A Plan to Develop Safe Autonomous Vehicles. And Prove It.

By Professors Amnon Shashua and Shai Shalev-Shwartz

The scientific knowledge and tools are now in place to develop and deploy fully autonomous vehicles (AVs) that operate without drivers in all situations (SAE Level 5). When implemented en masse, these vehicles have the potential to produce tremendous economic and societal benefits including greatly reduced traffic deaths, injuries, and congestion and to make less expensive, more flexible, and more productive transportation available to all.

But autonomous transportation is not just a product; it is an industry. For it to truly improve mobility, automakers, technology companies, and government regulators must come together to develop a holistic model to enable this industry.

At Mobileye, an Intel company, we believe we have developed an innovative framework for that model. It is centered on two fundamental pillars: 1) provable safety assurances and, 2) economic scalability. Without a clear model for these two critical parameters, all the investment in AVs runs the risk of ending up as simply a very expensive science experiment.

Consistent with Mobileye's philosophy of collaboration, we illustrate and open up technical details of this model in an academic paper recently published here. And today we publish this post to provide an overview of the Safety Assurance pillar of the model for a wider audience. We will illustrate the concepts of the Economic Scalability pillar in a future post.

Overview of the Plan

We believe that it is important for the automotive industry to collaboratively establish a methodology and standard for safety validation in partnership with global standards–bodies and regulators. The United States is among the countries leading the way with pending self-driving vehicle legislation and new USDOT Automated Vehicle Guidelines, making this a perfect time to engage in these collaborative next-step discussions.

Our proposed model provides a detailed, practical, and efficient solution for designing and validating an AV system that results in drastically improved safety. Here is an outline of the next-step areas we believe merit attention and the solutions we propose:

A predetermined set of rules to rapidly and conclusively evaluate and determine responsibility when AVs are involved in collisions with human-driven cars

- AVs will share the road with human-driven vehicles for decades. In order for AVs to unlock the safety, mobility, and traffic efficiency benefits mentioned above, they need to operate in a "normal" way (i.e., not restricted to very low speed or restricted to stay far away from other vehicles).
- Figure 1 illustrates the point. If one of the human-driven cars on the outside makes an error and cuts in on the AV, there is no way for the AV to avoid a collision. But the situation happens all the time, so if we forbid the AV from being in that situation, it will be useless. The conclusion is that any useful AV will be involved in situations that may lead to accidents, including mechanical failure and external forces. Complete avoidance of every accident scenario is impossible.

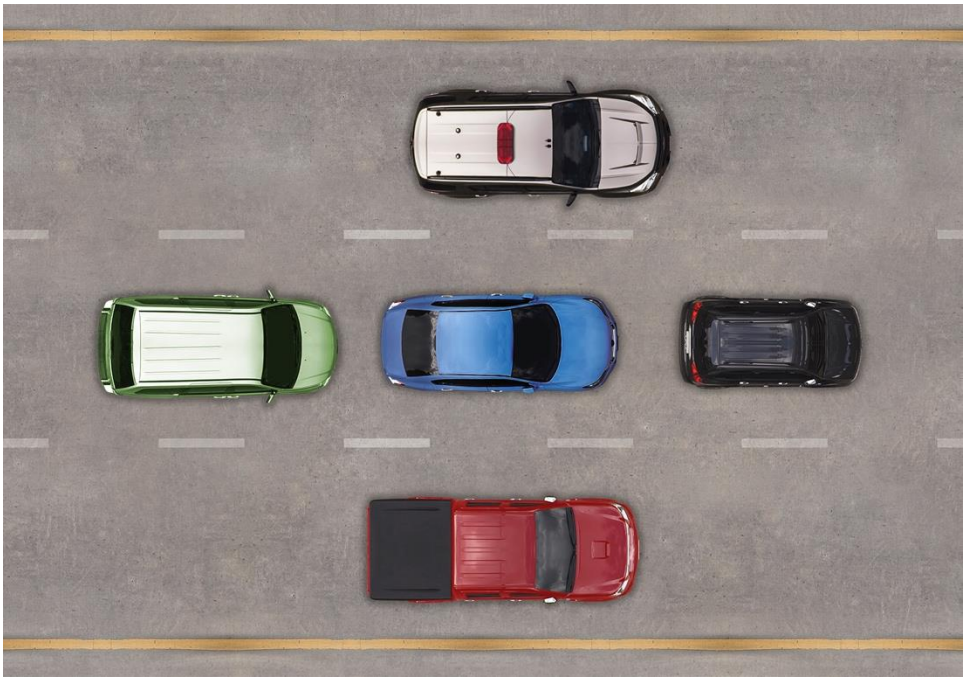

Figure 1. The central car can do nothing to ensure absolute safety.

- What will happen when a collision occurs? There will be an investigation, which could take months. Even if the human-driven vehicle was responsible, this may not be immediately clear. Public attention will be high, as an AV was involved.
- Our solution is to set clear rules for fault in advance, based on a mathematical model. If the rules are predetermined, then the investigation can be very short and based on facts, and responsibility can be determined conclusively. This will bolster public confidence in AVs when such incidents inevitably occur and clarify liability risks for consumers and the automotive and insurance industries.

- Here is how it is possible to do this:
    - Our approach formalizes the "common sense" behind who is responsible for an accident. It describes how we can formalize the set of driving scenarios, concepts of priority and give way/take way, and equations involving speed, distance, etc, that combine into a formal mathematical model for determining fault.
    - Since the AV has a set of highly accurate sensors, there will be data available to evaluate the exact environment before and when the collision occurred. In combination with the formal rules for determining fault, this data can be used to quickly, conclusively determine responsibility.

The formalized model for fault allows for the creation of decision-making (driving policy) software that avoids accidents caused by the AV system. It also enables validation in an efficient manner.

- AVs have 360-degree vision and lightning-fast reaction times, can analyze road conditions and available braking power, and are never distracted. Given this, plus the formalized model to determine fault, AV developers can design a system where the software can evaluate every command against this model.
- Mobileye has designed such a system, and it is called Responsibility-Sensitive Safety (RSS). RSS ensures that, from a planning and decision-making perspective, the AV system would not issue a command that would lead to the AV causing an accident.
- This system avoids the data-intensive validation process that most AV developers seem to be planning, which we see as not feasible (whether performed on-road or in a simulated environment). It is simply validated by proving that the system evaluates all commands against the predetermined set of mathematical rules.

How safe is safe enough? Our system can validate three orders of magnitude improvement to one traffic fatality for every one billion hours of driving vs. the human-driven vehicle rate of one traffic fatality for every one million hours of driving (i.e., a US traffic fatality rate of ~40 per year compared to ~40,000 in 2016).

- We don't believe society will accept machines causing accidents unless a drastically reduced fatality rate can be proven.
- While the RSS decision-making software is designed to not allow decisions that would result in an accident that would be blamed on the AV driving policy, there could still be accidents caused by mistakes of the sensor system (i.e., the information about the driving environment that is used to base decisions) or mechanical failure.
- We propose a Sensor Fusion system that includes three independently engineered systems, each relying on different technologies: 1) camera, 2) high-definition map, and 3) radar and lidar.
- The system can be validated to a miniscule error rate (supporting the rate mentioned in a heading) with a very reasonable data set of real-world miles driven—specifically 100,000 hours.

Details of Our Approach to Safety

At its core, RSS is a pragmatic method to design and then to efficiently validate the safety of an AV's planning/decision-making (driving policy) software.

As absolute safety is impossible as long as AVs share the road with human drivers, we start by defining an AV as safe if it doesn't cause collisions.

In order to prove that the AV won't cause collisions, we must first construct a mathematical model that determines blame when collisions occur, using a comprehensive set of driving scenarios. Essentially, we formalize the "common sense of fault" using mathematical equations. We can set these equations with certainty as the AV will have known reaction times and known ability to make evasive maneuvers through steering and braking.

The powerful result of formalizing this model of blame is that it enables us to construct the decision-making software of the vehicle to follow these exact parameters, continuously evaluate these rules, and never make a command that would put the vehicle at risk of causing a collision.

In practice, the AV needs to know two things:

- Safe State: This is a state where there is no risk that the AV will cause an accident, even if other vehicles take unpredictable or reckless actions.
- Default Emergency Policy: This is a concept that defines the most aggressive evasive action that an AV can take to maintain or return to a Safe State.

We coin the term Cautious Command to represent the complete set of commands that maintains a Safe State. RSS sets a hard rule that the AV will never make a command outside of the set of Cautious Commands. This ensures that the planning module itself will never cause an accident.

Following Scenario

To illustrate, let's take a common example. Consider the case of two cars driving in the same lane, one behind the other, along a straight road. If the front car brakes and the rear car cannot brake in time, the accident is the fault of the rear car. As seen in the figure below, we can calculate the exact distance corridor that the following car (blue) must leave to the lead car (red) to be in a Safe State. This calculation depends on data for certain variables, such as the difference in velocity between the two cars, which will be provided with high accuracy by the various sensors in the AV.

Safe Distance Formula

$$d_{\min} = L + T_f\left[v_r - v_f + \rho\left(a_a + a_b\right)\right] - \frac{\rho^2 a_b}{2} + \frac{(T_r - T_f)(v_r + \rho\, a_a - (T_f - \rho)a_b)}{2}$$

- L is the average length of the vehicles
- ρ is the response time of the rear vehicle
- v_r, v_f are the velocities of the rear/front vehicles
- a_a, a_b are the maximal acceleration/braking of the vehicles
- T_f is the time for the front car to reach a full stop if it would apply maximal braking
- T_r is the time for the rear car to reach a full stop if it would apply maximal acceleration during the response time, and from there on maximal braking

Figure 2. The above formula calculates the safe longitudinal distance between the rear vehicle and the front vehicle.

Cut-In Maneuvers

The above is a fairly simple scenario. It captures the intuitive principle that if a rear car hits a front car, it's always the rear car's fault. But what happens if the front car performs a reckless cut into another car's trajectory and the rear car then hits the front car from behind? In that case, it's the front car's fault.

RSS is well equipped to deal with this scenario using the same principles. Based on a set of variables, a safe corridor exists around the AV. If the human-driven vehicle (red car in figure below) violates that corridor before the collision occurs, it is that vehicle's fault. Conversely, the AV can continuously calculate the safe corridor around other vehicles and will never make a command that violates that space.

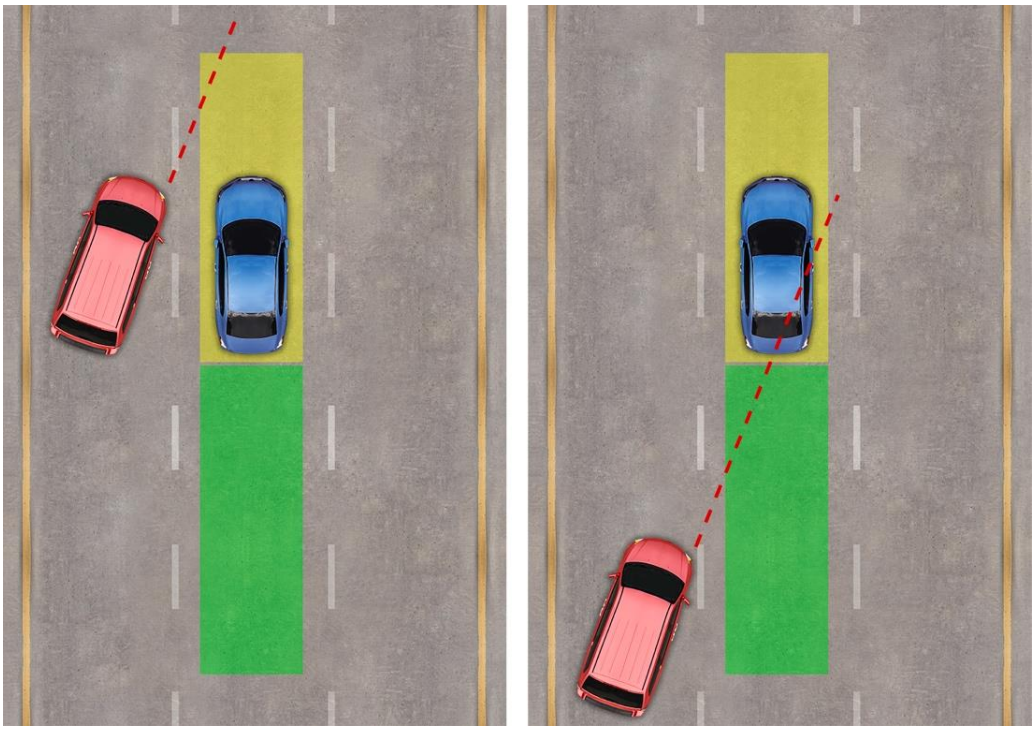

Figure 3: Calculated safe corridors around the vehicle will determine fault in the case of cut-in maneuvers.

Dealing with Limited Sensing and Occluded Objects

A very common human response, when blamed for an accident, falls into the "but I couldn't see him/her" category. A common answer is "well, you should have been more careful." A crucial component of RSS is that it formalizes what it means to be careful with respect to situations where sight is limited.

Consider the below figure. C_0 is trying to exit a parking lot onto a street where the view of traffic is blocked by a building. A human driver's behavior is to slowly merge onto the road, obtaining more and more field of view, until sensing limitations are eliminated. But there are assumptions that the driver makes, such as the likely speed of the traffic and at what point the traffic flow will be exposed. RSS calculates an assumption of the highest reasonable velocity, called V_{limit} (dynamic based on speed limit of road). With this information, if C_0 is the AV, it knows how much it can peek out and still give C_1 an opportunity to brake. If C_1 is going faster than V_{limit}, and a collision occurs, it is C_1's fault. Conversely, if C_1 is the AV, it will drive more cautiously and farther toward the center lane on the assumption that there may be a car that it can't see looking to exit the parking lot. In either case, the rules are clear, and the AV will not take a command that would put it at risk of causing a collision, even with an object that it can't see.

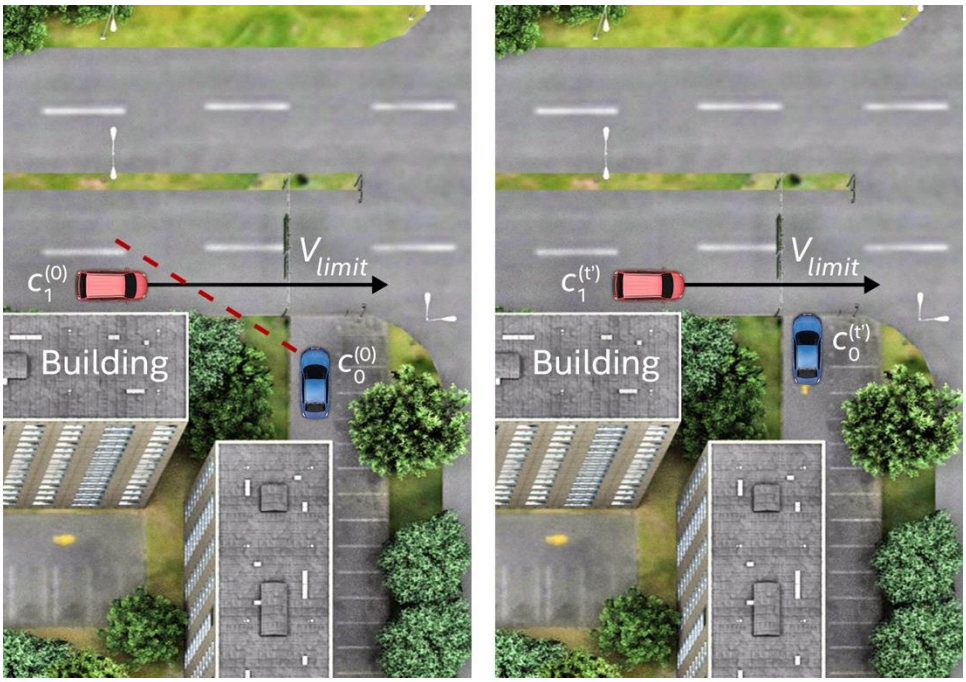

Figure 4. RSS encompasses situations where objects (vehicles, pedestrians, etc.) are occluded (hidden) by other objects.

This particular type of scenario is also very relevant to pedestrians. The blame in an accident with a pedestrian is almost always on the car, even in the case where the car did not see the pedestrian until very late. But there are exceptions. For example, under RSS, the AV accounts for the possibility that a pedestrian could dart out from between two parked cars, at some maximum speed V_{limit}. RSS will manage the AV's speed so that it can always brake in time to avoid a collision with the pedestrian, unless that pedestrian is at a velocity above V_{limit} (e.g., if the pedestrian was running super-humanly fast).

Extending RSS to All Road Structures

We believe that the formal model of blame that we've constructed (and the associated RSS software) covers a comprehensive set of driving scenarios (a large range is included in the appendices of the white paper). We also welcome engagement with regulators and standards bodies that would likely lead to additional scenarios to evaluate. Below are a few examples (by no means exhaustive) of additional scenarios that RSS considers:

Route Priority – This important concept enables RSS to cover situations that involve multiple road geometries that overlap in a certain area, such as highway merge lanes and roundabouts. In these cases, both vehicles must perform a cut-in to the frontal corridor of the other in order to merge. In this scenario, the vehicle in the priority lane is allowed to make the cut-in without violating RSS. Which route/lane has priority can be determined through traffic signs/regulations, which will be incorporated into our high-definition map called RoadBook™.

Two-Way Traffic – In two-way traffic situations, the center line provides a boundary, but this is naturally violated when one car crosses the center line to overtake a parked truck or to reverse into a parking spot. The white paper includes a section on how RSS covers these types of situations.

Traffic Lights – Intersections that are equipped with traffic lights are modeled using the concept of dynamic route priority. In other words, some of the routes have priority over others according to the traffic lights.

Unstructured Roads – RSS also covers scenarios involving unstructured areas like parking lots. The key concept here is predictability. An AV, like a human, can predict the path of a vehicle in an environment with no lane markings at all based on current trajectory. If a vehicle deviates from this prediction (by changing heading), it must be done with a long enough distance to allow the oncoming car to adjust as well. That distance can be calculated and if the vehicle changing heading violates that distance, it is at fault.

Conclusion

Mobileye has always tackled challenges through innovative engineering but also with a keen eye toward the ability to validate safety and enable products at a cost the customer can bear. This is a key reason the vast majority of the world's automakers have chosen us to supply today's cutting-edge life-saving safety systems, such as Automatic Emergency Braking. We now come together with Intel, a leader in the areas of scalable, purpose-built compute platforms, custom data centers, cloud computing, and next-gen communications to significantly broaden ADAS functionality and to enable AVs.

Intel and Mobileye have already begun work with BMW on a nonexclusive platform for AVs that follows these concepts. It is an example of an inclusive relationship of partners that each share a vision to put safety first.

We also clearly understand that without an economic model that is truly scalable, the real potential of AVs will never be experienced by the masses. Please stay tuned for our next post dealing with the issue of economic scalability for AVs.

For more detailed information on our model, please see the white paper by Professor Amnon Shashua, Professor Shai Shalev-Shwartz, and Shaked Shammah.