposing, it may enhance that possibility by supporting an application programming interface (API) that eases data extraction, or diminish that possibility through technical or legal rules (for example a prohibition on scraping) that

[15] (Remaining at the intuitive level, for the moment, we must look past the fact that there is nothing simple about causation, as Aristotle well demonstrated!)

discourage such extraction. For purposes of our analysis, affordances constitute a mode of acting-on that can serve designers of vehicle interfaces seeking to convey to users (drivers, passengers) what range of actions are possible and desirable in relation to the vehicle's operation. Unlike physical force, affordances are perceived and processed by users (human users, in this case) who act – often strategically – accordingly.

Returning to our mini case of a driver switching on headlights, we observe that the human actor physically exerts force on the button thereby initiating a causal chain resulting in the lights flashing on. When, further, we ask what made the human push the button, there may be various answers. One of them points to the interface, which has successfully exploited the affordance of "push-ability" in its design of the button in question. Other answers illustrate additional modes of acting-on.

Another plausible answer may cite purpose: the driver pushed the button because visibility was poor, night had fallen, it had started raining. A different answer may cite obedience to the law, which prescribes switching on headlights under certain conditions. Each of these cases reports on an intentional action taken by the driver, on a decision to switch on headlights. Although the model does not view the law, or light levels, or the miserable weather as components -- they do not *act on* the driver -- they surely inform the driver's action. The driver chooses to act (pushes the button) after having identified conditions or pertinent rules, interpreted them, and decided to act accordingly. The driver (user), as it were, as a free agent, is prime mover causing the headlights to flash on by pushing a button.

Now, imagine a subsequent model of the vehicle, in which the operation of headlights is automated via a small computer embedded within the vehicle controls. In this case, under the appropriate exterior circumstances, the algorithm's expression in lines of software code, implemented in an embodied computer, **acts on** relevant components resulting in the lights flashing on. The software code (and more abstractly, the algorithm) operate like legal rules. The model does not reify them as component actors; instead, their informational content, expressed as coded instructions, is embodied in material, electronic computers, which act on other system components, and so on. Without delving into metaphysical questions about the nature of free agency, the handoff model asserts a difference between automated headlight switches and human operated switches by noting that in acting on the coded rules, the material computer is not a prime mover but has been acted on by those responsible for the code, and preceding even, a particular trigger external to the system. Later in the article, the implications of some of our choices will become clear.

Defining Handoff: Given progressive, or competing versions of a system (S1, S2) in which a particular system function (F) shifts from one type of actor (A) in S1, to another (B) in S2, we say that *F was handed off from A to B*. In the headlights cases, for example, we could say that the function of switching on the light had been handed off from human actor to computer controller. As a matter of fact, well before the general hype over autonomous vehicles, a progressive shifting, or handing off of certain functions from human controller (driver) to vehicle components has been underway for some time. For example, the Electronic Stability Control System (ESC system) wrests control from the driver when it senses rear wheel activity that indicates "spinning out" (loss of directional stability) or front wheel activity that indicates "plowing out" (loss of directional control).[16] In these instances the car seizes control from the human driver, and takes responsibility for bringing the car back under control. The cars lateral acceleration and yaw rate, captured by onboard sensors, is compared to the driver's intended heading deduced from speed and steering angle measurements. If they are inconsistent the ESC system takes over.[17] The ability

[16] Federal Motor Vehicle Safety Standards; Electronic Stability Control Systems; Control and Displays, 72 Fed. Reg. 17,236 (Apr. 6, 2007) (final rule) (codified at 49 C.F.R. pt. 571). (requiring ESC systems on passenger cars, multipurpose passenger vehicles, trucks, and buses with a gross vehicle weight rating of 10,000 pounds or less.)

[17] "the ESC system measures the car's speed and its lateral acceleration, it can compute the radius of the circle. Since it then has the radius of the circle and the car's speed, the ESC system can compute the correct yaw rate for a car following the path. Of course, the system includes a yaw rate sensor, and it compares the actual measured yaw rate of the car to that computed for the path the car is following. If

to independently adjust brake torque on each wheel allows the ESC system to use uneven brake force—rather than steering input—to reestablish a yaw appropriate to the intended heading of the driver. Similarly, protections against dangerous secondary impact injuries—driver and passenger collisions with the inside of cars caused by a crash—moved from reliance on human actors to engage safety belts to foisting protection upon initially the driver (but over time front and backseat passengers in many models) through the introduction of passive restraints, such as airbags. These handoffs while aimed at improving safety have met with a range of value-based resistance, presaging the need for a model such as ours to identify and manage values during functional re-distributions.[18]

Finally, while not always illuminating, more often than not considering the impetus (what we refer to as **Trigger**) for two competing or sequential handoff configurations highlights specific values that may be both motivating the reconfiguration or implicated by it. For example, historian Peter Norton argues that during the 1960's traffic safety shifted from a paradigm of *control* to

---

the computed and measured yaw rates begin to diverge as the car that is trying to follow the circle speeds up, it means the driver is beginning to lose control, even if the driver cannot yet sense it." Id. at 34.

[18] See Wetmore, Jameson M. "Redefining risks and redistributing responsibilities: Building networks to increase automobile safety." *Science, Technology, & Human Values* 29.3 (2004): 377-405 (describing automaker resistance to passive restraints based on the belief that explicitly removing the driver and passenger from enacting safety (putting on a belt) *implicitly* threw open questions about who would be held responsible and ultimately liable for injuries and fatalities "[W]hile automakers were wary of accepting the responsibility of making vehicles more crashworthy in the 1960s, they were even more frightened of taking on the liability that would accompany their involvement in an air bag strategy.") at 390; and Wetmore, Jameson M. "Delegating to the automobile: experimenting with automotive restraints in the 1970s." *Technology and culture* 56.2 (2015): 440-463 quoting retired GM president "I do feel that when you have a passive system the responsibility lies more with the manufacturer and the service station that takes care of the system." at 447 citing Edward Cole, Presentation at "Public Hearing on FMVSS 208," Washington, DC, 27 April 1977, in NHTSA Docket 74-14 Notice 8 (Number 58), 20–27, at 26, in DOTDA.

one of *crashworthiness*. The control paradigm centered on preventing accidents through expert control delivered through engineers designing safer roads, expertly educated drivers and pedestrians, and targeted enforcement to prevent reckless driving. While the control paradigm concentrated on reducing the safety risks posed by drivers, pedestrians and roads, the *crashworthiness* paradigm, spurred by rising fatalities, ushered in a focus on reducing the damage of inevitable collisions and focused on reducing the damage caused by vehicle occupants colliding with the interior of the automobile. This paradigm put the design of automobiles in service of safety. The shift from "crash avoidance" to "crashworthiness" was part of an effort by safety advocates who sought to place greater responsibility for safety on the automobile industry and the automobiles they produced. The shift in paradigm ushered in (Triggered) the move from active (seat belts) to passive restraints. The explicit aim of safety advocates was to reallocate responsibility from the public whose attitudes and behavior had proved quite resiliently ill-suited to preventing secondary impacts, to technology that could compensate for human failings. Passive restraints were viewed as a response to the moral failings of drivers. Airbags and other passive restraints were to displace the immoral human actors. While airbags would not become standard until the 90's this 1960's paradigm shift triggered initial pressure for their development. The shift in responsibility for function to the automobile carried moral weight in terms of the task itself as well as the responsibility and potential liability for injuries.

Autonomous Vehicle Futures

Applying the handoff lens to autonomous vehicles encourages us to move beyond the idea of a gradual, linear transition of control from human to computational elements in service of producing a functionally equivalent artefact (a car) that performs a functionally equivalent task (driving). The SAE Standards for vehicle automation, for instance, describe this trajectory from traditional human control, through partial automation like automated lane keeping and braking (level 2), to vehicles capable of operating without humans under constrained conditions (level 4), to a mythical (or military) fully automated vehicle capable of operating independently under any and all conditions (level 5). This step-wise model encourages us to think about vehicles as discrete objects, whose development can be mapped and tracked with SAE levels, along a single trajectory of the degree to which they are controlled by computational systems, as technical and regulatory hurdles are gradually overcome. But the reality is different. Each step within these standards, along with tracking a level of automation, in fact embodies a different system of actors, components and agents; a different political agenda promoted by different stakeholders; and a different configuration of values for the users of those vehicles. Applying the handoff analytic to autonomous vehicles is useful here because it directs us towards thinking about these vehicles and the function they perform not as objects and tasks, but as complex systems of digital and physical infrastructure, and new and transformed components, with consequences for politics and ethics. Accordingly, we address the technical formations of autonomous vehicles not in terms of step-wise progress in automation, but according to systemic models or archetypes of deployment that represent different systemic visions of the future of autonomous driving.

There is no shortage of work on societal implications (technical, legal, ethical and social implications) of autonomous vehicles, by proponents as well as critics.[19] Proponents are quick to point out the societal costs of existing automotive structures and how autonomous vehicles will mitigate them. To name a few: the promise of reduced environmental impact -- not simply in terms of carbon pollution but also reduced needs for mining of materials and oil and associated geopolitical problems. Traditional cars impose negative consequences on cities through congestion and the privileging of roads over pedestrian centered travel; they are notoriously unsafe. The Rand Corporation has calculated that non-automated vehicles impose an externality cost of $1300USD of congestion on other road users per 10,000 miles travelled.[20] When measuring the value propositions of autonomous vehicles then, it is then important to consider forgone environmental benefits (like from vehicle ‘platooning’, and decreased travel times,[21] or

[19] See e.g. Markus Maurer et al (eds) *Autonomous Driving: Technical, Legal and Social Aspects* (Springer 2018)

[20] Rand Corporation *Autonomous Vehicle Technology: A Guide for Policy Makers* 68, xiv

[21] Ibid.

from decreasing the weight of vehicles that have increased over the years to improve safety performance in collisions). Along similar lines, Rand reports on research showing that if vehicles implemented even today's driver assist technologies, up to one third of collisions and fatalities could be avoided.[22]

Advocates of driverless vehicles acknowledge a loss of the occasional fun of driving and a few low skilled jobs but deem the trade-off against gains in safety, environmental, economic and efficiency – a small price to pay.[23] Further, the impact on labor systems rendered by autonomous vehicles can be absorbed by the new economy. Nostalgia for the experience of car ownership and driving is pitched as holding back the future.

For critics, whatever picture might be painted by utopian visions of autonomous vehicles, there are always things strategically omitted. While autonomous vehicles might drastically improve the use of urban space by making vehicle parking obsolete, they might also transform city streets into human exclusion zones, too dangerous or inefficient to share. While autonomous vehicles might make car-sharing a feasible and affordable travel model, the underlying political economy might yield a small cadre of high-tech leaders who are able to set terms and prices, impose greater costs on users, without the need to employ human drivers. While autonomous vehicles might dramatically reduce road accidents, they might also undermine public mass transit systems which are historically safer, more economical, and environmentally friendly.

A transition to driverless vehicles calls attention to values shifts that may not be immediately evident For instance, we typically interpret vehicles as 'private' places, even when in public space. That intuition has been confirmed to a degree by the US Supreme Court in *US v Jones*,

---

[22] Ibid.

[23] Gilles Duranton, 'Transitioning to Driverless Cars' (2016) 18 (3) *Cityscape: A Journal of Policy Development and Research* 193 (https://www.huduser.gov/portal/periodicals/cityscpe/vol18num3/ch12.pdf)

when it found the private property dimensions of a vehicle make it a protected space.[24] And while there are likely better arguments for why location privacy should be 'private', the property arrangements associated with vehicles remain critical in understanding the configuration of other values. Further, while proponents may disparage the sentimental attachment to driving, it's hard to deny the bond that exists between people and their cars. For many, cars are an expression of identity and personhood, an opening-up of geographical space, and a freedom to travel not subject to the vicissitudes or whims of others. The metaphor of 'being behind the wheel' signifies control, having direction, and the capacity to express intention. Thus, it is no exaggeration to say that for many people, owning and driving their vehicles constitutes the exercise of their autonomy. Autonomy, not simply in the ability to carry out an intention to drive and exercise control over a vehicle; it also means responsibility as owner and driver. The links of ownership, autonomy, and responsibility are reflected in the emergence of social institutions, such as licensing of human owners and human drivers to ensures their capacity for proper operation of a vehicle, identity for the sake of policing, and as a locus of responsibility for the carriage of a vehicle.

It is not our intention to expand and extend the scope of work on societal implications of human-driven and autonomous vehicles, which we have only superficially scanned; in our view, this work is comprehensive and thorough. Rather, the contribution of *handoff* is a richer account of cause and effect. Frequently, too much is elided when proponents and critics assert that human-driven or driverless vehicles will that this or that outcome, respectively. There are important parts of the account that are being omitted and that should be part of the story when we weigh up the consequences. One step we take to enrich the story is defining the archetypes, which immediately complicate any claims about the expected impacts of driverless vehicles. It forces an analyst to specify which archetype, or model, and thus also which commercial and political configuration. The handoff lens is a second dimension as it allows us to look at these archetypes

[24] *US v Jones* 565 U.S. 400 (2012); Michael Zimmer, 'Surveillance, Privacy, and the Ethics of Vehicle Safety Communications Technologies' (2005) 7(4) *Ethics and Information Technology* 201.

and register the structural, ethical, and political features, and their consequences for users and society, not because of a change abstractly conceived but because of complex systemic shifts back to which these features can be traced. This, at least, is the hope.

A caveat. As we make the case for a finer grained analysis that takes account of the diverse futures (we are saying, three) we are not attempting an account of societal implications, as a whole. Instead, we focus primarily on societal values, and within this class, for the most part, we limit the to property, privacy, and responsibility. With a focus on these, we use the handoff lens, scrutinize the details of socio-technical re-arrangements associated with those different archetypes for autonomous driving futures.

Interfaces

In using the handoff lens to describe the political and ethical consequences of different autonomous vehicle archetypes, we pay particular attention to the human-machine interface. What in a traditional car might be the controls (steering wheel, accelerator, brakes, transmission and dashboard, and perhaps also the interface to on-board computer, stereo and information environment), becomes, in autonomous vehicles, a highly contested and contingent technical arrangement. We see this interface as critically determinative of the politics and value propositions of autonomous vehicles. For instance, the interface determines the active role of the human component (labels of driver, passenger, occupant reflect the contingent and varying nature of the so-called “human in the loop”). The interface specifies the ‘mode’ of acting – be it through direct control inputs, remote control inputs, or coded responses to programs and sensors. The interface also facilitates communications between vehicle, broader communication and control system, and driver, as well as enables the ‘driver’ or control system to act on the vehicle. It determines or guides what information about vehicle operation (and potentially also data transmission for instance in a privacy policy) they receive, and whether that person is conceived of, and experiences themselves as, a ‘user’, ‘driver’, ‘passenger’, or something else. The Boeing 737 Max interface affords a tragic example. As we are learning, the 737 Max aircraft that crashed in both Indonesia and Ethiopia lacked specific safety features that communicated 'angle of attack' sensor and sensor 'disagreement' readings to the cockpit.[25] These indicators are connected to what are believed to be the sensors that failed in both instances. Beyond the failed technical transition of control to the computational system, these absent safety features show the broader political, and economic role in both the configuration of the interface, the role of the pilot, and their consequences for the crashes. That Boeing, the airlines, and regulators considered these safety features inessential, and only available with additional cost, demonstrates how the regulatory and economic imperatives conceptualize the pilot as superfluous to the control of the aircraft in this specific, and evidently critical, way.

With respect to 'driverless cars' however, many have noted that the term 'driver' becomes increasingly blurry. Relationships to vehicle controls and vehicle control may have little to do with occupancy, and where spatial arrangements that historically connote driving (front left seat in the U.S.) provide no affordances to drive. New proposed categories for human components in the reconfigured driving environment, like 'test drivers', and 'vehicle users', reflect these discontinuities with past practice.[26] For instance, the UK Pathways report identifies the possibility of non-occupant 'vehicle users' which includes individuals who are controlling the destination of a vehicle remotely without actually travelling in the car. The interface also influences (though does not determine) the responsibility and liability arrangements of autonomous vehicle use.[27]

---

[25] Gaby Del Valle 'Boeing and Other Companies put Safety at a Premium' < https://www.vox.com/the-goods/2019/3/22/18277694/boeing-737-max-ethiopian-airlines-lion-air-safety>

[26] UK Department of Transport, The Pathway to Driverless Cars: Summary Report and Action Plan < https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/401562/pathway-driverless-cars-summary.pdf>

[27] Mark Geistfeld, 'A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation' (2017) 105 *California Law Review* 1611; The question of responsibility is discussed here primarily in philosophical rather than legal terms. Insurance and legal liability rules are designed to apportion risk and fault according to a specific economic or behavioral calculus. The novel questions around both civil and criminal legal responsibility has been subject to a great deal of insightful

This is because it affects who might be a driver or in control of a vehicle (which is important in liability regimes), through constraining by affordance what parties or components are capable of performing controlling actions. In a similar way, the interface thus implicates the autonomy of users because it interrupts the capacity to express one's intention to control the car. Depending on the information flow between the vehicle to the user through the interface, it may also interrupt the capacity to generate a relevant intention to control the vehicle. This happens at one level by removing a steering wheel, and at another level by the vehicle's route to its destination being determined algorithmically. Privacy issues are also reflected in the arrangement of the interface as it is central to the information transmission that occurs, that is, it specifies what entity receives what information and when. Interfaces also reflect (rather than affect) the property ownership models of autonomous vehicles, in that, they not only define what specific control inputs are possible, but also embody broader questions of exclusion, license to use, and purpose.

In our view, the interface is not therefore simply a functional artefact, it is expressive of (while also sometimes dissimulating) the political arrangements embedded in the vehicle, and therefore the 'agenda' animating the relationship between the various component-actors involved, as well as the mode of how those actors engage each other.

---

analysis. While the question of human responsibility is an element of that calculus, the necessity of finding fault is also often avoided in liability systems through the introduction of no-fault or strict (product) liability systems. On the other hand, these systems often work in concert with negligence actions seeking to apportion fault to an appropriate party. The technical complexity of control hand-overs suggest the apportioning of legal liability between a vehicle manufacturer and human driver according to standards of performance (or negligence) may be difficult (unless there are extreme examples of negligence or product failure). This may result in product liability approaches, single insurance schemes where a single insurer covers both the driver and the manufacturer, or even no-fault compensation schemes. Ascertaining an appropriate liability regime is not the goal of this analysis however. Instead, we explore the question of how the information interface may result in defining the experience of responsibility for the operation of a vehicle.

Archetype 1 – 'Fully Driverless Vehicle'

On one account, the fully driverless vision of autonomous transport is seen as the best way to capture the economic and social benefits of autonomous vehicles. With a fully driverless car, 'passengers' or 'users' would be able to use travel time for other activities with no obligation to pay attention to the road or vehicle controls. Such vehicles typically require legislative fiat, which only a few jurisdictions have provided so far.[28] Several laws however, are being debated (or at least proposed) that would enable the use and sale of vehicles without the traditional vehicle control interface of steering wheels and pedals. The critical change to the vehicle interface in this configuration is the absence of direct controls and the introduction of rich systems of information exchange between human occupants ('users' or 'passengers' and the entities controlling the carriage of the vehicles (here 'operators'). As mentioned above, the cost of these vehicles indicates they are likely, in the near future at least, to be used in 'mobility services' whether privately, publicly or communally operated.

For instance, Waymo, in producing a commercial ride hailing service, uses a specifically designed Chrysler Pacifica mini-van, that, while still having traditional vehicle controls, is designed to reduce the user's direct control over the vehicle. (Indeed, the service was initially slated as having no driver in the front seat). In these vehicles, there is a digital screen for each back-seat passenger providing a 'god-view' (i.e. top down view with the vehicle in the center) real-time map showing the environment as detected in relatively low resolution, coupled with pulsing higher resolution images (of still relatively indecipherable dot representations of the physical environment). There is also a mechanical interface for back seat occupants, with three buttons - 'start ride', 'pull over', and 'help' - perhaps to give occupants a sense of ultimate control in terms of the capacity to

[28] See e.g. AV START Act

override the digital interface, although ‘pull over’ does not function as an emergency or panic button.

These in-car controls are coupled with the Waymo One App for smartphones, which operates in a similar manner to other commercial ride-hailing services. Destinations are input, prices are agreed, and feedback is provided following the common model of star ratings and selected verbal responses such as ‘good route choice’, and ‘clean car’ (although ‘friendly driver’ is probably no longer an option). All three interfaces, in-seat, mechanical, and app, give a ‘support’ option, where you can contact a Waymo employee, likely situated in a control or service center, who can offer guidance on *using* the vehicle, for instance instructing users on how to change destination.

These vehicles are owned and insured by Waymo, who calls the driving control system the ‘experienced driver’.[29] And while the software and hardware constellation is highly determinative of control over these vehicles, it is also possible that a company like Waymo would build an autonomous driving ‘platform’ that could be installed in other vehicles, perhaps shifting the identity of the ‘driver’ into the ‘software’ only, or a different combination of actors or components. Waymo One also charges prices similar to existing ride-hailing services like Lyft and Uber. While the prices may be set to incentivize use, ultimately the commercial orientation means that prices will be set to achieve maximum possible profit. If that remains the case, it is unclear what benefit this model of autonomous vehicle offers as a mobility system, other than enabling the operator to charge fairs without paying drivers or garaging a car.

Waymo is not, of course, the only entity exploring totally driverless cars or autonomous vehicles for ride-hailing. Uber has been experimenting with these vehicles and has partnered with Toyota. Lyft also has a partnership with tech company Aptiv, which have a fleet of BMW cars (that also include manual controls, a ‘safety driver’, and a display showing an approximation of the vehicle’s

[29] Waymo One press release (https://medium.com/waymo/riding-with-waymo-one-today-9ac8164c5c0e)

sensors), operating on a small number of 'routes'. Lyft also received a $500m investment from General Motors in 2017, indicating the possibility that General Motors may manufacture vehicles for autonomous ride-hailing, or that General Motors is becoming a ride-hailing business. General Motors has also acquired a 'driverless car' company 'Cruise' which is, for instance, building a driverless vehicle for Honda.[30] The cars that 'Cruise' seeks to manufacture have no in-cabin driver controls. Clearly, industry operators are adopting new, different, and complex positions in the autonomous vehicle eco-system, which involves new roles for manufacturers, service providers, and platform operators. These relationships between tech platforms, mobility services and manufacturers demonstrates the likely trajectory of commercial uptake of these vehicles, and the use to which they will be put. In other words, the high-level business arrangements become a lens onto the new political reality of such vehicles.

Managing the control and 'driving' of vehicles however, may require new actors or components beyond the software systems produced by tech and mobility companies. For instance, manufacturers such as Nissan, while also pursuing the manufacture of vehicles with no in-cabin controls, has introduced new components, interfaces, and modes of acting into their control systems. Their 'Seamless Autonomous Mobility' system uses a central control room with human 'mobility managers' who can intervene in vehicle control when facing complex obstacles.[31] This relocates an element of the driving interface to a remote location, and to a remote person, who makes decisions about vehicle operation. From the promotional material available however, this does not appear to mimic a traditional vehicle interface – i.e. it is not a vehicle driving simulator – but rather is a mapping system, where new routes can be plotted, and then delivered to the vehicle for execution through its own driving software. For instance, mobility managers draw a path on a digital map using a regular computer interface (i.e. a mouse), this acts on the driving system by determining the route, and the vehicle uses its autonomous driving systems to execute those directions. Mobility managers are typically engaged when a vehicle encounters an obstacle

[30] Andrew Hawkins, 'GM's Cruise will get $2.75 billion from Honda to build a new self-driving car' <https://www.theverge.com/2018/10/3/17931786/gm-cruise-honda-investment-self-driving-car>.

[31] Lawrie Jones, 'Driverless Cars: When and Where' (2017) March *Engineering and Technology Magazine*

that it cannot negotiate rather than an emergency situation. Control transitions to the 'mobility manager' thus occur typically when a vehicle is stationary, producing a communications dynamic that engages the mobility manager only at particular, necessary, times without dynamic control takeovers.

Other iterations of displaced drivers however, are different. Companies like 'Phantom Auto', for instance, enable autonomous vehicles controlled by humans in remote locations using vehicle simulators.[32] These people are not billed as 'drivers' either, but rather 'teleoperators'. (Although the job position advertised on their website is for a 'Class A Driver & Remote Operator'.)[33] In this situation, the teleoperator actually controls the vehicle rather than simply re-routing it. The difference in remote approaches highlight the questions of what constitutes 'driving' in these systems, how control 'components' are distributed, what the identities of these new components may be, how they are connected through distributed interfaces, and the different modes of acting those interfaces enable. For instance, when using teleoperators other components to the system becomes critical. That is, a key requirement of the 'Phantom Auto' system (unlike the Nissan 'SAM' system) is zero-latency video transmission, which could also provide a huge financial boon for telecommunications providers, over whose networks that data will flow, who have their own particular political stakes and incentives. Whatever the specific arrangement, it appears as if most autonomous vehicle providers are exploring similar remote driving capabilities.[34] To a certain degree, this looks like an outsourcing of the driving task. Vehicles still have human a 'controller', in the short term perhaps situated in vehicle, but in the long term likely remotely located. This shifts our understanding of 'autonomous driving' towards

[32] Alex Davies, 'Self Driving Cars have a Secret Weapon: Remote Control' 1 February 2018 *Wired* (https://www.wired.com/story/phantom-teleops/)

[33] https://phantom.auto/careers/?gh_jid=4073694002

[34] See e.g. Ford 'remote repositioning' experiment: <https://media.ford.com/content/fordmedia/fna/us/en/news/2015/01/06/mobility-experiment-remote-repositioning-atlanta.html>

metaphors of 'drone driving', with various distributed components, acting with their own specific role and interest, all contributing to overall system function.

With respect to values, privacy questions in this 'archetype' are inextricably bound up with the commercial or business model of the vehicle. For instance, in a ride-hailing service, there is little question that data produced through interaction with phone apps, or even in-cabin video-recording, is likely to be collected, used for internal optimization purposes, and potentially transmitted to other entities for the sake of behavioral advertising and profiling. Much like existing ride-hailing vehicles and taxis, these cars will record video footage inside that cabin. That internal-cabin footage is not likely to be used for the sake of driver safety, but perhaps for the sake of evaluating the experience of users, or deterring property damage and other forms of undesirable behavior.

In a commercial vehicle, vehicle data will likely be monitored in real-time by control center operators. In a privately operated 'fully driverless' vehicle, this may be simply recorded. But both private and commercial models likely provoke another emergent privacy issue, unique to this configuration of components and total system function, emanating from the possibility of remote drivers and teleoperators with access to vehicle sensor data of the vehicle. This is a new type of information flow. Indeed, if the legal environment requires constant monitoring of roadways, then teleoperators may become shadow drivers and passengers that 'tele-occupy' the same space as the vehicle 'user'.

The responsibilities of the various component actors in these driving systems may also require re-thinking. While a human occupant should perhaps not be considered a 'driver' in any meaningful sense, they may have different obligations as a 'passenger', 'user' or 'occupant', as may the other controlling components. If the vehicle interface *does* include a mechanical system for directing the vehicle to 'pull over' (note that the 'pull over' button in the present Waymo One car is actually just a way-pointing mechanism), does that place an obligation on the passenger to supervise the driver to the extent that if the vehicle is behaving absurdly, there is a responsibility to stop it? At what point does it become unreasonable for an occupant of an autonomous vehicle to not command the vehicle to stop? Does the passenger have an obligation to the broader public to ensure that the 'driver' (i.e. vehicle) is not acting dangerously or malfunctioning? When would this become the responsibility of a mobility operator, tele-operator, or control center manager? The point here is to highlight the question over the degree to which the modes of acting on, or engaging the vehicle, as well as which specific components are capable of acting through those different modes and when, all critical. A more speculative question is how might this change if automated vehicles become sufficiently advanced such that they can perform driving maneuvers too complex for a human to supervise such as very high-speed driving, or using continuous flow intersections? At what point on the spectrum of fully driverless vehicle activities do we abandon the idea of human responsibility altogether?

These questions of responsibility, as channeled through the interface, also raise questions about affects on autonomy rendered by this reconfiguration of components. However we think of autonomy in a privately owned traditionally controlled vehicle, the dramatic rearrangement of components and modes of acting inevitably challenge whether there remains a meaningful connection between control, intention, and autonomy.[35] Indeed, ceding control over vehicle functionality may not necessarily eliminate autonomy. There may be ways to create situational awareness and meaningful orientation and connection to the environment without exerting control over a vehicle in a traditional way. For instance, researchers have explored non-driving in-vehicle activities associated with work or leisure might maintain autonomy.[36] In the commercial context, perhaps providing feedback to company on a vehicle experience is sufficient to make a user feel as if they have agency?

On the other hand, within this configuration of components, it is also possible that a human user's intentions might be undermined in a meaningful way. Once again, ownership models have a

[35] Sven Krome, Jussi Holopainen, and Stefan Greuter. "AutoPlay: Unfolding motivational affordances of autonomous driving." *Automotive User Interfaces*. Springer, Cham, 2017. 483-510.

[36] Ibid.

significant influence here. With the proliferation of ride-hailing services, and 'drivers' compelled to use shortest path algorithms, there is a shift from the traditional norms associated with taxis, of passengers having an influence or at least say, over the route taken by the vehicle. Users of ride-hailing services however, especially in the case of ride-sharing or pooling, no longer express control in that way. The normalization of shortest path algorithms thus establishes a baseline that further automation does not disturb. Where loss of control over route may become an issue however, is where that route is influenced by agendas other than the passenger's agenda of travelling most directly to a destination. One can imagine commercial incentives taking people past particular destinations or restaurants, in the same way that entities could pay the Pokémon Go platform for Pokémon to be spawned in near their commercial establishments to increase foot traffic to their destinations. It is also possible to imagine 'public safety' agendas prohibiting travel to or through certain places, or 'policing' incentives using autonomous vehicles for the sake of de facto arrest.[37] These concerns have already come up with respect to commercial ride-hailing service drivers using shortest path algorithms, when the shortest path algorithm instrumentalizes the driver for mapping or other purposes.[38]

Archetype 2 – 'Driver Assist'

An alternative vision to 'driverless' cars that retains 'drivers' in drivers' seats, sees the implementation of increasingly sophisticated automation in privately owned passenger and logistics vehicles. This is often described as an Advanced Driver Assistance System (ADAS). Most vehicle manufacturers are pursuing these systems in one form or another. Driver assistance technologies typically use similar sensor arrays to fully 'driverless' vehicles. For instance: LIDAR,

[37] See e.g. Elizabeth Joh, 'Automated Seizures: Police Stops of Self-Driving Cars' (2019) *New York University Law Review Online* (forthcoming) available <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3354800>

[38] See e.g. Alex Rosenblat, *Uberland: How Algorithms are Rewriting the Rules of Work* (University of California Press, 2018)

ultrasonic, radar, and video camera for 'computer vision'. That said, Tesla notoriously relies more on computer vision as those vehicles are not presently equipped with LIDAR.

The ADAS approach involves a dynamic control relationship between automation and human,[39] with a precarious balance of autonomy, control and responsibility.[40] Central to this model, at least at this stage, is an understanding that because vehicle automation is limited, 'control transitions' between automated and human drivers are necessary in situations where obstacles or emergencies require human attention. Both the Uber and Tesla fatalities attest to that requirement.

Authors have tried to classify and build a taxonomy for different types of control handovers, including: step-wise handovers (first longitudinal then latitude etc), driver monitored (i.e. driver has hands on wheel and a countdown happens), and system monitored (the vehicle decides when the human is ready to resume control).[41] Other categories include 'scheduled' or 'unscheduled', as well as system or user initiated. Vehicles may also be able to alter how a handover is performed according to an assessment of the attention or activities taking place within the vehicle, and a profile of the human driver's capacities. Authors note however, that somewhere within these activities, questions of responsibility between driver and manufacturer become blurred.

Effective control transitions require extremely complex interfaces and informational interactions between the vehicle and human. Timing and clarity of communications are critical, and impact

---

[39] Natasha Merat and John Lee, 'Preface to the Special Section on Human Factors and Automation in Vehicles: Designing Highly Automated Vehicles with Driver in Mind' (2012) *Human Factors: The Journal of Human Factors and Ergonomics Society*

[40] Ibid.

[41] Roderick McCall et al 'Towards a Taxonomy of Autonomous Vehicle Handover Situations' (2016) *Proceedings of Automotive UI* 193-200.

the capacity to regain control and situational awareness within time to navigate the obstacle.[42] This raises questions over how much engagement between the user and vehicle is necessary? To what degree is, or *should*, the user be included in the control loop? And what differences in politics and values do these decisions express? The practical dimension of these transitions is a continuing research question:

> Managing automation mode transitions when a driver may be distracted poses numerous questions. In switching to an automated mode, how and when does the vehicle communicate to the driver the tasks for which the system is now responsible? To what extent is the driver monitored to ensure that they are sufficiently engaged with the driving task when the vehicle has control (Eye tracking? One hand on steering wheel?)? How long does the distracted or sleeping driver need to achieve sufficient awareness of the driving situation such that they can safely re-engage with the driving task? What information and cueing mechanisms will be most effective in managing this process? How does the vehicle manage if the driver is unable or refuses to resume control? In returning control to the driver, does the vehicle always return to full manual control (no automation) or does the vehicle step down through automation levels gradually? While engineers deliver technical solutions to enable automated driving, the answers to each of the questions may be critical in ensuring that drivers' experience of automated vehicles is safe and enjoyable.[43]

Some have argued that this issue can be simply resolved by identifying the vehicle as the 'driver' when autonomous mode is engaged, and the human as the 'driver' when manual driving is occurring.[44] However, this is overly reductive.

---

[42] See e.g. Mok, B., et al. Timing of Unstructured Transitions in Automated Driving. Intelligent Vehicles 2015, June, Seoul, Korea; Mok, B., Johns, M., Lee, K., Miller, D., w Sirkin, D., Ive, P., Ju, W. Emergency, Automation Off: Unstructured Transition Timing for Distracted Drivers of Automated Vehicles. Intelligent Transportation Systems 2015; Mok, B., Johns, M., Miller, D., Ju, W. Tunneled In: The Effects of Active Secondary Tasks on Unstructured Transitions from Automation. Accepted to CHI 2017

[43] Michael Fisher, Nick Reed, Joseph Savirimuthu 'Misplaced Trust?' (2014) *Engineering and Technology Magazine* (http://cgi.csc.liv.ac.uk/~michael/automotive_preprint.pdf)

[44] Mark Geistfeld, 'A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation' (2017) 105 *California Law Review* 1611.

The design of interfaces capable of effecting these transitions is part of the 'human factors' research domain of autonomous driving, with each configuration produces its own value consequences. Authors have commented that 'Finding the right balance between requiring the human to be ready to intervene at a moment's notice and realizing the benefits of this technology is likely to be a challenge.'[45] For instance, does the interface communicate sufficient information to enable 'drivers' to make informed choices, or does it communicate sufficient information to keep 'drivers' compliant with a vehicle's suggestions (i.e. suggesting driver autonomy has a negative correlation with safety)? Does the interface obtain and demand attention in any moderately questionable situation, risking habituation, or does it only interrupt the 'driver' in emergencies?

These questions become more complicated when 'drivers' are not required to supervise the vehicle operation, despite having the controls to do so. Even though the 'safety driver' implicated in the Uber fatality was using a personal device at the time of the accident, the future of autonomous driving my include users being encouraged to perform tasks other than driving to maintain a level of engagement and alertness.[46] This raises questions as to what devices can or should a 'user' be engaged with? What will this component of the driving system look like and communicate? Might the use of non-vehicular devices reduce the communicative efficiency of the vehicle interface? Can personal devices be connected to the information ecology of the interface? Might prohibiting personal devices decrease the autonomy of the user, and the 'economic' benefit of automation. These questions elevate the importance of in-cabin entertainment systems, their capacities, and the control that various entities have over them as components of total system functionality. Regulatory environments also become highly influential components of the configuration of these systems, because road rules will influence what tasks a human 'driver' / 'owner' / 'user' can or must perform at any particular time.

---

[45] Rand Corporation *Autonomous Vehicle Technology: A Guide for Policy Makers* 68

[46] Miller, D.B., et al. Distraction Becomes Engagement in Autonomous Vehicles. Best Student Paper in Surface Transportation Track at 2015 Annual Meeting of the Human Factors & Ergonomics Society, October 2015.

Interfaces designed for the possibility of control transitions may communicate via auditory,[47] visual, mechanical, haptic, or in combinations. These are new modes of acting on the human occupant, to trigger some type of human behavior. The goal is to produce a feedback loop between vehicle and user (however they are to be defined) for the sake of ensuring the intentions of the user are executed, and that execution occurs safely. Research exploring how an automated vehicle might communicate its knowledge and perceptual capacities at any particular time is continuing.[48] We argue that those interfaces, what they communicate and fail to communicate, and what control inputs they allow or prohibit, have consequences for values.

As noted, because of the diffusion of control across the interface between the human component and the automated driving component, regulatory components may prescribe the actions that parties can take in these situations. Such regulation addresses that the identity of components for the sake of control is blurry. With respect to compliance then, ADAS vehicles may require surveillance of 'drivers' in order to ensure they are paying sufficient attention for safe vehicle operation. This is novel privacy issue, potentially requiring real-time in-cabin surveillance cameras with behavior, emotion, and fatigue detection, or measuring the amount of contact a person has with vehicle controls. Control transitions will also likely require fine-grained recording of control behavior, such that manufactures, who are likely to be *de facto* responsible for accidents when cars are in autonomous mode, can pursue 'drivers' for negligence. The privacy issues associated with ADAS style vehicles are thus less defined by the commercial norms we might see in driverless cars, and more associated with increased policing and roads enforcement, and a potentially antagonistic relationship between drivers and vehicle manufacturers, as

[47] David Beattie et al 'What's Around the Corner?: enhancing driver awareness in autonomous vehicles via in-vehicle spatial auditory displays' (2014) *Proceedings of NordiCHI* (https://dl.acm.org/citation.cfm?id=2641206)

[48] Evangalene Pollard, Phillipe Morignot, Fawzi Nashashibi 'An ontology based model to determine the automation level for co-driving' (2013) *Proceedings of the 16th International Conference on Information Fusion* (https://ieeexplore.ieee.org/document/6641334)

mediated by insurers. In other words, connected with establishing the identity of components and what their responsibilities are.

The question of responsibility in ADAS systems is also different, and perhaps more complexm than in fully driverless implementations. Because we anticipate the 'liability' issues associated with this vehicle configuration are likely to be resolved in a relatively pragmatic way, we are interested in the more philosophical dimensions of whether it is the vehicle or the human sitting behind the wheel that is responsible for ensuring safe transitions of control.[49] This is an extremely complex question that will depend, primarily, on the degree is an occupant is obliged to pay attention to the road. But we should also recognize that it may even be entirely impossible for a human driver to effectively perform a 'control transition', as there is no evidence that humans can effectively resume control in the context of an emergency. Perhaps if a human successfully negotiates a dangerous control transition it should be considered lucky rather than a fulfilling of responsibility? Private vehicle ownership also introduces new responsibilities that seem analogous to those associated with contemporary non-automated vehicle ownership. For instance, maintenance for autonomous vehicles might include the installation of software or firmware updates. It may be that a 'user' has an obligation to ensure that the 'driver' is performing at its highest capacity (i.e. using the latest version of its software), in the same way that a person is responsible for ensuring the maintenance of a vehicle (or lack of maintenance) does not cause a danger. On the other hand, this may remain a vehicle manufacturer exercise. However, it occurs, software updates are a complex problem with dramatic consequences if done incorrectly.[50]

[49] We acknowledge that these issues of responsibility vs liability may not always be so simply determined. See e.g. Andrew Selbst, 'Negligence and AI's human Users' (2019) *Boston University Law Review* (forthcoming) with respect to tortious liability, and Sabine Gless et al, 'If Robots Cause Harm, Who is to Blame? Self Driving Cars and Criminal Liability' (2016) 19(3) *New Criminal Law Review* 412 with respect to criminal liability.

[50] See e.g. Deirdre Mulligan and Kenneth Bamberger, 'Public Values, Private Infrastructure and the Internet of Things: The Case of Automobiles' (2016) 9(1) *Journal of Law and Economic Regulation* 7.

ADAS systems also introducee specific 'autonomy' consequences. The use of ADAS reportedly causes drivers to feel a reduction in autonomy and a loss of control over vehicles.[51] Research also suggests that 'user experience' and 'user acceptance' are at their highest with limited levels of vehicle automation.[52] In measuring autonomy across the interface, it becomes important to ask what the interface permits or prohibits. What is its agenda? The interface of a prestige car may be focused on comfort. In a ride-hailing service it may focus on customer experience, in a logistics vehicle such as a long-haul truck, perhaps efficiency and discipline. These agendas will define what inputs from the human are desirable or necessary. In a privately-owned vehicle, driver assist technologies are typically understood as supporting a driver's intentions, despite shared control over the vehicle.[53] This might not be the case for truckers.

Archetype 3 – 'Connected Cars' / 'Internet of Cars'

The computer science research paper 'Robust Physical-World Attacks on Deep Learning Visual Classification' published in CVPR 2018 suggested that the computer vision processor on board an autonomous vehicle could be tricked to not recognize a road-sign with a simple application of black and white stickers.[54] 'Computer vision' is one of the sensor systems common on-board

---

[51] Alexander Meschterhakov et al 'Experiencing Autonomous Vehicles: Crossing the Boundary between a drive and a ride' (2015) *CHI EA* (https://dl.acm.org/citation.cfm?id=2702661) ; (Kraus, S., Althoff, M., Heissing, B., & Buss, M. (2009). "Cognition and Emotion in Autonomous Cars," Intelligent Vehicles Symposium, IEEE.)

[52] Christina Rodel et al 'Towards Autonomous Cars: The Effect of Autonomy Levels on Acceptance and User Experience (2014) *Automotive UI* (https://dl.acm.org/citation.cfm?id=2667330)

[53] Michael Fisher, Nick Reed, Joseph Savirimuthu 'Misplaced Trust?' (2014) *Engineering and Technology Magazine* (http://cgi.csc.liv.ac.uk/~michael/automotive_preprint.pdf)

[54] Kevin Eykholt et al, 'Robust Physical-World Attacks on Deep Learning Visual Classification' (2018) *CVPR* (https://arxiv.org/pdf/1707.08945.pdf)

autonomous vehicles. Whereas systems like RADAR and LIDAR excel at determining the form, range, and velocity of objects, computer vision generates assessments of what those objects are. I.e. this shape is a human or bicycle, this sign specifies a speed limit of 30, or that vehicles must stop at an intersection. While the research paper offers a meaningful demonstration of the fragility of computer vision systems and deep learning algorithms more generally, its effect on the deployment of autonomous vehicles systems is relatively limited for at least two reasons.

First, fully autonomous vehicle systems are typically deployed only in areas pre-mapped to very high resolution. Vehicles thus use their sensors to position themselves and move about within an already well-understood physical space that includes knowledge of road signs. Second and related, it is unlikely that autonomous systems would, in their practical deployment, rely on computer vision to recognize road infrastructure like street signs. Instead, street signs, if not specifically mapped, may be part of the system controlling the vehicle. This archetype of integrated infrastructure is sometimes called ‘connected cars’ or the ‘internet of cars’. While there is, of course, a spectrum between them, this archetype represents an alternative to ‘autonomous’ models where vehicles rely primarily only on their own sensor arrays to navigate the physical world. In the ‘connected cars’ approach, vehicles constantly share information with other vehicles, cloud computing infrastructures, and the roads infrastructure itself. These ‘connected car’ (or Vehicle-to-Vehicle (V2V) / Vehicle-to-Infrastructure (V2I) / Vehicle-to-X (V2X)) approaches are also relevant for both ADAS and ‘fully driverless’ models, but nonetheless produce novel rearrangements of politics and values. Connectivity means not only V2X, but also personal mobile connections and on-board computer systems. The critical differences between the ‘internet of cars’ vision of V2X and general connectivity through personal mobile devices and existing on-board communications systems however, is the ‘direct control for time-critical flow-related interventions—for which some degree of system-level coordination is required for safe operation.’[55] Accordingly, the ‘X’ becomes an important component of overall system functionality.

[55] Ibid.

Presently, very few vehicles use robust V2V or V2I communications.[56] But it has been an element of the autonomous driving vision since its inception. Early imaginations of driverless vehicles involved infrastructure as part of the communications and control systems for cars. For instance, an exhibit at the 1939 Worlds' Fair sponsored by General Motors, included electric cars with the circuitry embedded directly into the roadway.[57] That approach was tested on short stretches of highway in the late 1950s, with detector circuits buried in the pavement transmitting radio signals to guide the position and velocity of vehicles equipped with appropriate receivers and actuators.[58] Again, in the 1960s, Ohio State University pursued research in autonomous vehicles with the model of electronic devices embedded in the roadway, and similar research was done in the UK with magnetic cables that successfully transported a Citroen DS at 130km/h around a test track. Keshav Brimbaw suggests that the US Bureau of Public Roads investigated the construction of electronically controlled highways in multiple jurisdictions. In the early 1990s, it is claimed that Daimler had constructed vehicles that could travel semi-autonomously on highways, for 1000s of kilometers, at high speed, effecting lane changes, with minimum human intervention.[59]

[56] Bryant Walker Smith, 'A legal perspective on three misconceptions in Vehicle Automation' in Gereon Meyer and Sven Beiker (eds) *Road Vehicle Automation* (Springer 2018)

[57] Keshav Brimbaw 'Autonomous Cars: Past, Present and Future – A Review of the Developments of the last century, the present scenario and the expected future of Autonomous Vehicle Technology' (2015) *Proceedings ICINCO.*

[58] Ibid.

[59] Behringer, Reinhold, and Nikolaus Muller 'Autonomous road vehicle guidance from autobahnen to narrow curves" Robotics and Automation, IEEE Transactions on 14, no. 5 (1998): 810-815; Franke, U., S. Gorzig, F. Lindner, D. Mehren, and F. Paetzold. "Steps towards an intelligent vision system for driver assistance in urban traffic." In Intelligent Transportation System, 1997. ITSC'97., IEEE Conference on, pp. 601-606. IEEE, 1997.

So far, V2V communications technology has been primarily tested and promoted in the context of ‘road train’, ‘peloton’ or ‘platoon’ style transport configurations, designed to reduce environmental pollution by using the aerodynamic efficiencies of travelling closer together or behind a truck. Engineers argue that almost every aspect of ‘driver’ decision making would be improved by V2X connectivity.[60] The real-time information flow is intended to inform the ‘driver’s’ (however defined) situational awareness, improve dynamic traffic management and general planning exercises (for instance, facilitating continuous flow intersections, electronic traffic signaling, and efficient right of way sharing), enhance the general capabilities of vehicles (i.e. higher speed, coordinated cruise control, and minimal distance between vehicles which reduces emissions by increasing aerodynamics and improves traffic flow).[61] This is unlikely possible simply with V2V communications. V2X thus includes roadway infrastructure fitted with sensors and microcontrollers that some have named an intelligent vehicle grid that enables a ‘vehicular cloud’.[62] Only with integrated and transmitting roadways are some of the more advanced exercises in automation, such as continuous flow intersections and high-speed travel, considered possible. In some models, this involves hyper-marketisation of roadways, with vehicles bidding for ‘slots’ to pass intersections, or lane time in higher-speed transit corridors. In other words, dynamic tolling of road use with high granularity. There are accordingly risks of economic stratification of road users (i.e. exacerbations of the disparities associated with contemporary toll road use). In other words, these visions include a dramatic proliferation, in fact an explosion, of components, triggers, and modes of acting. It may be possible to track these changes with close analysis of an idiosyncratic deployment of connected cars, however at this level of abstraction, it can be described as controlling components shifting away from identifiable discrete entities and into a diffuse network or architecture, with multiple, simultaneous modes

---

[60] Hani Mahmassani ‘Autonomous Vehicles and Connected Vehicle Systems’ (2016) 50(4) *Transportation Science* (https://pubsonline.informs.org/doi/pdf/10.1287/trsc.2016.0712)

[61] Ibid

[62] Mario Gerla ‘Internet of Vehicles: From Intelligent Grid to Autonomous Cars and Vehicular Clouds’ (2014) *IEEE World Forum on Internet of Things* (http://koasas.kaist.ac.kr/bitstream/10203/187604/1/81062.pdf)

of operation. This explosion of an individual in a vehicle into an infrastructural vehicular cloud raises enormous questions for human and societal values. As one commentator reminds us:

> Furthermore, we are often reminded that autonomous vehicles are essentially computers on wheels, and the cabin/cockpit is merely the user interface. As such they record every event what they "see," their trajectories over space and time, who is traveling, with whom, and so on. The sheer amount of information encoded in these 3D trajectories over time and space is well beyond anything transportation scientists are used to handling. From activity analysis, to network modeling, to traffic flow characterization, and last-mile delivery optimization, individual trajectories contain a plethora of valuable information for a variety of purposes.

Each of the new modes of acting and each communicative relation between components within the V2X paradigm also produces data that can be tracked. The massive quantity of data streaming between vehicles and infrastructure creates profound opportunities for 3rd parties to participate in a new vehicular / infrastructural data economy, with the capacity to interact with vehicles (and the people within them) in real-time, for new purposes, and using new interfaces. Data governance questions are thus particularly acute in the 'connected vehicle' paradigm. Clearly, the transmission of data is essential to the proper function of all the vehicles within the network. The data produced and distributed in a connected vehicle may include 'technical data about the car and its components, data about road, weather and traffic conditions, data about the driving behavior of the car drivers, location data but also data about the use of entertainment, navigation and many other services by the car users.'[63] But beyond the mere functioning of vehicles, data governance will also be influenced by commercial imperatives, and thus provoke competition regulation. The actual configuration or architecture, 'connected cars' in any configuration therefore raise many of the governance questions that are already being asked in relation to 'smart cities'. These are the general privacy concerns in this paradigm.

[63] Wolfgang Kerber, 'Data Governance in Connected Cars: The Problem of Access to In-Vehicle Data' (2019) *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* (forthcoming) (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3285240)

To that end, proposed models for data architectures include ‘shared servers’, ‘in-vehicle – i.e. consumer choice’ data storage, manufacturer controlled servers, and even ‘peer to peer’ proposals. If viewed from a competition lens, open access to vehicle data is likely preferred, whereas a more privacy protective approach might preference an ‘in-vehicle’ data storage system. On the other hand, vehicle manufacturers prefer the ‘extended vehicle’ model, wherein the data produced is part of a continuing system between the vehicle and the manufacturer’s servers, because it gives them more control over the vehicle interface and thus a commercial advantage. Another alternative pushed by third parties is the ‘on-board application model’ which potentially gives users the capacity to choose where their data is stored (potentially in-vehicle servers) meaning the user can choose whichever service provider that already has an agreement with the manufacturer to access their in-car interface / information infrastructure. With ‘neutral server’ interfaces, the design of existing cars means that data would first go through the manufacturer who then decides which information gets passed onto the neutral server.

These various data governance approaches also define who can have access to the informational environment of the car. With the extended vehicle model only the vehicle manufacturer controls the informational experience. In the shared model, multiple parties have access. But vehicular information ecologies are already changing. Car manufacturers are already ceding control over vehicle interfaces to third parties. It is possible this may extend to technologies actively controlling vehicle functions. For instance, there may be an android driving platform installed in an Audi, who becomes the information gatekeeper in that arrangement. That means that a V2x system may also enable location and user-specific commercial messaging.

But this diffusion of components of a driving system into a broader network of multiple vehicles, infrastructure, manufacturers, and platforms also make questions of responsibility more diffuse and complex. Some have talked about how the introduction of autonomous vehicles will transform the driving behavior of all road users into a single ‘driver’ – the operating system of all vehicles operating on a certain network.[64] This represents a dramatic reconfiguration of what was once a relatively individualistic and private exercise into a broader technological, dynamic, and likely corporate network. What obligations will an individual have to the network generally? What if the sensors in one car fail because of improper maintenance causing damage to other vehicles relying on that communications network? What are the responsibilities of the network components? What if the network itself fails? In the division of fault and responsibility between vehicles and drivers, it is forgotten that roads agencies are often involved in litigation surrounding vehicle accidents.

It is worth noting that in insurance litigation for vehicle accidents, claimants often pursue roads agencies and government departments as to whether a road was appropriately signed or maintained. There may be questions about whether the road design was satisfactory. That is, perhaps the camber of the road too steep for the curve? Perhaps the speed limit too high for the visibility? These questions will inevitably have to be reformulated in the connected roadways context. For instance, was there too much latency in the communications network? Why was a specific class of communications privileged over another class in any particular situation? Did a piece of infrastructure not transmit clearly enough? Did the agent managing an intersection give inappropriate priority to a particular vehicle? What responsibility will state agencies, or the entities operating the transport 'platforms', as new components in the system with new roles, bear for the proper operation of static infrastructure and communications networks?

The autonomy question is more complex still. Connected car approaches inevitably involve loss of individual control over cars, as well as loss of control over the general informational environment. This is presented as part of the trade-off associated with using a 'smart city' style platform. For instance, in order to obtain the greater public benefits of high-speed travel or continuous flow intersections, the loss of autonomy becomes the cost of obtaining access to the

[64] Mark Geistfeld, 'A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation' (2017) 105 *California Law Review* 1611.

benefits of the 'smart city'. The agreeability of that trade-off will depend on the general agenda of the 'smart city'. This may be highly commercialized, or it may be a primarily public arrangement. The consequences however, will inevitably include a shift from a human driver in a vehicle into, on one extreme, a public, communal, social infrastructure designed for common benefit, or alternatively into a corporate, extractive infrastructure.

Conclusion

In this paper, we have introduced the *handoff* analytic, and attempted to demonstrate its utility through an analysis of autonomous vehicles. The goal has been to demonstrate that the broader lens offered by handoff affords unique and critical insights into the operation of these systems, in terms of new components and modes of acting, that have dramatic consequences for both human and societal values. In our view, this is a critical ameliorative to a focus on the ongoing transition of control into computational components, instead showing the structural, political and ethical stakes of those changes.