Redrawing Canada's Safety Lines for AV/CV

Jason Millar[1] and Ian Kerr[2]

Policy Provocation

SAE Level 4/5 automated and connected vehicles (AV/CV) will blur the bright lines that currently delineate Canada's interjurisdictional regulatory structure for motor vehicle safety. Significant delegation of the driving task to AV/CV will require us to redraw the boundaries between federal and provincial regulatory powers in a significant way. A networked mobility system will also require us to redraw the conceptual boundaries that define automotive safety itself by causing us to confront a much broader range of societal issues alongside safety, including revitalized tension between private and public-sector values.

Objectives

1) To explain how AV/CV will upend Canada's current interjurisdictional regulatory scheme for motor vehicle safety.

2) To elucidate the complexities that AV/CV will introduce in achieving the appropriate regulatory balance between individual, collective and corporate mobility system decision-making.

3) To examine competing interests and the potential for value alignment in the design of intelligent transportation systems.

4) To underscore an opportunity for a federal coordinating and rulemaking role in setting design standards and engaging a broad set of stakeholders to resolve novel social issues and appropriate jurisdictional responses.

---

[1] Dr. Jason Millar is an Interdisciplinary Postdoctoral Fellow at Stanford University's McCoy Family Center for Ethics in Society and Apple, and a Research Affiliate at the Center for Automotive Research at Stanford.

[2] Dr. Ian Kerr is the Canada Research Chair in Ethics, Law and Technology at the University of Ottawa's Faculty of Law.

Big Picture Jurisdiction

Painting with broad brushstrokes, Canadian motor vehicle safety requirements come in two different stripes. Federal regulators seek to ensure road safety by focusing on vehicles and their interoperability, ensuring that all vehicles are designed and manufactured according to an agreed upon set of standards so that human operators can drive (and crash) them as reliably and safely as possible. Meanwhile, provincial and municipal regulators seek to ensure road safety by shaping the behavior of individual road users. They train, license and prosecute human operators, post signs, paint lines, build speed bumps, sidewalks, install traffic lights, parking meters, crosswalks, railway crossings, run public awareness campaigns, and hire law enforcement officers, primarily with individual human road users—rather than the vehicles themselves—as their target.

This rough jurisdictional split between (i) regulating the vehicles and (ii) regulating the manner of their use is premised on the recognition that human nature sometimes collides with motor vehicle safety as well as broader range of social values—human frailty imposes limits on the design and regulation of the mobility system.

Human Nature and Motor Vehicle Safety

With humans behind the wheel and in the streets, much of our existing motor vehicle safety regulations respond to a limiting and inescapable fact: human road users are, by nature, an unpredictable and unreliable crowd. Though they rarely set out intentionally to cause harm, humans regularly bend, forget and break the rules of the road. Drivers and pedestrians get angry, sad, depressed, giddy, tempted, celebratory, curious, distracted, bored, inattentive, nervous, sleepy, confused, selfish, lazy, pressured and drunk.

In fact, these predictable human failings are at the very core of justifications for pursuing a SAE Level 4/5 AV/CV mobility system. The central claim is to dramatically improve safety and advance overall mobility for all road users by eliminating the human factor from the driving task. But, as the National Association for City Transportation Officials (NACTO) reminds us, "this future is not guaranteed."[3] From a design perspective, eliminating human drivers from the mobility system means delegating a huge amount of driving and decision-making to the mobility control system—i.e., the vehicle and the broader connected mobility infrastructure. The challenge will be to eliminate human failings while, at the very same time, aligning core human values within the infrastructure of the system.

Delegated individual vehicle decision-making will likely eliminate many of the accidents that are currently caused by human frailty: keeping to the posted speed limit; lane changes; maintaining a safe following distance; lane keeping; and negotiating intersections, pedestrian crosswalks,

[3] NACTO 2017 at p.11.

interactions with cyclists, and so on. This will directly improve safety.[4] But it will almost certainly result in new injuries and deaths that would not have occurred with human operators.[5]

But automating the mobility system also entails automating a broad set of controls in the AV/CV mobility system that were previously up to individual drivers. Put another way, we were previously only able to exert limited control over individual human driving decisions—an automated and connected mobility system enables almost perfect control over vehicle movement and decision-making within the system. This redistribution of control will have significant ethical implications and will challenge current approaches to interjurisdictional regulation of motor vehicle safety, as stakeholders in the mobility system *other than human drivers* acquire responsibility and accountability for the driving function and broader decision-making in the AV/CV mobility system.

Blurring Jurisdictional Lines

As described above, the aspects of safety governance associated with the operator of the vehicle are primarily within the jurisdiction of provincial and municipal regulators. Federal regulation tends to deal with the safety risks associated with the vehicles. But what happens when the driving function and decision-making are no longer carried out by humans but instead by the vehicles and the broader networked mobility systems? Will provincial and municipal authorities begin to interact with manufacturers and network providers so as to continue regulating the driving function? Or, will the current role for federal authorities expand? How will the private sector influence the way that these jurisdictional lines are redrawn? Regardless of the best policy option, it seems clear that the current jurisdictional lines will become blurred.

Clarifying the lines will create new opportunities for substantial interjurisdictional cooperation. It may also require the resolution of significant capacity challenges at various levels of governments, including the provision of appropriate financial commitments, human resources, expertise, and the ability to implement agreed upon changes. Resolving these capacity deficits will involve more than goodwill. It will be expensive, labour intensive and time sensitive. It is unclear at this time whether many of the players currently involved in the regulation of the driving function (or in the development of the AV/CV systems, for that matter) will have the capacity to participate in the development, implementation and maintenance of new mobility systems without some kind of reallocation of resources. Nonetheless, the issues raised by an AV/CV mobility system are significant and unavoidable.

[4] NHTSA 2015.

[5] Lin 2013.

Novel Societal Considerations

Eliminating the human driver requires delegating a number of critical mobility system decisions to the control software running the system, which generates novel societal considerations. For example:

- How will mobility and traffic safety be distributed in the new mobility system?
- Will mobility networks be used to automate law enforcement, and how might this impact privacy and other civil liberties?
- To what extent must intelligent mobility systems align with local values (e.g. provincial and municipal) reflected in urban design?
- Must Canada's mobility infrastructure adapt to accommodate the needs of industry? Who pays?
- How can the private sector's desire to monetize elements of new mobility systems be balanced alongside individual, governmental and societal interests?
- How can we ensure that intelligent mobility systems meet the needs of equality seeking groups?

This brief list is only a tiny sample of the many system level decisions that mobility planners, regulators, manufacturers and the many stakeholders who interact with the new mobility system will confront as a direct result of eliminating the human driver. Early indications of these tensions can already be seen in the adoption of existing automation and artificial intelligence applications.

Consider the two Case Studies that follow.

Current Illustrations and Discussion Prompts

Case Study 1: Gaming Waze

Waze is one of Google's mapping and navigation applications.[6] It runs on mobile phones and offers detailed turn-by-turn navigation. In addition to standard navigation prompts, Waze allows users to "report" traffic slow-downs, police "radar traps", construction, potholes, roadkill and any number of other driving hazards that might impede the regular flow of traffic, thus allowing drivers (and the app) to minimize driving delays. Like other navigation apps, Waze selects its routes based primarily on the shortest door-to-door travel time. This core design decision has led to conflict between citizens, city officials and Waze. The conflict arises when Waze decides to

[6] https://www.waze.com/

start routing traffic along city streets that used to be relatively quiet and off the beaten path. Local citizens become unhappy about the sudden increase in traffic along their previously quiet streets, while authorities contemplate the redistribution of traffic risks, as safe zones are compromised by private software and navigational service providers.

To combat Waze, some citizens have recently resorted to "gaming" the algorithm—reporting false accidents, construction, and other fictional delays that would trick Waze into rerouting traffic somewhere else. Likewise, city officials have resorted to installing new signage, such as "no left turn" restrictions, in attempts to limit access to certain roads at rush hour. An article in WIRED went so far as to outline several effective ways to defeat Waze's algorithm, such as installing speed bumps, no entry signs, narrowing streets, etc.[7] With this, we see that automation technologies are fueling ad hoc, post hoc, and potentially expensive regulatory responses.

Reports of these conflicts—between Waze and local officials/citizens—are multiplying.[8] This worsening conflict stems from a design decision that delegates routing decisions to Waze based only on the human value of *time efficiency*. Without proper regulatory oversight, Waze and other private sector entities fail to take into account other, locally determined, traffic management decisions that are clearly important in the broader mobility system, such as maintaining relatively traffic-free roads in dense pedestrian areas or neighbourhoods.

It is unclear how these novel safety risks should be regulated in an increasing AV/CV environment. But it is clear that mapping applications like Waze will eventually control the flow of traffic within the mobility system and could have any number of societal values at its core—safety is only one of them. For example, it is unclear what would happen if Waze decided to shift away from time efficiency, basing its algorithm instead on design values that would route traffic through vendor sponsored commercial areas. Or if they decided to shape the level of service within the mobility system based on discriminatory socio-economic factors, and so on. Waze is currently a voluntary service, but will eventually be at the heart of the mobility system control center. Thus, to be clear, the adoption of particular design values becomes even more significant with SAE level 4/5 automation, where passengers, municipalities and other public entities may lose control over key mobility system control and routing decisions.

Policy Considerations:

1) What design values are permissible for mobility system control/routing software like Waze in developing its local traffic routing and mobility distribution preferences?
2) What role might the various levels of government play in ensuring control/routing software like Waze is designed to take into account local mobility preferences and ordinances?

[7] Marshall 2016.

[8] See, for examples: Weise 2017; Paulas 2017.

3) Do existing regulatory frameworks adequately address potential problems as control/routing software is delegated more and more of the navigation function? If not, what additional structure needs to be in place to do so?
4) Should AV/CV navigation software be prescribed by law?
5) What are the guiding values and what governance frameworks should be used to enforce them?
6) To what extent does this case illustrate a problem that is unique to AV/CV? Are there lessons learned in other contexts useful to AV/CV?

Case Study 2: Jurisdictional Bias in Machine Learning Algorithms

Machine Learning (ML) algorithms are credited with many recent autonomous driving successes, which suggests they will be a key technology in any automated and connected mobility system. ML relies for its training and subsequent successes on very large curated data sets; those data sets are used to both train and validate the ML algorithm. Those data sets can also serve as problematic sources of bias in the resulting behavior of a system.[9] For example, in 2015 Google's image classification system, an ML system designed to classify objects in pictures, was widely publicized as biased because of its tendency to classify African-American faces as "gorilla".[10] One underlying cause of this bias was that the data set used to train and test the system did not contain enough facial images of African-Americans to train the system to consistently distinguish between African-Americans and gorillas. This example of ML bias raises significant challenges for an automated and connected mobility system that crucially relies on sensors and software—rather than people—to accurately identify objects in the vehicle's path.

Mobility systems that rely on ML for their functioning will be trained using specific data sets, and those data sets will often be representative of the populations or jurisdictions from which the data are gathered. For example, let us imagine a data set collected in and around Kanata, the current heart of Canadian autonomous vehicles testing. In Kanata (representative of a broader driving norm in urban centres in the Province of Ontario), drivers tend not to stop voluntarily for pedestrians except at stop lights or designated pedestrian crossings. Because Ontario pedestrians therefore tend to cross only when designated to do so or when no cars are around, a Kanata-trained ML algorithm would predict and function accordingly. Calgary pedestrians, on the other hand, tend to cross at will, expecting a car to stop whenever they are trying to cross, even when they are not at a designated crosswalk.[11] Consequently, if the Kanata-trained system is deployed in downtown Calgary, subtle but potentially significant safety risks could arise. The difference in pedestrian behaviours between jurisdictions could result in a system that puts Calgarians at elevated risk, even if only slightly. In other words, where behavioural norms differ between

[9] AI Now 2017.

[10] Zhang 2015.

[11] Except, of course, when the car has a green light on a major roadway.

jurisdictions, a system designed for one set of such norms will bias against expected outcomes in other jurisdictions.

This kind of jurisdictional bias can extend beyond safety considerations to broader mobility system design issues that can affect equity and access. Consider a system intended for Montreal, and trained to maximize the availability of autonomous rideshares (or AV/CV deployed public transportation) based partly on data representative of the demand patterns in a particular urban U.S. setting, say San Francisco. That data set could represent socio-economic trends in mobility patterns typical in San Francisco, but could result in bias when deployed in Montreal. The bias could result in certain segments of the Montreal population having greater access to mobility than other, underrepresented segments of Montrealers. Thus, cross-jurisdictional issues of fairness and equality are relevant to the use of data sets to train and validate autonomous and connected mobility systems.

Policy-Relevant Questions:

1) How/Should ML algorithms like those imagined in Kanata be designed to take into account local driver/pedestrian norms in other jurisdictions? Should vehicles have adjustable settings? Who should be responsible for choosing the default settings? Who should be permitted to change the settings?
2) What role might the various levels of government play in ensuring ML algorithms are designed to take into account local traffic norms and ordinances?
3) Do existing regulatory frameworks adequately address potential problems as ML algorithms take over the driving function. If not, what additional structure needs to be in place to do so?
4) Should AV/CV ML software be prescribed by law to minimize cross-jurisdictional bias?
5) What are the guiding values and what governance frameworks should be used to enforce them?
6) How do we ensure “training” data is generated locally to help mitigate jurisdictional bias in AV/CVs?
7) To what extent does this case illustrate a problem that is unique to AV/CV? Are there lessons learned in other contexts useful to AV/CV?

References

AI Now. (2017). *AI Now 2017 Report*. Online: https://ainowinstitute.org/AI_Now_2017_Report.pdf

CCMTA-CCATM. (2016). *Automated Vehicles in Canada*. Online: https://ccmta.ca/images/publications/pdf//CCMTA_WhitePaper_Automated_Vehicles_in_Canada_2017.pdf

NACTO. (2017). *Blueprint for Autonomous Urbanism.* Online: https://nacto.org/publication/bau/blueprint-for-autonomous-urbanism/

NHTSA. (2017). *Automated Driving Systems 2.0: A Vision for Safety*. Online: https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/13069a-ads2.0_090617_v9a_tag.pdf

NHTSA. (2015). *New Car Assessment Program (NCAP)*, 80 Fed Reg 78522 at 78552, December 16, 2015 Online: https://www.federalregister.gov/documents/2015/12/16/2015-31323/new-car-assessment-program

Lin, P. (2013). "The Ethics of Saving Lives with Autonomous Cars Is Far Murkier Than You Think." *WIRED*. (JULY 30). Online: https://www.wired.com/2013/07/the-surprising-ethics-of-robot-cars/

Marshall. A. (2016). There Are Better Ways to Kill Traffic Than Lying to Waze. *WIRED* (July 5). Online: https://www.wired.com/2016/07/better-ways-kill-traffic-lying-waze/

Paulas, R. (2017). "For the Good of Society — and Traffic! — Delete Your Map App." *NYMAG*.com (Dec 11). Online : http://nymag.com/selectall/2017/12/waze-and-google-maps-create-traffic-in-cities.html

Waymo. (2017). *Waymo Safety Report: On the Way to Fully Self-Driving*. Online: https://storage.googleapis.com/sdc-prod/v1/safety-report/waymo-safety-report-2017-10.pdf

Weise, E. (2017). "Waze and Other Traffic Dodging Apps Prompt Cities to Game the Algorithms." *USA Today* (Mar 6). Online: https://www.usatoday.com/story/tech/news/2017/03/06/mapping-software-routing-waze-google-traffic-calming-algorithmsi/98588980/

Wirtschafter, E. (2017). Driving Apps Like Waze Are Creating New Traffic Problems. *KALW* (Mar 23). Online: http://kalw.org/post/driving-apps-waze-are-creating-new-traffic-problems#stream/0

Zhang, M. (2015). Google Photos Tags Two African-Americans As Gorillas Through Facial Recognition Software. *Forbes* (July 1). Online: https://www.forbes.com/sites/mzhang/2015/07/01/google-photos-tags-two-african-americans-as-gorillas-through-facial-recognition-software/#530ebbca713d