

SUBMISSION TO NTC: [In-service safety for automated vehicles](#)

M Haines. CEO VANZI

295 Esplanade East, Port Melbourne, Vic 3207

m.haines@vanzi.com.au

0418 121 561

8 August, 2019

Ethical Driving Rules

First and foremost, a Self-Driving Car (SDC) must obey the law and drive defensively. There are no unresolved ethical questions if it fails to do so.

This paper looks in detail at the ethics involved where the *SDC is not at fault* AND it has *time to consider* evasive action due to another party or thing illegally intersecting its trajectory. From the perspective of the SDC, this is 'the party at fault' (recognising the ultimate fault may rest with another party, including an 'act of God').

Often this problem is likened to 'the trolley problem' where several people are tied to a rail line, and you have to decide to let them be killed by an approaching train, or pull a switch to send the train down a sideline killing only one person.

Variants include knowledge of the character and/or age of the people involved.

However, the situation where an SDC is not at fault is very different.

In the case of the people on the train track, they have been forcibly tied to the track. In the great majority of real-life cases, those in the path of the SDC will be there due to their own fault (going through a red light, speeding, crossing in front of the SDC without right of way, etc). A vastly different moral issue.

Often, an SDC will be able to take evasive action (including simply stopping), avoiding the need to make an 'ethical decision'. On other occasions, events will happen so fast that (given the speed and trajectories of the parties), or due to other circumstances, there is no way to take any action to avoid a collision.

Rarely (for an individual vehicle driving defensively within the law) will the SDC be forced to take evasive action that has no 'good' outcome (only a 'least-worst' outcome) where it has the time to consider alternative courses of action and is able to reliably act within a set of pre-determined guidelines!

Practically, *in such a situation, the only facts that can be known for certain upfront by society* (for the purposes of setting any ethical rules), and by the SDC in the moments of the accident (in order to follow the rules), are the *roles* of the people involved:

1. The 'Party at Fault'

2. The 'Occupants of the SDC'
3. The 'Occupants in other vehicles'
4. 'Unprotected Bystanders' (on foot and on mobile devices such as skateboards, mobility scooters, bikes, motorcycles etc).

With these parties in mind, the paper explores an ethical hierarchy that the SDC should be programmed to follow (where possible) in order to decide: *who to save when it is not possible to save everyone*, in all manner of circumstances. It also considers the use of sidewalks and bike paths to avoid potential harm, as well as the practicality of making a decision to cause minor injury to one party to avoid major injury and/or death to others.

As well, it looks at how to ensure the 'driver' remains in both legal and actual control of the vehicle until the SDC is capable of driving 'unsupervised' (within any defined area and/or conditions).

First, we have to know who or what has responsibility for making the decision. This comes down to 'who or what has operating control of the vehicle'.

Who or What is in Control?

[A previous paper](#) argued that the road rules should permit only two operating modes:

1. Person in legal and actual control (with and without various Driver Assist technologies), and
2. System in legal and actual control (within specified areas and conditions)

It was suggested that the best way to ensure a person remains in actual control is to have them retain control of steering at all times. Under this rule, 'hands-free' driving would not be permitted until the system could take full control.

It was also argued that once full control is handed to the system there should be no requirement for a person to supervise it – simply because people make very poor 'passive monitors'. They quickly become distracted and cannot orient themselves and react in time to avert a sudden emergency.

The only permitted handover from the system back to the person would be through a 'managed' process that ensured the person had effective control before the system was released from responsibility. This could include the car stopping safely on its own if there was any doubt about the person's capacity to drive.

Under these rules, any emergency would have to be handled by 'the system' while it remains in control.

Responsibility of Controller to Obey the Law and Drive Defensively

As a matter of principle, whoever or whatever is in legal operational control must obey the law.

There is also a moral/ethical (and in some jurisdictions legal) imperative to drive 'defensively'. For example, to drive slower with a 'watchful eye' when passing a school, or where children are playing, or while overtaking or passing a stopped bus, or when driving through a shopping strip, or in a car park; and to keep a safe distance from the car in front, depending on the speed; etc.

If a Self-Driving Car (SDC) under the control of 'the system' breaks the law and/or is found to be driving without due care, in the absence of mitigating circumstances 'the system' will be wholly responsible for the consequences. Of course, 'the system' will not be liable for any fines and/or damages. It will be one or more of the owner/operator/manufacturee/parts/software supplier(s)/maintenance provider, depending on who contributed to the failure, and to what extent.

Mitigating factors could include failure of infrastructure, such as a collapse in the road due to a burst water main. Hacking is unlikely to be regarded as a 'mitigating factor', as the manufacturer ought to ensure cyber-risks are counter-measured.

It is possible for a designer, manufacturer, parts supplier or maintenance provider to be prosecuted for criminal negligence, but unlikely. The likelihood is that any system failure will be managed in accord with recall regulations, with activity focussed on what went wrong and how to fix it.

It is up to each manufacturer to ensure its SDC system obeys the law and drives defensively. This includes taking appropriate evasive action when others are at fault.

SDC must be able to take Evasive Action when Others are 'At Fault'

Perhaps hundreds of thousands (millions?) of times every day in the US alone people take evasive action: braking hard to avoid hitting a stopped vehicle; or braking or accelerating hard to avoid a car that fails to give way at an intersection; or swerving into a parallel lane or an opposing lane to go around something or someone, or even onto the footpath. In most cases, they avoid any collision or incur only minor damage that goes unreported.

Such actions also help to mitigate potentially more serious accidents that are reported, so only property is damaged, or people sustain only minor injuries, instead of major injury or death.

Unfortunately, there are no reliable statistics for these saved incidents or mitigated accidents, but they are likely to be at least an order of magnitude greater than reported accidents.

Clearly, any SDC will have to be able to manage these incidents and accidents at least as well as most people - if they are not going to add to the toll.

In fact, they may have to do better than people to avoid liability.

In many jurisdictions, there is a recognition that the duty of care to others is reduced in an emergency. It is called the '[sudden emergency doctrine](#)'. It is based on the recognition that, in emergencies, people cannot be expected to always react in the most appropriate way - due to

the fact that they have little or no experience in dealing with such situations and are prone to panic or other human responses that limit rational thought.

In the case of an SDC, it is likely that, as the response has to be 'pre-programmed', it will be expected to adhere to agreed ethical standards.

The next sections discuss what those standards ought to entail.

Ethical Questions only arise in very limited situations

Ethics is an issue in only a tiny fraction of 'incidents', that are nonetheless significant from a moral perspective.

They only arise when:

1. the SDC is not at fault (driving defensively within the law)
2. another party causes it to take evasive action
3. there is sufficient time to react
4. there are alternative paths, but
5. every path leads to someone being injured or killed.

The cause may be, for example, something has fallen off another vehicle into the path of the SDC, or another party is likely to illegally intersect the SDC's path (as when a person or animal runs out, or another vehicle fails to give way when it should, or an oncoming vehicle overtakes toward the SDC in an unsafe manner, or loses control, etc).

The principal question is: who to save when it is not possible to save everyone?

Practical Limitations

In practice, who to save comes down to which path to take: straight ahead, swerve left, or swerve right; and to brake or accelerate?

As it is impossible to compute the outcome of multiple collisions, the ethical decision must be limited to the path the SDC first chooses, without regard for what happens after any subsequent collision. This is no different to the choices people make, except an SDC may have more time and be better able to compute the best alternative due to the extended range of its sensors, access to a detailed 3D model of the immediate environment, its processing speed, and better 'awareness' of its own dynamic capacity to respond in the circumstances, unaffected by emotional factors.

In implementing any strategy, just like people, an SDC will also be constrained by the physics of the situation. In general, the faster it is going and the less traction it has, the narrower will be its 'arc of response' (the range within which it can veer left or right).

Before we can decide on the most appropriate path, it is first necessary to answer a subsidiary ethical question: should the footpath and/or bike paths be used as escape routes?

Use of the Footpath and Bike Paths as Potential Escape Routes

While driving onto any footpath or bike path is inherently unsafe, every day around the world, people do it to avert a collision or mitigate the outcome of an accident, even if the law does not specifically allow it.

In the case of SDC, a specific decision needs to be made to either prohibit or permit the use of the footpath and/or bike path to avert a collision *in an emergency that is not of the SDC's making*, as this will need to be pre-programmed into the car.

Ethically, there is no rationale for treating an SDC any different to a human driver. It means that whatever decision is made for the SDC, it should apply to human drivers also.

Unfounded Concerns relating to the Use of the Footpath and Bike Paths to Avert an Accident

If the law specifically authorises the use of the footpath and/or bike path as an emergency escape route, it is recognised that some people may become unnecessarily fearful that they are being exposed to a new risk, namely: collision with a car on the path.

However, the fear is unfounded for several reasons:

First, people on foot, bikes and other open mobility devices are already exposed to the risk, legally or not.

Secondly, though it happens, few people will encounter the risk in the whole of their lives due to the rarity of the event. You only have to consider how often it has happened to you and people you know. This is a consequence of two facts: 1) the need to drive onto a footpath or bike path to avoid collision is itself a relatively rare event; and 2) apart from city centres and shopping strips during certain hours, any point on a footpath is rarely occupied, and then only very briefly. As a result, (outside crowded centres), the chances that a pedestrian is in the path of a car in any incident is quite remote. And even when they are in the path, they often jump out of the way. Bike paths will be crowded in different places at different times. Even so, most of the path will be unoccupied most of the time.

Thirdly, and most importantly, the authority to use the footpath and/or bike path would be subject to the overriding requirement that such action be 'safe'. In practice, this means that no unprotected person (eg people on foot, or bikes or other open mobility devices, etc), suburban fence or shopfront should be within the stopping range of the vehicle when it swerves onto the bike path and/or mounts the kerb. (The reason for excluding collision with fences and shop fronts is that it is impossible to know if anyone is on the other side).

Other Countermeasures

In places and at times where people are in significant numbers but cars still need access to the road; emergency use of the footpaths and bike paths may be outlawed, but the speed limit can be reduced to mitigate any accident. Or, people can be given the right of way, forcing cars to slow and be on alert. In this case, the whole road is turned into a footpath/bike path, with cars permitted to use it with due care. This 'traffic calming' measure has been trialled with some success in a number of cities, Copenhagen amongst them. Melbourne is another city that has a long-term plan to develop a hierarchy of roads where pedestrians take precedence.

Consequences of any Ban against using Footpaths or Bike Paths as an Escape Route

The decision to ban or not to ban will likely be a 'local' (city x city) decision.

If any regulator acquiesces to fear and bans emergency use of the footpath and/or bike paths by any vehicle, it will doubtless lead to:

- more traffic violations as people are prosecuted for crossing into a bike path and/or mounting the kerb to avoid an accident *that is not their fault* (which seems senseless if no one is hurt), and, of much greater concern
- unnecessary property damage, injury and death as drivers and SDC obey the law and crash into people, animals, cars and other things they could have avoided by using the vacant footpath or bike path (which seems unethical). We don't know to what extent this may happen, as we don't know how many accidents are presently avoided or mitigated using the footpath and/or bike paths.

The ethical questions that will have to be answered by each city are:

1. Do we hold back the use of SDC until they are proven safe to use the footpath and/or bike paths in an emergency (foregoing their overall benefits such as more spare time, greater mobility for all, and better safety generally)? or
2. Do we allow wide use, but prohibit emergency access to the foot and bike paths? In this case, it would mean accepting a (possibly minor) increase in property damage, injury and death (due to the SDC inability to take evasive action), in order to gain the vastly greater overall benefits of using SDC? or
3. Do we allow both wide use and access to the footpath and bike paths for emergencies, with the risk that (most likely rarely) someone may be severely injured or killed on a path as a result of an SDC taking evasive action that was not safe? and/or
4. Do we institute other traffic control measures to mitigate the risks?

The answer to these questions may change if the data shows more than a very minor an increase in accidents relative to the overall reduction in other accidents due to the use of SDC.

Technical Challenge of using Footpaths and Bike Paths as an Escape Route

Unfortunately, approving the use of the footpath and bike paths to avoid an accident is only the easiest part of the problem solved.

Separately, each manufacturer must determine if its SDC can safely mount the kerb at the specific location of any incident (likely based on detailed 3D models of the area, including fences, shopfronts and the profile of the kerb and gutter) and stop before hitting a person, fence or shop front, given the car's velocity and dynamics. If safe stopping cannot be determined with a high degree of assurance as the accident unfolds, it would simply mean that the SDC would have to be programmed to not use the paths to avoid the specific collision.

This raises a number of ethical issues.

If a manufacturer restricts its SDC from using the footpath and/or bike paths as an escape route (because it cannot do so safely) and as a result, has an accident that could have been

avoided, does society indemnify them if, overall, their safety record is much better than for human drivers?

If one manufacturer solves the problem effectively, should this solution be proprietary, or should it be made mandatory and licenced to all the other SDC producers? Can this even be done if the solution is embedded in the whole SDC system?

Overall, property damage, injury and death rates ought to drop through the use of SDC that obey the law and drive defensively. This may suggest to some that, if it will be a rare event, 'why bother developing detailed 3D models and algorithms to use the vacant footpath or bike paths in an emergency'.

The simple answer is that ethically if we can, we should... as long as the costs of doing so are not out of proportion.

Fortunately, the technology to model and recognise the natural and built form of the city and the behaviour of moving objects (including people) is improving at a rapid rate, so the challenge of utilizing the footpath and bike paths to avert or mitigate an accident is becoming easier to meet. By the time SDC are ready for general use, it may no longer be a technical problem.

Assuming each city makes its own decision to allow or ban the use of the footpath and/or bike paths to avoid an accident, we can now address the principle ethical question.

Deciding Who to Save where there are no 'Safe Paths'

This situation has been likened to the 'trolley' problem. In this problem, a train is on a single track, with a number of people tied to the rails up ahead. There is also a siding with only one person on the line and you are given a switch that could send the train onto the siding, saving more people at the expense of killing another person. The questions it raises are: is it ethical to pull the switch, and would you?

These questions have led people to argue that the decision may depend on other factors, like the ages of the people involved; or even their value to society. For example, are they a doctor or a criminal.

However, no SDC will ever face such an ethical dilemma.

In the case of the SDC, the person who wants the car to act as their agent ('pulling the switch') is not a 'detached bystander' but is sitting in the car and has a direct interest in the impending collision.

More importantly, unlike the people tied to the train line in the trolley problem, in the overwhelming majority of real-life cases, it will be their own fault that puts people in the path of the SDC (eg running out in front, or backing out or overtaking without due care, or failing to give way when they should, etc).

These facts change the moral dilemma significantly.

It is also clear that we can never know enough to make an ethical decision based on age, or any other personal characteristic. It may not even be possible to know the number of people in each path (eg, because they are inside a vehicle, or otherwise obscured from view).

The only facts that can be known for certain upfront by society (for the purposes of setting any ethical rules), and by the SDC in the moments of the accident (in order to follow the rules), are the *roles* of the people involved:

1. The 'party at fault'
2. The 'occupants of the SDC'
3. The 'occupants in other vehicles'
4. Any 'unprotected bystanders'.

With these parties in mind, the challenge is to develop an ethical hierarchy that the SDC should be programmed to follow (if possible) to decide: to who to save when it is not possible to save everyone?

Take the case of an oncoming motorbike rider who has fallen while overtaking and is skidding towards the SDC at high speed. Assume, even with hard braking the SDC will likely run over and kill the rider, with likely, only minor injury to the occupants. On the other hand, veering around the skidding rider to save them, would send the SDC into the path of an oncoming truck at high speed, likely killing the occupants, destroying the car and causing major damage to the truck.

There does not seem to be any rational argument why society should mandate 'self-sacrifice' by the occupants of the SDC in these, or similar circumstances.

Ethically, we may rule that where there is a choice, the SDC should try to save the occupants from major injury or death ahead of the 'party at fault'.

There are also arguments that the SDC should put the safety of its occupants ahead of all others, based on the principle of 'self-preservation'.

This argument is supported by English tort law which accepts that a person is entitled to protect themselves from harm caused by others, even at the expense of harming 3rd parties. cf. [Scott v. Shepherd, 96 Eng. Rep. 525 \(K.B. 1773\)](#). In this case, a person threw a lit torch into a crowd. Another person picked it up and threw it away to protect themselves, causing harm to a third party. It was held that the person who originally threw the torch was the cause of all the harm that ensued and hence liable for damages.

The original thrower is analogous to the 'at fault party' (who causes an SDC to have to take evasive action to avoid injury to the occupants). Based on Scott v Shepherd, it could be argued that the SDC is entitled to save its occupants by swerving into a pedestrian or bike rider, etc. to, say, avoid a high-speed head-on collision with a truck that has veered into its path.

As well, product liability law would suggest a supplier of an SDC should try to keep its occupants safe in the event of an accident caused by another party.

These principles and precedents give support to Mercedes decision (since backtracked somewhat) to prioritise occupants in these circumstances.

On the other hand, the principle of 'self-preservation' is not absolute. There is also a general duty of care towards others. In some jurisdictions, eg Texas, USA, the law specifically requires drivers to keep a look out for danger on the road and to take steps to avoid or mitigate an accident, even if turns out that another party was ultimately in the wrong. It means that if you can avoid likely major harm to one party at the cost of possible minor harm to yourself, or even another party, you ought to take that option.

The ethical argument is that, as a society, we should try to limit the flow of harm.

In keeping with this principle, we may decide that if the choice is between likely major injury or death for the occupants or other innocent bystanders, the SDC should limit the flow of harm and always save the bystanders.

It may be easier to accept the argument (at least as it applies to 'unprotected bystanders') if we look at an alternative countermeasure. No one could argue that it is unethical to erect an impregnable guardrail along every footpath and bike path to protect people on foot or on an open mobility device. In this case, any car faced with an oncoming truck could not evade collision by running onto the footpath or bike path due to the hypothetical guardrail. The result being an 'unprotected bystander' is saved and the 'occupants of the SDC' are killed. Their unfortunate death would not be due to the fault of the SDC, but to the truck driver (or other cause for the truck being on the wrong side of the road).

By establishing the ethical rule that an SDC should prioritise the safety of 'unprotected bystanders', it is no different to erecting an impregnable (virtual) fence around them.

The benefit of the virtual fence (created by the rule) is that it can be removed instantly if it is clear that there are no people, suburban fences or shop fronts within the SDC's 'stopping distance', allowing it to use the footpath and bike paths as an escape route when it is safe to do so.

Based on this view, the SDC should put the safety of 'unprotected bystanders' ahead of both the 'occupants' and the 'party at fault'.

Even if there was no such rule, it is hard to see any court allowing a claim against the manufacturer that the SDC should have swerved and caused major injury or death to other innocent parties, in lieu of the occupants and/or the party at fault who were actually injured or killed!

Given that the accident was not the SDC's fault, and that vision from its cameras and data from its other sensors showed that it had no alternative, the SDC system ought to be exonerated from causing the occupant's major injury or death in these circumstances. It would simply mean the occupants time was up, due to factors beyond their control, or the control of the SDC. In this case, the people who caused the accident would be held liable for all damage and harm, like the initial thrower of the lit torch.

It should be remembered that it will be extremely rare, if ever, that any individual is involved in an accident where there is time to react with alternative paths, but none are safe. Mostly,

there will be a safe path with the accident avoided or mitigated. In other cases, there will be no time to react and the accident will just happen. However, every so often, an SDC somewhere will be faced with this ethical decision, so it needs to be resolved.

Importantly, it needs to be resolved based on human values, so it is understandable by everyone and can be put into law, and followed by all SDCs as any accident unfolds.

As a [TED talk by techno-sociologist Zeynep Tufekci](#) explains: intelligent machines can fail in ways that don't fit human error patterns — and in ways we won't expect or be prepared for. "We cannot outsource our responsibilities to machines," she says. "We must hold on ever tighter to human values and human ethics."

The last thing we want is to have SDCs making 'black box' decisions on a case by case basis. Not only would this lead to unnecessary litigation, it would leave all SDC occupants and other road users uncertain as to how the SDC will respond.

On balance, it would seem that the ethical response is to take a path that saves people in priority:

1. All 'unprotected bystanders'
2. The 'occupants of other vehicles'
3. The 'occupants of the SDC'
4. The 'party at fault' (that cause the SDC to take evasive action).

For elderly people, or those with a terminal illness, or others with a strong sense of self-sacrifice, it would theoretically be possible to set up a procedure for any sole occupant (or group) to instruct the SDC to change the priority, and put themselves last. This could not be used to commit suicide, as it relies on another party doing the wrong thing to cause a life-threatening accident. From wide discussion, it is likely few may ever choose to push the 'self-sacrifice' button, which may indicate the rule correctly reflects the ethics of most people.

The next ethical question is about relative harm. What if the choice is not between major injury or death for any one of the parties, but between 'likely' minor injury to one party vs. 'likely' major injury or death to another?

Ethics of Causing Minor Injury to One Party vs Major Injury or Death to Another

First, we should accept that collision with 'unprotected' bystanders ought to be avoided at all costs. The reason is that even a minor bump may knock them over causing a hit to their head, which can easily result in a major brain injury or death. This is also in keeping with the idea of erecting an 'impregnable barrier' around them, as previously discussed.

This would mean, for example, that an SDC could not swerve onto a footpath or bike path, even if it meant saving (say) a fallen motorcyclist from almost certain death - if there was any risk of hitting an unprotected person or going through a suburban fence or shop front.

The problem is what to do where people are protected in vehicles.

SDC Crashing into other Objects to Avoid Serious Harm to 'Unprotected' People

The simplest response is to outlaw any evasive action that could hurt the occupants of any vehicle. However, it would likely result in more major injuries and deaths to unprotected 'at fault' parties which are now avoided at the expense of property damage and nil or minor injury to occupants of other vehicles.

For example, at a speed that would likely kill a pedestrian or fallen rider; due to the use of seat belts, crumple zones and perhaps airbags, a car could safely swerve (to save the pedestrian who may have walked out in front of it, or the fallen rider) and hit a pole or parked car, or even a slow-moving truck in the opposite lane, without undue risk to the occupants of the car (or the truck).

Ethically, it is hard to argue that the SDC should kill the wayward pedestrian or fallen rider (even though they are 'in the wrong'), to save the occupants from (at most) likely minor harm.

The problem, in this case, is not the ethics. It is the uncertainty of the outcome. How can we know in advance what the likely outcome of a collision will be?

The following analysis applies only where an SDC is 'not at fault' and must consider crashing into a solid object as a result of 'unprotected' people causing it to consider taking evasive action.

This is where it gets a little messy!

In the case of poles, trees, embankments and other solid objects, the consequences of the impact can be directly assessed by the SDC manufacturer. This should add little extra cost or complexity as manufacturers must already assess different impact scenarios to determine the car's safety rating.

Given the car's safety systems, based on actual accident data, simulation and test crashes, the manufacturer can determine the risk to the SDC's occupants at different occupancy loadings, impact speeds and angles (head on, rear, driver side, passenger side, and four corners). This process should result in a series of published thresholds for each type of SDC below which there is 95% chance of nil or minor injury to the occupants in different scenarios.

It would then be left to the SDC system to assess its loading, the nature of the other objects (eg pole, tree or other solid object) within its 'arc of response', the relative speeds and angles of impact and decide that, in a specific instance (based on the manufacturer's published thresholds), it will swerve to avoid hitting the fallen rider or wayward pedestrian and collide with another solid object instead.

The result may be that the occupants sustain nil or minor injury (say bruising from the seat belts), while the car is perhaps written off... but a fallen rider or pedestrian is saved from likely major injury or death.

Once the thresholds are approved, assuming all the SDC's safety systems are working correctly, if a crash (that is not the fault of the SDC) does occur below the stated thresholds, and video and other sensor data from the SDC (and perhaps other sources) confirm the circumstances, the manufacturer ought not to be liable for any harm to the occupants. (In this case, the occupants would need their own travel/accident insurance for their own injuries).

But only if the SDC is following a legislated rule.

This rule should state that the SDC should avoid ‘likely’ injury or death to the party at fault, even at the expense of ‘likely’ minor injury to the occupants – based on the manufacturer’s published thresholds.

If the occupants sustain major injury or death (instead of the expected minor injury), there may be a claim against the SDC manufacturer if it is due to a malfunction in the safety system. Or, if the accident investigation showed that the thresholds were not correct and in fact exposed occupants to much higher risk of major injury and death than warranted.

If the SDC failed to take evasive action and caused injury or death to the ‘party at fault’, the manufacturer would have to show that any evasive action would have most likely caused major injury or death to the occupants or other parties. This could be done via simulation (using data from the SDC’s sensors), based on the published thresholds. Again, this would be subject to the investigation confirming the correct operation of the system and appropriateness of the thresholds.

Any unoccupied SDC of any type should be required to crash in order to avoid hurting anyone. The owner of the SDC would then have a claim against the ‘at fault’ party, whose identity is hopefully captured via its sensor recordings, or perhaps via V2X tracking if they fail to stop.

If any action by the SDC was not deemed reasonable, the incident may result in some liability attaching to the SDC manufacturer and/or other providers of the system components. The benefit being that it would also provide a learning experience to improve the system for all other cars of the same make (and possibly others) – unlike people, where only the driver gets to learn from their experience, if at all!

Crashing into Parked Vehicles to Avoid Harm to ‘Unprotected’ People

From the SDC manufacturer’s viewpoint, a parked car should simply be another object that is tested when setting its impact thresholds.

While drivers do crash into parked cars to avoid hitting (say) a pedestrian or someone on a mobility scooter illegally crossing the road, under the proposed rules, the SDC should do this only if the impact is likely to result in minor injury to the occupants, at most.

As a general rule, if the SDC occupants are likely to sustain only minor injury, that will also be the likely result for anyone in a parked car that it hits. But it may not be the case.

Here it may be best to rely on the balance of probabilities in deciding whether or not to allow an SDC to crash into a parked car. A simple check of parked cars reveals that very few are occupied at any time and then only briefly in most cases.

For someone to be seriously hurt in a parked car, they must be inside it at the precise location and time a car has to take evasive action, where it has time to react and there are no other escape routes and the collision causes the person in the parked car major injury, while only causing the occupants of the SDC minor injury. This will be a very rare event.

On balance, it is likely there will be many more occasions (but still rare), where ‘at fault’ people are saved as a result of an SDC swerving into a parked car than are seriously harmed by being inside the parked car at the time.

Even so, due to their mass and likely consequential damage, it may be unwise to permit any truck or bus (over a certain size – to be legislated) to crash into any parked car (above a certain speed – to be legislated), in order to avoid an accident caused by another party.

With this supplementary rule, a Self-Driving Truck or Bus (above a certain mass – to be legislated) could swerve around you into a parallel lane, or into the opposite lane, or even onto a bike path or footpath – if safe to do so. It could even hit a pole, tree or embankment to save you if the manufacturer’s guide assessed that the occupants (if any) would suffer no more than a minor injury.

This rule would mean that SDT/B (above a certain size) would only be prohibited from swerving into a parked car above a certain speed. Assuming the accident was your fault; if you were injured or killed as a result of the SDT/B being prevented from swerving due to this rule, and it had no other escape route, the SDT/B would have no liability.

Crashing into Operating Vehicles to Avoid Major Injury or Death to an ‘At Fault’ Person

Once again, if any vehicle is unoccupied, it should be treated as ‘property’ for the purposes of applying the rules. In the case of operating vehicles, this would require Vehicle to Vehicle (V2V) communications to broadcast each vehicle’s status. In the absence of a broadcast, any operating vehicle must be assumed to be occupied.

Which raises the last ethical question to be faced by SDCs: the possibility of crashing into another occupied operating vehicle to avoid major injury or death to an ‘at fault’ party.

Clearly, this should only occur where the likelihood of major injury or death to any other people is negligible.

For example, say a bus is just pulling out when a child runs in front, to beat it across the road, but does not see the SDC approaching from the opposite direction, and falls in its path. Assume too that cars line the side of the road so there is no safe escape route.

In this case, if the SDC assesses that even with hard braking it would hit the child, it should swerve into the bus as it pulls out, and brake hard – but only if the occupants were likely to sustain minor injury (given the estimated speed at impact and based on the SDCs thresholds). In this case, there would be little risk to the occupants in the SDC or bus, and the child would be saved.

It would be different if the approaching vehicle was a car (rather than a bus), and the SDC needing to take evasive action was a truck or bus. Again, it may be better to legislate that above a certain mass, and speed at impact, no bus or truck should swerve into any operating vehicle to avoid hitting an ‘at fault’ party.

Determining and Applying Thresholds

Rather than have each manufacturer make a determination in regard to other vehicles, it may be safest to mandate impact speeds (to front, rear, either side and four corners) for different classes of vehicles (small, medium, large, SUV, Van/Small Truck, etc) above which it would be illegal to swerve into them to avoid major injury or death to an 'at fault' party.

The lowest impact threshold would prevail in all cases. For example, if a small vehicle was assessed to have a safe impact speed of 30kph and a large vehicle a safe impact speed of 40 kph, the vehicles would be precluded from swerving into each other (to avoid major injury or death to the 'at fault' party) if the assessed impact speed was greater than 30kph.

These speeds would need to be the subject of extensive research on an industry basis, with a wide margin for error.

The alternative is to outlaw any evasive action that could result in a collision between operating vehicles, no matter their relative speed. This would rule out even low-speed collisions in a car park, where a driver swerves into an oncoming car to avoid a person who has run in front, as well as many other low-speed collisions that happen when people swerve to avoid hurting someone who has absentmindedly appeared in their path (getting worse with the use of mobile phones!)

While it will not be possible to mitigate all possible accidents, ethics would seem to dictate that some attempt should be made to set maximum safe limits for a collision between different classes of operating vehicles, below which there is a 95% chance of nil or minor injury only.

People on foot or on any form of open mobility device, including a bike, skateboard, mobility scooter, motorbike, Segway, golf buggy, etc. would be classed as 'unprotected', where any collision would be deemed likely to result in major injury or death.

Determining Liability

With the availability of rich sensor data; in any accident, it should be possible to know if the SDC system was operating correctly and to use simulation to assess how well it followed the rules.

Rules for Ethical Driving

This paper has suggested that the only facts that can be known for certain upfront by society (for the purposes of setting any ethical rules), and by the SDC in the moments of the accident (in order to follow the rules), are the *roles* of the people involved:

1. The 'party at fault'
2. The 'occupants of the SDC'
3. The 'occupants in other vehicles'
4. All 'unprotected bystanders'

With these parties in mind, the paper has explored an ethical hierarchy that SDC should be programmed to follow (where possible). In summary, it is:

1. Obey the law and drive defensively to avoid all property damage, and harm to animals and people. If that is not possible...
2. Take a path that avoids any collision, at the expense of breaking the law. If that is not possible...
3. Take a path that avoids injury to any party, at the expense of property damage and/or harm to animals. If that is not possible...
4. Take a path that avoids collision with any unprotected bystander, suburban fence or shop front and also avoids injury to any occupants of any vehicle, at the expense of likely minor injury to the party at fault (based on published thresholds). If that is not possible...
5. Take a path that avoids collision with any unprotected bystander, suburban fence or shop front and also avoids likely major injury or death to any occupants of any vehicle or the party at fault, at the expense of likely minor injury to the occupants of any vehicle (based on published thresholds). If that is not possible...
6. Take a path that avoids collision with any unprotected bystander, suburban fence or shop front and also avoids major injury or death to any occupants of any vehicle, at the expense of likely major injury or death to the 'party at fault' (based on published thresholds). If that is not possible...
7. Take a path that avoids collision with any unprotected bystanders, suburban fence or shop front and also avoids likely major injury or death to any occupants of any other vehicle, at the expense of likely major injury or death to the occupants of the SDC and the party at fault (based on published thresholds).

Provided that:

1. Classes of vehicles (to be legislated) may only swerve to avoid likely major injury or death to the party at fault and crash into each other or a parked car at a speed below the lowest threshold (to be legislated by class) applicable to the vehicles involved.
2. Having decided the safest path (based on these rules), the SDC must continue to adjust its trajectory (as best it can) as the relative velocity of each object changes (just as a person may start to steer in one direction and find that something is moving in front, and so steer away)... and keep doing this until the SDC either gets clear or hits something.

These rules have been stated definitively for the purposes of discussion. It is recognised that only through debating the detail will we be able to settle upon a set of rules that is both ethical and practical.

Who's in Control of the Car on the Road to Autonomous Driving?

This paper is not intended to provide either legal or engineering advice. It is aimed at advancing public policy in regard to Automated Driving.

While there are still many technical challenges confronting developers of Self-Driving Cars (SDC), also called Autonomous Vehicles, the two most important social issues yet to be settled are the questions of 'control' and 'ethics' - as they relate to compliance with the road rules, safe driving and responsibility for accidents.

This essay considers one possible approach to the question of 'Control'. The thorny problem of 'Ethics' is tackled in an upcoming post.

Control

There are two sides to control: 1) who or what is 'legally' in control of the car, and 2) who or what has 'actual' control.

Ideally, the same entity (the car or the person) should have both legal and actual control at all times.

Definition of Actual Control

Actual control is taken to mean control of steering at a minimum, combined with the ability to take immediate control of braking and accelerating. Steering is critical to 'actual control' as, in order to steer, the driver is required to pay attention to the road at all times.

As an example, if Active Cruise Control (ACC) is activated, the car can automatically brake and accelerate to maintain speed and/or distance from the car in front, up to the speed limit. However, the person must still steer.

In these circumstances, the person retains both legal and actual control because they can override the ACC at any moment, by touching the brake, accelerator or cruise control knob.

Separately, Advanced Driver Assistance Systems (ADAS) such as 'Forward Collision Avoidance' (FCA) and 'Lane-Keeping' (L-K) can be activated in the background, as safety features.

Forward Collision Avoidance

The FCA is there solely to assist the driver. Its purpose is to stop the car before a collision; if the driver fails to react in time.

If the FCA fails to respond as warranted, the driver may have a claim against the manufacturer and/or maintenance provider. This would be a civil matter. It could also trigger a recall, depending on the nature of the failure.

The manufacturer (or maintenance provider) would have no liability to third parties for any failure in the FCA. The reason is simple: the driver remains in legal control of the car at all times, and should not drive in a way that allows a crash to happen that is their fault (eg running into the back of the car in front).

If a collision occurs, the driver would be liable for all damages (subject to any mitigating circumstances).

If a collision is avoided by the FCA, immediately the car is stopped, all functions should be returned to the control of the driver in order to proceed.

Lane-Keeping

In the case of lane-keeping, the system may first give a warning to the driver that the car is drifting to the edge of its lane, and then take active control of steering to bring it back on to the centre-line - if the driver does not respond to the warning.

To ensure the driver remains in actual control, once back on the centre-line, the system should immediately return control of steering to the driver, ideally with a verbal warning that they are in control of the vehicle and must steer at all times. This should be an industry standard process.

Conversely, the driver must be able to over-ride lane-keeping (to actively drive into another lane), say by using the indicator, or touching the brake or accelerator as they veer away. For example, to avoid a potential collision with an object in its path that has not been picked up by an FCA system that is still in 'beta' mode. This capability ensures the person in the driver's seat has actual control (as well as legal control) at all times. Again, this should be an industry standard process that ensures the driver is making the manoeuvre intentionally.

System may take Legal and Actual Control

As a further safety measure (as GM is proposing with its 'Super Cruise' system), if after 'x' times or 'y' seconds (which should be an industry standard number in each case), the driver fails to take back and maintain control, or if the driver monitoring system determines the person is no longer alert (ideally, also in accord with some industry standard); the car could pull over and stop when it is safe to do so.

At the time that the car takes over full control (to stay in its lane or pull over), the system would have actual control and ought to then have legal liability for compliance with the road rules and for any accident during the time it remains in actual control. This puts the onus on the manufacturer to ensure its system is safe to carry out such manoeuvres, including signalling and safe merging. Unless and until it accepts this liability, it should not be permitted to offer the system as an option.

Traffic Violation and Automatic Reporting

Any time the system has to take actual control of steering, accelerating and braking to pull over, it ought to be a major traffic violation for the human driver. (The system only takes over when the driver has illegally relinquished actual control by letting the car drift in its lane after repeated warnings and/or a specified period of seconds).

Perhaps too, the violation should be automatically reported by the system to police, together with the vehicle's location. As well, the car may be disabled for, say, 30 mins?

Depending on the reason for the loss of control (e.g. biometrics indicate a heart attack, stroke or loss of consciousness), an ambulance could also be summoned.

In-car cameras and phone videos are already being used in evidence, so having the car automatically report the loss of control (after repeated warnings/time) would seem a logical extension of this trend.

This may be contentious on privacy or civil liberty grounds, but a dangerous driver is breaching the civil liberties of everyone else on the road (by endangering their lives), and ought to be liable for that breach. If there is no breach, there is nothing to report. (Reporting specific breaches of the law is very different to central monitoring of all drivers at all times. While it can be argued central monitoring does breach civil liberties; in cities at least, it appears inevitable to eliminate congestion, but that is a separate issue).

An argument against automatic reporting is that it would discourage our worst offenders from using ADAS. However, there is no reason why FCA, L-K and 'Super Cruise' systems (including automatic reporting of loss of control) could not be made mandatory following a serious traffic violation - as an alternative to loss of licence; especially important where the person needs to drive to earn their living. Mandatory breath test interlocks for drink-driving offenders are already required in Australia.

Current Guidelines for ADAS Permit Separation of Legal and Actual Control

The current US NHTSA/SAE guide includes 0-5 levels of automation, with L5 being full automation in all circumstances.

Unfortunately, it is possible for the driver to have legal control, but not actual control at Level 2. This arises where both centre-line lane keeping and active cruise control are activated together.

At this level, the person is still required to pay attention to the road and to what the car is doing. However, they may have no role in actually driving the car, and may even be able to 'let go the wheel' for a period of time in some cases. It means, in effect, that the car is driving itself, and the person is assigned to a passive monitoring role. On an interstate freeway, this state of affairs could last for hours.

It is well recognised that people are very poor at passive monitoring tasks. And, the better the technology is at driving and the longer it keeps control without intervention, the more likely people will be lulled into a false sense of security (it is 'false' if the tech cannot be relied on to handle all eventualities while activated).

This seems to be the most likely cause of the Tesla accident on 7 May 2016 where the car ran into the side of a truck that had turned in front of it on the freeway. Had the driver been required to actively steer the car, it is most likely he would have kept his eyes on the road, realised the car was not braking, and done so himself.

It is difficult to see any safety advantage (for the driver or the community) in being able to take your hands off the wheel and (due to human nature) your eyes off the road, while you remain in legal control of the car. An accident can occur in just a few seconds of inattention.

Hands-free driving at L2 appears to be a dangerous novelty. As it clearly separates legal control from actual control, it ought not to be allowed.

Applying the Lessons Learned

Tesla seems to have learned its lesson and has now implemented its latest technology to operate in the background, without actually controlling the car. It means the company can still gather all the data it needs to refine its systems (by comparing the driver's actual responses with what the system would have done had it been in control), without imperilling the driver, or anyone else.

Once Tesla is satisfied that its systems can operate safely (within specified areas and conditions), it is to be hoped that they (and the law) will accept that as soon as the car has actual control of all dynamic functions (steering, accelerating and braking) that it also has legal control.

Benefits of Aligning Legal and Actual Control

With lane-keeping (but not lane-centring) combined with active cruise control and forward collision avoidance we get the best of all worlds until full SDCs are released:

- Competent drivers are not lulled into relying on beta tech that could fail to keep them safe; because to steer, they must remain on alert, ready to brake or accelerate as needed.
- Delinquent/impaired drivers are protected from their own bad behaviour, and more importantly, so is everyone else protected when they fail to keep control of steering and braking.
- Other drivers and, as importantly, the police are alerted to the fact that the person is no longer in control of the car (as it wanders inside its lane for a short time, before pulling over), so both can take appropriate action. In the case of other drivers by giving the car a wide berth, and in the case of the police by apprehending the delinquent driver. Also, medical help could be summoned if the biometrics indicate ill health is the problem.
- Car companies can continue to test the full range of their driverless tech without putting drivers or anyone else at risk due to 'passive monitoring syndrome'.

Once testing demonstrates they can safely accept full liability, manufacturers can gradually expand the areas and conditions where their cars are rated to operate in 'autonomous mode' - providing all the benefits of SDC, without incurring undue risk in the process.

Obviously, this takes away a bit of 'fun' from the semi-autonomous driving experience. However, this technology should be first and foremost for safety.

New Regulatory Guideline

It would seem that for safety's sake, as well as to limit lawsuits, it would be better to dispense with 'levels of automation' and recognise only two legal modes of operation:

1. Driver Mode: Driver in Legal and Actual Control (of steering as a minimum, as outlined above, combined with zero to any level of ADAS support)
2. Autonomous Mode: System in Legal and Actual Control (within specified areas and conditions, from highly restricted to unlimited)

It is then up to each manufacturer to determine in regard to 'Driver Mode', what combination of ADAS support they will provide by model; and in regard to 'Autonomous Mode', the areas and conditions in which any of their cars may operate.

Autonomous Mode is when the 'Fun' Starts

It is only after the car has both legal and actual control that the 'fun' should start; with no requirement for the person to pay attention to the road, or the operation of the car.

A person should only need to become re-engaged with the driving task when the car warns that it is approaching the limit of the area or conditions where it is rated to operate (e.g. say, it is only rated to operate on freeways and is approaching the exit, or it appears that snow is likely, and it is not rated to operate safely in the snow, or it encounters an unfamiliar situation that forces it to slow or stop to allow the driver time to respond).

Managed Change in Control

Except where the system takes over control in an emergency (eg to avoid a forward collision), any change in control (from the person to the car, and the car to the person), ought to be a 'managed' process, with clearly defined steps (much like a handover between pilots in an aircraft).

The 'handover' process should be seen as 'safety critical' and developed as an industry standard, so any person is able to drive any autonomous car in manual mode (if they have a driving licence) without having to familiarise themselves with a new handover process each time they get into a different car.

Ideally, drivers should know how long they have before they are expected to be ready after a warning that they will be required to take back control, what steps they have to take to resume control, what tests the car will carry out to determine that they are capable of taking control and that they do in fact, have control on handover; and what happens when they are deemed to be incapable, because they are (say) too sleepy, or drug or alcohol affected, or otherwise impaired (eg taken ill, etc).

Standard Safety Processes

Ideally, the industry and regulators should be collaborating to design a set of standard processes relating to vehicle control:

1. No vehicle should have the capability to take full control of all dynamic functions without assuming legal control; at the same time releasing the person from any requirement to monitor the road or the car.
2. To ensure the driver remains in actual control of steering, if for safety's sake, the system needs to re-centre the car in its lane, the system should immediately return control of steering to the driver, ideally with a verbal warning that they are in control of the vehicle and must steer at all times.
3. If after 'x' times or 'y' seconds (to be regulated numbers), the driver fails to take back and maintain control, the car pulls over when safe to do so.
4. Driver 'alert' tests based on monitoring biometrics and/or change in driving behaviour that require the car to take control and pull over.
5. Automatic call the police and/or for help when the car is forced to take control and pull over.
6. Except where the car takes over emergency control, the managed handover process from driver to car, and car to driver, at either the request of the car, or the driver.