"How Good is Good Enough?" in Autonomous Driving

Hans-Peter Schoener
"Insight from Outside" - Consulting
66 PUBLICATIONS **295** CITATIONS

“How Good is Good Enough?“ in Autonomous Driving

Dr. Hans-Peter Schöner[1]

Abstract

A new approach for quantifying „good enough“ for the behaviour of autonomous vehicles (AVs) is presented here. It is deducted from general behaviour guidelines for (human) behaviour in traffic, and especially the wording of §1 of the German Traffic Code (“Straßenverkehrsordnung, StVO”) is taken into account. The German traffic code generally requires constant *caution* and *mutual consideration*; in a second paragraph, it more specifically requires, that “no one is **harmed**, **endangered** or unnecessarily **hindered** or **bothered**.” The here presented approach proposes a way how to specify and finally quantify those key words. These quantifications can be derived from normal (i.e. safe) traffic practice, which has ever since been established based on the human capabilities to cope with traffic situations.

First, this quantification defines the guidelines for a safe own behaviour of an AV, for normal (avoid hindering or bothering others, but never endanger them) and extraordinary (never harming others) traffic situations. It establishes behaviour rules and thus testable performance guidelines for a **single vehicle**.

Second, based on the fact, that *everybody* should comply with §1, certain behaviour of traffic partners can be trustfully *expected* in any **interactive traffic situation** (even an AV should not be harmed or endangered by any other traffic participant), but a safe and robust reaction on (rare) expectable situations (collision free, if only being hindered or bothered by other participants) needs to be ensured. The quantifications can help to define the functional design space for AV behaviour in normal traffic scenarios, with the goal of “not endangering” traffic partners. This leads to more robust behaviour than the “no collision” goal for extraordinary situations. The concept allows to deduct testable performance goals in reference situations, with reasonable passing criteria.

The quantification of the key words is mainly based on the **reaction times** for safety-relevant actions in traffic scenarios, in combination with **manageable reaction patterns**. Human reactions are generally limited by necessary perception, interpretation, reasoning and action times. Thus, the requirement for AVs must ensure equal or better total performance. This cannot (and is not required to) be tested in every conceivable situation; but typical traffic reference situations must be agreed on, which define borderline cases. Some outer conditions must be summarized by testing for the extreme case. Any testing procedure should include verification of the ability for autonomous stopping; for the worst case of total sensor loss under extreme weather conditions, the **“Blind stopping procedure”** is proposed.

It must be accepted that there exist conceivable situations, in which a single traffic participant will not be able to avoid an accident (loss of own controllability), once a situation has evolved to a certain criticality. The quantifications help to draw a limit line between “avoidable” and “unavoidable” (force majeure, natural disaster) accidents. In order to stay away as much as possible from getting into such uncontrollable situations, an AV needs to demonstrate a perception of “**level of danger”** and a sense for its own capabilities (“**self-awareness”**). Testing for these skills should be part of release and certification procedures.

[1] Insight from Outside – Consulting, www.ifo-consulting.com

1. Background and Motivation

Since several years an international discussion on the testing, verification and certification of autonomous vehicles (AVs) is going on [1], [2]. In Germany, the PEGASUS-project [3] is on the way with the goal to define “how good is good enough?” and “how can we verify this?” with respect to the introduction of AVs. Internationally, many other activities and conferences are dealing with this topic.

One of the key reasons for the long-lasting discussion is the insight, that the often-announced goal of an accident-*free* road traffic with autonomous vehicles is theoretically impossible to reach: since there are many independently acting traffic participants, and since the traffic is exposed to uncontrollable environmental conditions, it cannot be designed to be 100% accident-free [4]. Assumptions for the operational design space of AVs need to be made, in order to come to a feasible design solution, with much fewer accidents to be expected than in the actual purely human-controlled road traffic. These assumptions imply, however, that there might be accidents, which will occur predictably under certain (rare, but conceivable) conditions. Product liability legislation requires, that nobody should provide a product which is undue unsafe by its nature; in order to resolve this apparent contradiction, a generally accepted guideline for designing a *sufficiently safe* AV behaviour in traffic is indispensable.

One important aspect is the interaction of traffic participants. A safe traffic can only be designed if all participants comply to certain rules. Recently, in [4] and [5] a complete set of formalized design rules, which mimic human behaviour in traffic, are published; the design rules are called “Responsibility-Sensitive Safety (RSS) model” by the authors. They explicitly state, that “the RSS model contains parameters whose values need to be determined through discussion with regulatory bodies and it would serve everyone if this discussion happens early in the process of developing autonomous vehicles solutions”.

The approach discussed in this paper has been independently developed from the RSS model, but the ideas are close to each other and complement each other. This paper strives to provide a new view on how to define the parameters for defining “good enough”, being derived from normal traffic practice and specifically based on §1 of the German Traffic Code (“Straßenverkehrsordnung, StVO”), which is the general guidance for (human) behaviour in traffic. Besides this, this paper gives suggestions on how to test AV skills and monitor AV behaviour in order to provide evidence of safe behaviour for release and certification purposes.

2. The Goal of Defining “Good Enough” in this Context

This paper focusses on the creation of guidelines for fair dealing with rare challenging situations caused by the environment of AVs. These situations present a continuous spectrum from easy manageable to impossible to control, even for human drivers, but also for autonomous vehicles.

They can be divided specifically into two categories:

- suddenly emerging, challenging environmental situations (“emergencies”)
- extreme and/or unexpected behaviour of other traffic participants
  (behaviour outside of “law and order”)

These challenges do not derive from specific technical solutions, but they are caused by the properties of the application “public road traffic”. They deal with the situations which have “to be expected” by the traffic participants including autonomous vehicles. The guidelines should define a minimal operational design domain, ODD.

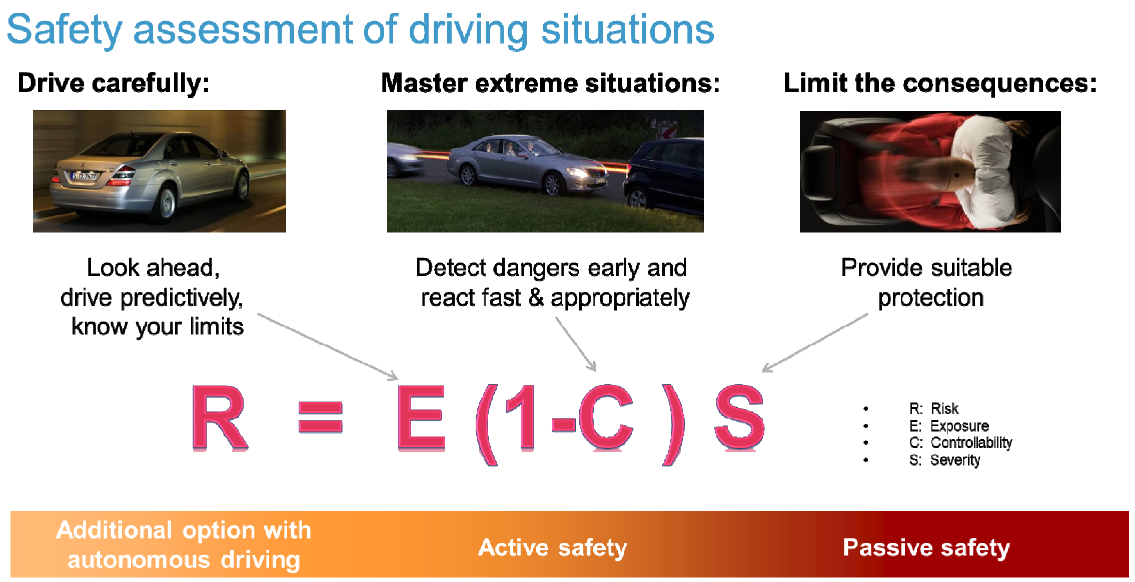
Thus, this paper does not aim to give guidelines for

- completeness and robustness of technical solutions,
- suitability of user interfaces, etc.

These topics are related to the specific technical design solution. This should be left to the providers of such functions; competition should aim to develop the best solution within the design space. And it should not hinder anybody to provide solutions which even might expand the minimal ODD.

As already pointed out in [1], the risk of dangerous traffic situations derives from three factors (as for any other accident types), see figure 1:

- S: severity of an accident, once it has happened
- C: controllability (avoidability of an accident), if an accident-prone situation comes up
- E: exposure to accident-prone situations



Safety assessment of driving situations

Drive carefully:

Look ahead, drive predictively, know your limits

Master extreme situations:

Detect dangers early and react fast & appropriately

Limit the consequences:

Provide suitable protection

$$R = E\,(1\text{-}C\,)\,S$$

- R: Risk
- E: Exposure
- C: Controllability
- S: Severity

Additional option with autonomous driving | Active safety | Passive safety

Figure 1: Risk results from exposure to, controllability of and severity of accidents [1]

Driver Assistance Systems DAS (up to level 2 of vehicle automation) have been developed in the past, in order to help the driver to control critical situations (bring the factor C closer to 100%). But they always left the task to the driver to avoid coming into safety critical situations (no influence on the factor E). Since in level 2 systems, most systems are designed to not interfere with the driver unnecessarily, the intervention of the DAS thus was limited in most cases to situations, when an accident is getting almost unavoidable. For this reason, there has been a big focus on very intense interventions (emergency braking) of the driver assistance system, striving to increase the controllability C of the situation. Autonomous vehicles, however, have a chance to perform automatic vehicle reactions in a much earlier state. This allows to intervene in a more comfortable way, which finally has the effect of reducing the exposure E to truly critical traffic situations. For autonomous vehicles of level 3 and higher, only reducing the exposure E to critical situations becomes the final reason for reducing the risk of accidents, since the controllability C should be already inherited from level 2 systems.

3. Limits of Controllability

Figure 2 shows the typical characteristics of controllability distributions in a critical situation [1]: the red line shows the controllability of unassisted human drivers, the green line the curve of assisted human drivers, and the blue line the curve of automated collision avoidance systems. All three curves show, that there is a minimal reaction time which limits the controllability; typically, automated systems can react faster than human drivers, but they still cannot anticipate accidents. If the available reaction time is shorter than the necessary reaction time, controllability C goes down to zero because of physical reasons. In practical traffic this can be cargo falling off from a preceding vehicle, or an object being dropped from a bridge.

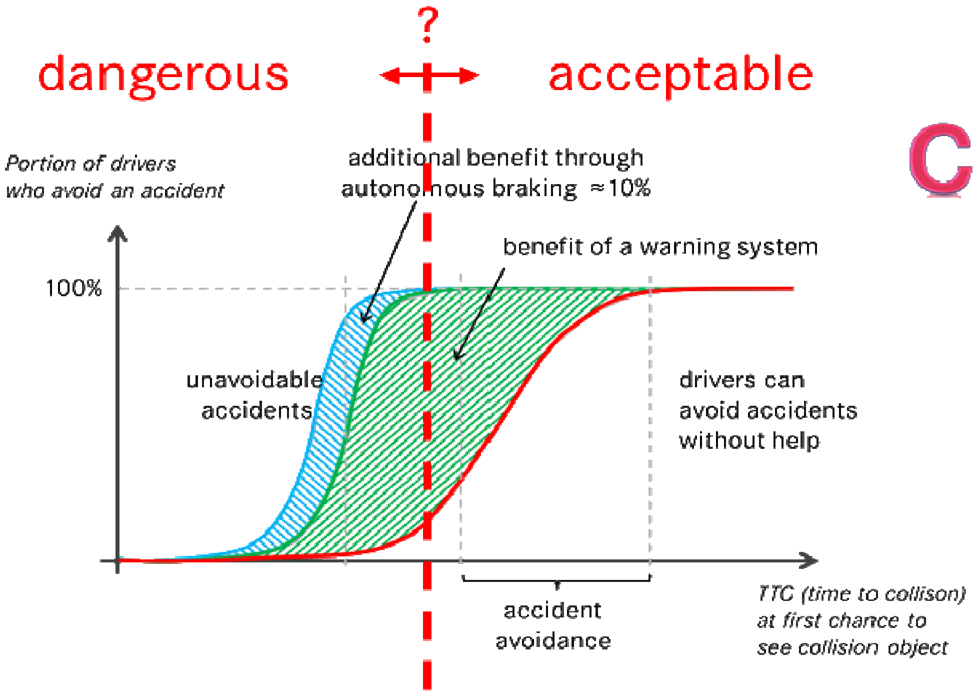


Figure 2: Limits of Controllability [1]

It must be accepted that there exist conceivable situations, in which a single traffic participant (no matter if human or AV) will not be able to avoid an accident, once a situation has evolved to a certain criticality. A situation which becomes close to being uncontrollable, is considered dangerous. An agreement on a value for a limit in the available reaction time helps to draw a line between “avoidable” and “unavoidable” (force majeure, natural disaster) accidents.

The only way to avoid accidents caused by such uncontrollable situations is, to make sure to never come into such critical situations (which is equivalent to limiting the exposure E). However, there are two main categories of situations which a single traffic participant cannot rule out by himself.

3.1 “Emergencies” in Traffic Environment

In real world, sudden extreme situations can emerge; for this reason, they are called “emergencies”. They are not to be expected in normal daily operations (low exposure E), but they can have extreme consequences (high severity S); thus, their risk (without controllability) might be significant.
In traffic, such unusual outer conditions might be:
unexpected severe weather, landslides, road and bridge failures, sudden technical vehicle faults, terroristic or suicidal actions, lost cargo on the road, …
Even the best environmental sensing methods cannot provide practical solutions for such risks. For every technical solution (as complex it might be) there is always an imaginable situation which will still lead to undesirable consequences. Commonly accepted methods to cope with such risks are a necessity for a “reasonable safety level” of AVs.

3.2 “Law and Order” in Traffic Behaviour

Traffic is governed by general laws and behaviour rules (order). If everybody would comply to all laws and rules, the behaviour of other traffic participants would be relatively easy.

Traffic laws are considered unconditionally valid. So, there is no question that autonomous vehicles have to adhere to the laws. And the AV planning software can assume in planning its own behaviour that every other traffic participant is adhering to the laws as well.

Traffic rules, however, regulate the *orderly* and thus *safe* behaviour of traffic. Normally, everybody should stick to those rules as well. But in practical traffic, the rules are often not obeyed completely: for practical reasons in order to solve unusual situations, to ensure traffic flow (e.g. if a lane is blocked, even a solid line as lane marking needs to be crossed), or just by loose interpretation of the rule (how to stick exactly to a speed limit), or by mistake. Thus, traffic rules cannot be used as limits for the assumed behaviour of other traffic participants.

But what kind of behaviour can you really expect from other traffic participants? If you always consider worst case behaviour, traffic probably would collapse by over-cautiously driving participants. Some roads are designed and operated specifically (freeways) in order to reduce the variance of behaviour to be expected by other traffic participants on these roads. Guidelines (derived from those traffic rules, and dependent on road classes) are needed for the design of interactive AV behaviour.

In German and Austrian jurisdiction [11] – for example – a concept of “trust” has been established, which defines certain risks as acceptable: a traffic participant can trust in the fact, that other traffic participants do obey all relevant rules – he is not obliged to consider any unusual endangering misbehaviour when deciding about his own actions within the given rules. (However, if indications of misbehaviour of others are observable, he still must react by mitigating the risks from this misbehaviour.)

4. Safe Behaviour in Traffic

All over the world, driving behaviour in traffic is learnt when getting a driver’s license. Safe behaviour is typically defined in very general abstract rules, but it is learnt intrinsically by practical training in traffic situations. Thus, in order to teach an AV a safe driving style, we cannot build on a given and complete set of exact rules.

Generally speaking, safe behaviour depends on many parameters (see figure 3): There are controllable factors, which can be influenced by the driver, namely: driving speed, longitudinal distance and lateral distance to other objects. And there are uncontrollable (independent) parameters, namely: weather conditions, traffic conditions, motion state of other objects, road curvature and lane markings, safety infrastructure along the road, and many more.

Driving safely means to use the controllable factors in order to stay away from critical situations (keep the exposure E small), which might emerge from complex driving situations; these factors have to be adopted constantly if easy situations turn to more difficult ones by changing outer conditions (see red arrows in figure 3). Safety depends also on limitations due to own driving abilities, which might be related to visibility, driving dynamics skills, tire conditions, etc. In summary, staying away from accidents and even severe accidents requires the recognition of a **level of danger** of the current situation, considering all those factors.

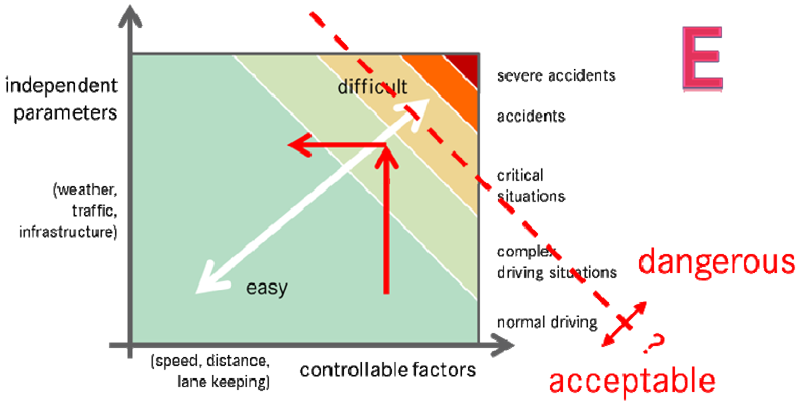


Figure 3: Safe behaviour for avoiding dangerous situations [1]

4.1 §1 of the German Traffic Code as Helpful Example

The German Traffic Code ("Straßenverkehrsordnung, StVO“) states in its first paragraph the general behaviour rules in the following way:

„(1) Die Teilnahme am Straßenverkehr erfordert ständige *Vorsicht* und *gegenseitige Rücksicht*. (2) Wer am Verkehr teilnimmt, hat sich so zu verhalten, dass kein anderer *geschädigt*, *gefährdet* oder, mehr als nach den Umständen unvermeidbar, *behindert* oder *belästigt* wird.“
Translated into English this reads:
"(1) The participation in the traffic requires constant *caution* and *mutual consideration*. (2) Anyone who participates in traffic must behave in such a way that no one else is *harmed*, *endangered* or, more than under the circumstances unavoidably, *hindered* or *bothered*."

These rules are the general basis (usually relevant in absence of specific traffic rules) for law suits and court judgements with respect to traffic accidents in Germany [6]. With the help of some interpretation, specification and quantification these sentences can be transformed into more specific driving rules. They also destinguish between different behaviour categories, which are important for priority discussions.

4.2 Specification of the Key Words in the Traffic Code

Here are some specifications, how the key words used in the traffic code should be interpreted, or which aspects should be considered in this context:

Caution: Looking far enough ahead,
watch out for any signs of unusual
or unexpected behaviour of other traffic participants,
know your own skills and limitations,
anticipate the consequences of your actions,
know all the prevailing rules (driveable space, adequate speed, right of way, …),
which implies to first know exactly where you are,

Mutual consideration: Consider and anticipate behaviour of others,
understand typical signalisation of others,
drive predictably for other traffic participants,
consider how your own behaviour is interpreted by others,

Harming: Always avoid own accidents !
(accident: undue touching of other participants or objects, leaving of driveable space)
Do not cause accidents of others !

Endangering: Leave enough space & time for safety reactions,
do not cause situations with too short "time to collision" (TTC),
do not drive with too short "temporal headway" (THW),

Hindering: Leave enough space & time for comfortable reactions,
do not force others to accelerate / steer uncomfortably,
do not hinder normal traffic flow,
which implies to obey all prevailing traffic rules,
signal your braking and lane change intent clearly,
be cooperative (actively open gaps for safe distances to others)

Bothering: Do not stop in the way of others,
do not distract others,
avoid situations which may lead to increased level of danger.

These specifications are derived from normal traffic practice; of course, this is open to further specifications and additions. Most of these words refer to behaviour which aims at reducing the exposure to critical sitations, as explained in the context of figure 1. The lower four words from harming to bothering define an order of criticality in the sense of figure 3. The knowledge of prevailing rules and of the own performance limits (in category *caution*) is considered safety-critical; without correct knowledge a safe and situation-adaptive behaviour is impossible.

The "Responsibility-Sensitive Safety (RSS) model" defines five specific behaviour rules in the following way [5]; their relation to the above specifications is marked with the respective key word:

1. Do not hit someone from behind. (*harming*)
2. Do not cut-in recklessly. (*endangering*)
3. Right-of-way is given, not taken. (*mutual consideration, endangering, hindering*)
4. Be careful of areas with limited visibility (*caution, mutual consideration*)
5. If you can avoid an accident without causing another one, you must do it. (*harming*)

4.3. Quantification of the Key Words in the Traffic Code

The next step is to quantify the specifications of the key words. All values given here are considered as parameters for starting or promoting a concrete discussion. They need to be defined in projects like PEGASUS, or in similar international projects or standards resulting from such.

The quantification of the key words "endangered" and "hindered" is mainly based on human reaction times for safety-relevant and comfort-relevant actions in traffic scenarios. Human reactions are generally limited by necessary perception, interpretation, reasoning and action times. The requirement for AVs must ensure equal or better total performance.

Human reaction times in detail depend on many factors, however they cannot be validated in every conceivable situation separately; thus, typical traffic reference situations must be agreed on. Typical human reaction times for such situations can be measured in simulator experiments; on the other hand, a conservative limit for human reaction time can be derived from existing Driver Assistance Systems (which produce warning sounds leaving reaction times which are statistically proven to be save enough). Such rough estimates should be good enough to be used for other situations as well, if no better knowledge is available. The definition of a limit for perceived "dangerous" situations and "hindering" situations can be verified by experiments as well.

Key word Quantification (suggested parameter values, only for start of a discussion !)

Caution: have a *quantitative* knowledge about own abilities and limitations:
know your own position, driving state and vehicle state,
know your preview horizon (from map) and your sensor range,
in summary: have "self-awareness";
know all rules and conditions on your route *redundantly*:
from an actively managed map (for preview)
and from sensors (interpreting traffic sign, lane markings etc. for final validation),
know variable signals (lane lights, traffic lights) *redundantly* by camera & Wifi;
have an *up-to-date* knowledge of track conditions and suitability for AVs:
location of emergencies and difficulties (objects on the road),
location of emergency vehicles and road works,
weather forecast;
define and apply a reasonable measure for a "level of danger".

Mutual consideration: signal your lane change intent clearly (3s before),
signal unusual reactions clearly (heavy braking by flashing brakes),
be able to anticipate behaviour of others from their motion pattern,
be able to understand signalling (lane change, warning lamps, emergency vehicles);

Harming: No accidents *at all* driving alone,
No accidents *caused* (with participants) if TTC > 1.5s;

Endangering: keep "time to collision" TTC > 1.5s,
keep "temporal headway" THW > 0.7s;

Hindering: obey speed limits (stay below +5%; assume others to stay below +20%),
stay in lane (with mirrors stay away 20cm from lane markings),
do not force others to accelerate / steer uncomfortably
$|a_{long}| < 2.5 m/s^2$, $|a_{lat}| < ??? m/s^2$,
be cooperative (open gaps for others so they can cut in with THW > 0.7s),

Bothering: do not stop in the way of others (do not reduce the driving lane more than 10%),
do not drive unnecessarily slowly (less than 10% below traffic flow without reason).

Some quantifications can be derived from normal driving discipline in traffic measurements, as seen in figure 4: from the distribution of temporal headway in cut-in situations in normal highway driving a THW-limit of 0,7s seems reasonable.

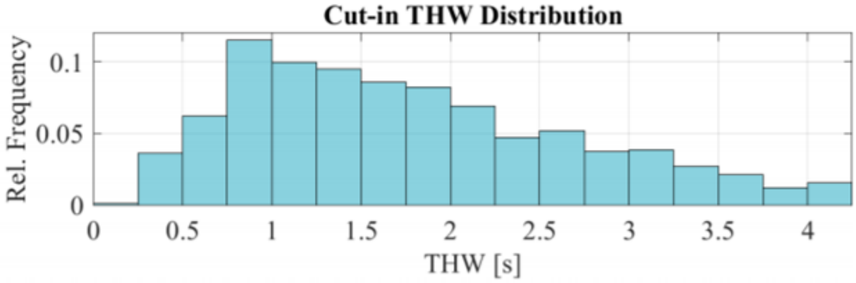


Figure 4: Example for established driving discipline, here temporal headway in cut-in-situations on highways [7]

5. Corollary for *Single* Vehicle Traffic Behaviour

In order to sort the implications of the behaviour rules more clearly, let us concentrate in a first step on guidelines and performance tests for a single AV, without any other interfering traffic participant. This again can be divided into driving under normal conditions and under extraordinary conditions:

- Single vehicle, normal traffic situations; passing criteria "do not endanger":
  make sure you know all relevant rules for your location, obey the rules, stay in your lane, drive with comfortable dynamics, monitor your own system, stay away from own limitations, drive only in locations where your limitations comply with the requirements of the location.
- Single vehicle, extraordinary traffic situations; passing criteria "do not harm yourself, do not endanger others":
  react adequately on objects in your track, react adequately on surprises (sudden new information, invalid map information) with respect to trajectory planning, recognize emergency vehicles and work areas, react on changing weather and road conditions, react adequately on sudden technical failures, … and in all cases: do not react in a way which would endanger others or other objects.

5.1 Single Vehicle Performance Tests

Examples for performance tests which should be deducted to verify compliance with those guidelines are given below:

- Precise localisation of the vehicle, even under difficult conditions (no or bad GPS satellite connection)
  ➔ know your position and deduct the prevailing rules and conditions for this location correctly; show adequate behaviour when localisation degrades.
- Safe and comfortable driving, even under worst case street curvatures for a given road class
  ➔ Do not drive faster or closer than you have under control, w.r.t. speed – distance – lateral distance, without heavy accelerations longitudinally and laterally.
- Awareness of information horizon (sensors, map), ability to react on reduced horizon (sensor deficiencies due to weather or lighting conditions, loss of or wrong map information, loss of communication, denied road clearance)
  ➔ Know your own look ahead and safe stopping performance limitations (safety and comfort limits of acceleration, deceleration, curve speed, sensing horizon) depending on weather and lighting conditions, and show adequate reaction based on this information.
  ➔ Sensing horizon at high speed requires knowledge from a map, with sufficiently recent information – show adequate behaviour when the quality of map information degrades (outdated or missing map data, loss of road clearance)
- Ability to react on objects on the road (lost cargo: a definition of relevant objects to be detected is needed) or unexpected lane deviations (road works, potholes, …)
  ➔ show your detection ability and adequate behaviour at highest speed and large sensing horizon
  ➔ show that behaviour changes due to awareness of reduced sensing horizon is effective
  ➔ If you do not know for sure (for example, redundant information is inconsistent) – assume the worst, show adequate behaviour depending on road type
  ➔ prove that you have a (sufficiently) safe method to handle worst case scenarios!

Component and system robustness are not considered here in detail. Such design dependant reliability tests and verification of fall-back solutions should be covered under ISO 26262.

5.2 Autonomous Driving Starts with Autonomous Stopping

Bringing the vehicle into a safe state out of any situation is one of the most challenging performance requirements for AVs. An autonomous vehicle should not drive without making sure that it can safely stop the vehicle, even under worst case conditions. Even if there is a safety driver in the vehicle, safety concepts should require that the vehicle is able to stop without a take-over of the driver (there might be less stringent worst-case conditions for level 3 than for level 4 or level 5 AVs).

Stopping requires bringing the vehicle down to zero speed from a given initial speed within the distance, which is physically needed for deceleration. This distance defines the *necessary decision horizon* which is the look ahead perception horizon minus vehicle reaction distance (distance required for evaluation of the free space incl. safe trajectory planning and initiation of the deceleration). The necessary decision horizon depends mainly on the possible or accepted value of deceleration. For normal driving without urgent need of bringing the vehicle to a stop, a slower deceleration (goal: comfortable, and do not hinder others) should be considered than for emergency braking (do not endanger others). Scheduling of take-overs in AVs of level 3 at highway speeds should normally be based on map information, since sensor information would not provide a sufficiently large *available decision horizon*.

Thus, the map-based available decision horizon must be large enough to allow for

- stopping with comfortable decelerations (not *hindering* anybody), or
- stopping after a "non-responded take-over request" with a manageable deceleration cascade for followers (not *endangering* anybody)

The sensor based available decision horizon must be large enough to allow for

- autonomous stopping with a manageable deceleration cascade for followers (not *endangering* anybody)
- assuming tbd. minimum-size critical objects to be detected by the sensor set

The maximum acceptable speed can be derived from these conditions, in dependence on the availability of reliable map information. Test scenarios, in order to verify the ability to bring the vehicle to a full stop (with the adequate braking procedure), or in order to verify the correct reaction on the loss of reliable map information, can be derived from these considerations. In level 3, stopping in the lane should be acceptable, since eventually a driver can take over; level 4 and 5 might require finding a suitable stopping point without hindering others (if possible).

5.3 Blind stopping procedure

Sudden sensor degradation, or even a complete sensor set failure, might be caused by outer conditions (as well as by technical faults). No failure rate characteristics are (to my knowledge) available to date which allow to claim, that the sensing system will be available with very low failure rates for continuous AVs operation under any outer condition. Even robust radar systems, as well as any optical system, can have severe performance losses in very short time during extreme rain or heavy snow fall.

As long as no other evidence can be provided by an AV manufacturer, it is suggested that the vehicle performance in this worst-case scenario should be guaranteed and tested with the "Blind Stopping Procedure" as follows:

In addition to the normal driving trajectory, the vehicle should continuously generate a planned trajectory for stopping using a manageable deceleration cascade for followers (not endangering anybody), based on available map and sensor information. In case of a complete sensor failure, the vehicle should be able to switch to this trajectory based on last available valid sensing data plus inertial and wheel data (by ESP/ESC control); the vehicle should be able to come to a full stop without leaving the driveable surface of the road ahead. If sensing system and ESP/ESC-system are well designed (i.e. have independent power supplies), this concept would also help in designing for high technical system reliability. Such an emergency stop should be clearly indicated by warning lamps and V2X communication.

5.4 AV Highway-Pilot: Road Clearance and Map Learning

A highway (“Freeway” or “Autobahn”, with separated lanes for different driving directions and restricted access for bicycles and pedestrians) is a designated, specifically protected and constantly managed traffic area. Here vehicles can drive with less safety margins than on other roads because of these conditions: wider lanes, normally no risk of oncoming traffic, normally no unaware pedestrians, etc..

For the AV concept of “Highway Pilot” as considered in the PEGASUS project, a clearance for autonomous driving has to be obtained from a central server for using autonomous mode. The clearance is related to a dated map of the road section. The information in this map can be assumed valid for planning purposes (to ensure comfortable reactions), however deviations from the map must be considered and map information has to be confirmed by vehicle sensors (to ensure safety reactions). Any safety relevant deviation from the map information needs to be reported back to the central server, in order to provide a constantly learning map. The map information is provided in several layers [8], and it includes:

- road layout: road network, lanes and their interconnections, driveable surfaces, …
- traffic rules: traffic signs, lane markings, locations of variable signals, …
- road condition: moisture, snow and ice coverage, friction coefficient, …
- traffic condition: congestion, accidents, debris, emergency vehicles, road works, …

Especially, the last two categories can change in a very short time, and they might be essential for safe behaviour.

One essential parameter for the safety assessment is the update speed of the map and the information delivery rate to the AVs. It is impractical to require that a complete collision avoidance with fast changing objects based on this map information can be realized. However, the performance of this crowd-based learning needs to be specified in order to define an adequate level of contribution to the safety concept.

Examples for conditions for a release of such a system:

- The clearance for AVs must be reissued/reconfirmed at least every xx minutes.
- Every AV must participate in a feedback loop as condition for AV clearance.
- If several independent background servers are available, exchange of safety critical information must be exchanged to each other within an update rate of xx minutes.

For a fast building-up of a reliable learning map network, it should be considered that even level 2 vehicles could significantly contribute in the feedback loop, at least for some of the information categories.

6. Corollary for Behaviour in Interactive Situations

Based on the fact, that *everybody* should comply with the rules from the traffic code, certain behaviour of traffic *partners* can be expected in *any* interactive traffic situation. In the interpretation of the traffic code in [6] this is called “trust” in the safe behaviour of others.

- Since an AV should not be harmed or endangered by any other traffic participants, very low TTC or THW are not to be expected for its own planning purposes (some extraordinary scenarios might be specified with extreme values);

- Managed highways could have a different limit for certain to be expected behaviour than other roads (needs to be specified with the street class).

On the other side, a safe and robust reaction on (rare) expectable situations (collision free, if only being hindered or bothered by other participants) needs to be ensured. Reference situations for such behaviour should be defined, based on the general behaviour rules. In general, for city driving with all kind of different traffic participants, it is a large task to iterate through all possible interaction types. However, for the Highway Pilot this leads to a limited number of possible interactions; this should be the focus in section 6.3.

6.1 Testing of Borderline Interactive Situations

As explained in chapter 3, interactive scenarios will lead to uncontrollable situations when criticality parameters are increased. Shorter available reaction time will always decrease the ability to control a situation. This can be caused by shorter initial interaction distances in the beginning of a critical scene, or by higher relative speed, or by higher acceleration/deceleration. Since there is a monotony on the dependence of such parameters, borderline situations will guarantee controllability on one side of the borderline case in this parameter space.

The principle for testing for safe behaviour in interactive scenarios should be divided into two cases:

- Interactive scenario, normal traffic situations; pass criteria “do not endanger”:
  Define traffic situations, in which one participant needs to perform a challenging task in traffic, which might involve getting into the intended path of other participants. The *first* participant needs to show in the borderline case that he would not act in a way which would endanger others. For scenarios less critical than the borderline case, the first participant could – without blaming him – perform his manoeuvre.
- Interactive scenario, extraordinary traffic situations; pass criteria “do not harm” (collision avoidance):
  Define traffic situations, in which one participant behaves at the lower borderline limit to endangering a second participant. The second participant needs to show controllability in this situation (i.e. without collision); but for more endangering or even harming behaviour of the first, the second participant may not be liable if he cannot avoid an accident. More than borderline dangerous behaviour of the first participant *may be* compensated by the second participant, but it *cannot be guaranteed*.

6.2 Safety Impact of Behaviour Interaction in Traffic

Using the concept of behaviour within the categories of the traffic code, several combinations of outcome of the interactive situations are possible (see figure 5).

Interaction of participants, acting both with normal, bothering or even hindering behaviour should not lead to any accidents. If one participant acts in an endangering way, accidents may be possible, even if the other participant acts normally, or bothering, or even hindering. That means, there are safety margins which can be considered as “forgiving behaviour” for endangering behaviour of the other participant. Even if both participants act in an endangering way, the situation can end without any accident, although this is less probable. Harming action by any participant leads to an accident, by definition.

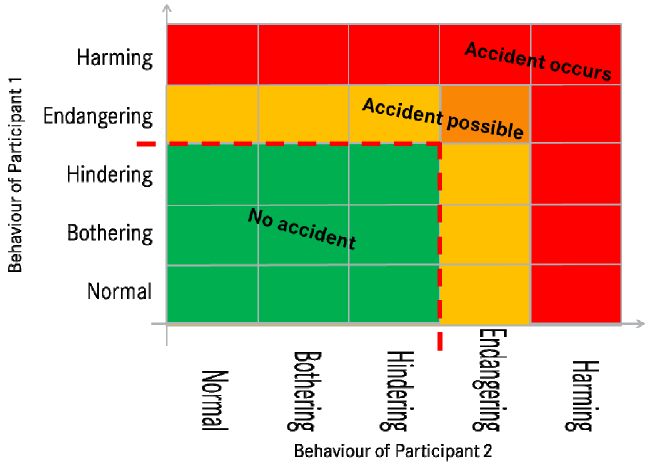


Figure 5: logic of traffic safety: dependance on behaviour interaction

The visualisation of this concept also shows, that it would be not safe to limit the behaviour of all participants to “not harming” for the borderline cases. There would be no margin for small variations in the behaviour. The region of “not endangering” others is important to avoid accidents with a certain robustness.

6.3 Defining Borderline Cases for the Function “Highway Pilot”

In [8] a complete set of interaction scenarios with two vehicles for possible collisions on highways is presented, see figure 6. It shows 16 different ways which can lead to longitudinal or lateral collisions, resulting from different initial constellations. In partial overlap with this, in [9] the three different basic scenarios of following a preceding vehicle, of cut-in and cut-out manoeuvres with sudden view on previously invisible third vehicles/objects on the road are described as challenging scenarios.

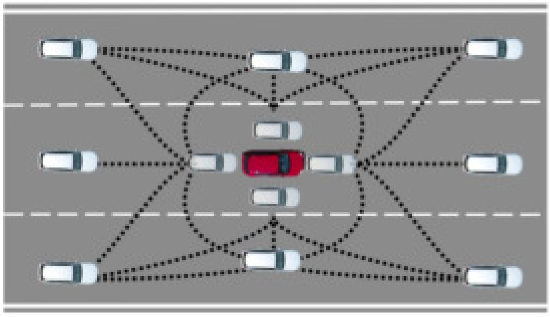


Figure 6: complete set of two-vehicle highway interaction scenarios [8]

For all these different scenarios, borderline cases at the limit of *endangering* behaviour (of the causing traffic participant) can be constructed. These include the following most important examples:

Normal traffic situations:

- Driving towards a known end of a traffic jam or a known object on the road (display adequate braking behaviour);
- Cooperative lane change, as seen from the lane-changing vehicle, especially when one lane in front is ending or blocked, under dense traffic conditions (display adequate braking and/or lane change behaviour).

- Cooperative lane change, as seen from a vehicle in the receiving lane, especially when one lane in front is ending or blocked, under dense traffic conditions (display adequate cooperative behaviour).

The passing criteria for the behaviour against other participants should be, not to endanger the other participants.

Extraordinary traffic situations:

- Cut-in of a vehicle with high differential speed and/or heavy braking in front (challenging the vehicle on the receiving lane).
- Following and full brake of preceding vehicle because of a blocked road ahead; preceding vehicle may be stopped violently by crashing into the end of a traffic jam (define borderline challenge of the following vehicle).
- Cut-out of preceding vehicle gives view on a blocked road (or even an oncoming third vehicle; define borderline challenge of the following vehicle); since the oncoming vehicle is endangering (and heavily against the rules), an outcome without accident cannot be guaranteed.

The passing criteria for the behaviour against other participants should be, not to harm (collide with) others.

7. Learning of Safe Driving

Human drivers develop their full driving skills only after some time of driving experience. We accept that new unexperienced drivers start driving in public traffic without this driving experience. So, the chance of making little mistakes or needing longer reaction times than experienced drivers is higher than for the average of the driver community. Designing the traffic behaviour with reserves for such unexperienced behaviour is mandatory to allow for robust traffic management under such conditions. This way, small mistakes will not lead necessarily to a collision, but a resulting endangering situation might be mitigated by the forgiving (robust) behaviour of others.

The acceptance of unexperienced behaviour only works for a small number of traffic participants. Gaining experience over time by learning mechanisms is essential to make this work in the whole community. We expect from and trust in human drivers that they learn more and more safe behaviour during their first years of driving.

Transferring this concept to the introduction of AVs requires to establish learning mechanisms for AV driving software as well. The industrial concept however is to introduce AVs after an intense development phase. During this phase, feedback loops for learning safe behaviour are built into the integrated development and testing processes of the vehicle developers. Even after the release of AVs, the monitoring of vehicle behaviour and constant improving of the product need to go on.

Let us have a closer look on how the learning of driving skills works [10].

7.1 Human driver

When a driving examiner assesses the driving skills of an applicant for a driver’s license, he considers the driver’s behaviour with respect to his judgement of the entire driving situation including

- own skills of driving
- conditions and abilities of vehicle
- own position and driving state, traffic, weather
- recognition of prevailing rules, their applicability and limitations

- checking of evolving driving situations in compliance with intention and plan
- ability to update a navigation plan in due time
- handling of surprises
- correct assessment of own (and other’s) reaction time

Generally, in driving examinations, no assessment of collision avoidance abilities of the applicant is verified. However, the examinant assesses by his experience, whether the applicant has a sufficiently well-established (implicit) set of behaviour rules based on general values from training in a driving school.

Besides this, in many countries driving with an experienced driver (supervised driving) has been established as an additional measure for fast improvement of the driving skills of young drivers. Even insurance rates are reduced for young drivers using this method.

What happens during the supervised driving phase is as well not focused on improving collision avoidance skills of the driver but improving the mitigation of dangerous situations (reducing the exposure to collision-prone situations). The supervisor helps the unexperienced driver by focusing his attention on early indicators of possibly risky situations and behaviour (of the driver or of others). Especially the knowledge of the own limitations (for example: with respect to fast and correct judging of situations; implications of the weather conditions on the own abilities) are an essential part of the support of the supervisor. Again, there is no clear algorithm on how to recognize such situations; the intuitive skills of the human supervisor are essential for the effectiveness of this learning procedure.

7.2 “Safe Driving Score” for Human Drivers

In order to copy this successful human strategy for accident reduction to autonomous vehicles, the main challenge is to assess complexity and criticality of the situation (“level of danger”) and judgement of the own capabilities (“self-awareness”) based on adequate measures. With these goals, the definition of measures for the safety of driving is essential. This must include also low criticality levels, not only the well-established measures for collision prone situations (using time-to-collision, TTC).

For a human driver, the driving style could be assessed, for example, by giving him feed-back on key parameters like distance, speed and look-ahead time (see figure 7) – as would probably a supervisor do. Even a fuzzy algorithm, which would judge these parameters based on simple assessment methods, would give the driver a feed-back, which he intuitively understands. The driver would learn a safer behaviour; this should work like driving with better fuel economy based on similar fuzzy indications. A “safe driving score” like this one could (at least partially) do a similar job compared to a human supervisor.

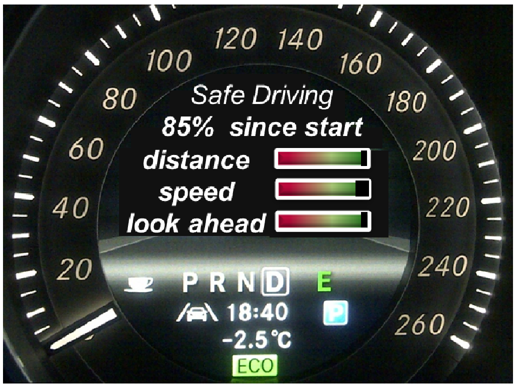


Figure 7: Possible “safe driving score” for human drivers [10]

Keeping a safe distance or a safe speed depends a lot on outer conditions. The driver needs to learn, under which conditions he gets good or worse scores, even with the same driving style, when using it in different situations.

It is hard to implement an assessment of the own driving skills (self-awareness) in a score for the driver. This must be left to the driver himself.

7.3 Self-Awareness of Autonomous Vehicles

For an autonomous vehicle, the driving style is programmed into the vehicle; it makes more sense to assess the robustness of autonomous driving with respect to several categories of outer driving conditions. These categories could be the map quality, the visibility of tracks (lane markings), the traffic conditions, or the weather conditions (see figure 8).
It should provide insight into the question:

How robust *was* the safety of autonomous driving?

For scores above 100%, the vehicle would be able to adopt the driving style accordingly. If one of the categories would fall significantly under 100%, this would be a reason to hand back the driving task to the driver (in level 3 vehicles) or to suspend the driving completely (for level 4 and 5 vehicles).

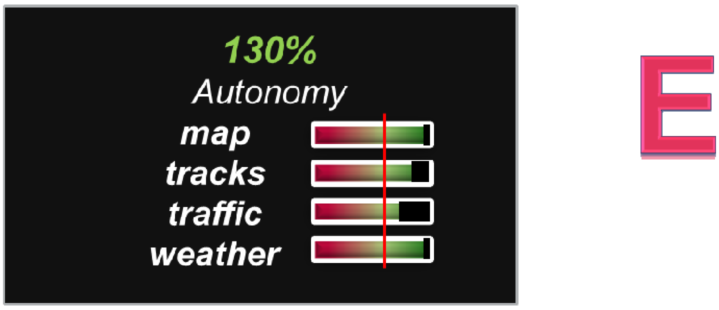


Figure 8: Possible "safe driving score" for autonomous vehicles [10]

In detail: looking back for the last driving minutes, the *a-posteriori* assessment of driving segments could cover the following topics (and explain, why the vehicle changed to a more defensive driving style with respect to the ➔ indicated aspects):

Map

- was the map w.r.t. drivable surface, traffic signs, traffic jams in accordance to the experienced reality and up-to-date, were there any unexpected objects or surprises, any wrong map information?
  ➔ adequate speed and distance to others, look ahead distance

Tracks

- Were the tracks well visible, were they marked as expected?
  ➔ adequate speed

Traffic

- Did the traffic behave in a predictable way, were there any dangerous objects or driving styles, dangerous cut-ins or partially blocked lanes?
  ➔ adequate time gap and longitudinal & lateral spacing

Weather

- Were the sensors able to cope with the weather, was the weather forecast correct?
  ➔ adequate speed, distance, time gap

These parameters provide fuzzy, but evaluable measures for a "safe driving score", either continuously updated, or in summary after a complete ride. They can support the developer in monitoring the performance of the AV driving software, and to a certificatory it provides a measure for judging, whether the AV has a correct sense of the difficulty of the driving task it just performed. Comparing the human assessment of the just monitored driving task with the scores generated by the machine would provide a similar trust in the performance of the driving software as inquiring a human driver's-license-candidate at the end of the test drive on his impressions on his own driving performance. In this way, this type of "safe driving score" is a measure for "self-awareness" of an autonomous vehicle. It provides evidence for its perception of the safety of the driving situation and the way it is responding to it. The score can also be used for Artificial Intelligence learning and improvement of safe driving.

7.4 Implementation and Evaluation

How can such a score of self-awareness be implemented?

The path planner in any AV controller should have such measures already available, because the choice of the best trajectory and the suitable speed needs basically the same rating. However, such performance measures would normally not be available for vehicle validation. At least for certification purposes, but perhaps also as a monitoring feature to increase user trust it should make sense to make these measures available.

The path planner normally would not look back, in case of a surprise or other information only available at the end of a driving scene. So, the scores of the path planner probably would need some extensions to cover the full possible scope.

How can it be evaluated?
Short and long-time evaluations of the scores can make sense:

- Checking of parameters after every critical or even complex situation, whether the vehicle has realized its criticality (checking the self-awareness of the vehicle).
- Monitoring the evaluation of parameters over time to check the learning of the vehicle (verifying the learning ability of the vehicle).

A verification of the ability of the vehicle to perform such self-awareness should be considered as an important part of an AV driver's license (or certification procedure). The autonomous vehicle should this way make evident that it is able to judge its own capabilities and limitations, and that it is able to stay safely away from these limitations under any driving condition.

8. Summary

This paper describes a concept for defining "Good Enough" behaviour of AVs in challenging traffic situations. It starts by deriving a consistent set of principal behaviour rules from the behaviour guidelines in §1 of German Traffic Code. The exact parameters of this approach still need to be quantified by experts, to be derived for example from human behaviour in traffic, in situations which show safety by "*proven in practice*".
The behaviour in normal situations should be oriented on the parameters for "not *endangering*" others, instead of on "*no collision*" (as is standard for level 2 systems). This gives an extra safety margin necessary for the distribution of behaviour to be found in real traffic, like deviations from "acting to the rules". On the other hand, it releases responsibility for collisions in situations where endangering behaviour of *other* traffic participants caused the accident.
Borderline test situations can be derived from the quantified behaviour rules as reference situations for demonstrating safety for certification. Worst case test conditions are proposed for extreme cases which are needed to be defined by the community, not by the single manufacturer.

These rules should be good enough for introduction of AVs, however the vehicles should be able to show that they implement a sense for "level of danger" and "self-awareness", for assessing the performance and effects of AVs as traffic participants. Based on these measures, learning and improvement over time can be implemented.

References

1. H.P. Schöner: **Challenges and Approaches for Testing of Highly Automated Vehicles**; 3rd CESA Automotive Electronics Congress, Paris 2014
2. U. Steininger, H.P. Schöner, M. Schiementz: **Requirements on tools for assessment and validation of assisted and automated driving systems**; 7. Tagung Fahrerassistenz, München, Nov. 2015
3. https://www.pegasusprojekt.de/en/about-PEGASUS
4. https://www.mobileye.com/responsibility-sensitive-safety/rss_on_nhtsa.pdf
5. http://arxiv.org/abs/1708.06374
6. König, P.; Dauer, P.: **Straßenverkehrsrecht**. Beck'sche Kurz-Kommentare, Band 5, pp. 422…424; Verlag C.H.Beck, München, 2013
7. Krajewski, Robert, e.a.: **The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems**. in 21st IEEE International Conference on Intelligent Transportation Systems, IEEE, Ed., 2018.
8. Bock, Julian, e.a.: **Data Basis for Scenario-Based Validation of HAD on Highways**. 27th Aachen Colloquium Automobile and Engine Technology 2018
9. H.P. Schöner: **Simulation for Development and Testing of Autonomous Vehicles**; 18th Stuttgart International Symposium, Stuttgart, 2018-03-14
10. Schöner, Hans-Peter: **Self-Awareness of Autonomous Vehicles**. Symposium on Testing and Certification of Autonomous Vehicles, Nanyang Technological University, Singapore, 2018-09-14
11. https://de.wikipedia.org/wiki/Vertrauensgrundsatz