An Introduction to the Safety Force Field

David Nistér, Hon-Leung Lee, Julia Ng, Yizhou Wang

MAR19

An Introduction to the Safety Force Field

David Nistér, Hon-Leung Lee, Julia Ng, Yizhou Wang

NVIDIA

This document is an introduction to the Safety Force Field. The precise math is detailed in [1].

We all want a vehicle that monitors the surroundings and shields us from collisions. The Safety Force Field is the practical realization of that. It is not possible to guarantee safety regardless of what other actors do, but it is possible to guarantee that we do not contribute to an unsafe situation. If all actors in traffic had such a guarantee, no unsafe situations or collisions would occur. The Safety Force Field achieves this guarantee while allowing normal everyday driving, even the kind of driving that is necessary in practice to, for example, make a lane change in congested traffic.

The Safety Force Field takes the vehicle's understanding of the surroundings and determines a set of acceptable actions. It relies on a constructive computational mechanism that determines which actions contribute to maintaining obstacle avoidance safety. The acceptable actions never create or escalate an unsafe situation. This allows us to monitor and protect against unacceptable actions. The computation can be combined with any driving software as a layer in the motion planning that monitors and prevents unacceptable actions.

The Safety Force Field follows from one core principle as opposed to a large set of rules and exceptions. The guarantees it provides have been mathematically proven. It seamlessly handles highway driving, cluttered urban situations, and unstructured zone driving. It cleanly separates obstacle avoidance from a long tail of complicated rules of the road. It considers longitudinal and lateral constraints together, and the constraints can also be visualized in a very direct way.

The Safety Force Field in One Dimension

The Safety Force Field is built on a simple core concept:

Actors in traffic should apply a safety procedure or equivalent action before it is too late.

Ideally, we would like all actors to do that but in particular, we put this constraint on our autonomous vehicle. To see why this makes sense, let us start with the simplest possible example: our autonomous vehicle is driving forward in one dimension towards a stationary obstacle in front of it (such as a wall). We will use space-time plots like the figure below to illustrate.

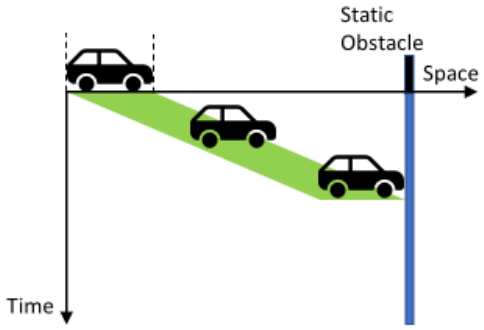


Fig 1. A space-time plot of a vehicle driving at constant velocity into a stationary obstacle.

We use the horizontal axis as space and the vertical axis as time. We use a bounding polygon for the vehicle. In this simple example it is just an interval in space that covers from the front to the back of the car (between the dashed lines). Note that we may include some air in that bounding polygon to provide additional margin. We think of that bounding polygon as the volume around the car that should not touch anything else. If the vehicle drives forward at a constant speed, the space-time plot traces out a parallelogram until we reach the stationary obstacle (as in the figure). Note that any stationary obstacle becomes a vertical shape in the space-time plot. Now, of course, we need the vehicle to apply the brakes before it is too late. In this case, the safety procedure is to ramp up braking quickly and firmly, and keep braking until stopped. If this is done just in time, the vehicle will stop exactly when it is about to touch the wall. In the space-time plot, the area traced out by the vehicle will become perfectly vertical at the stopping point right at the wall, as seen in the figure below.

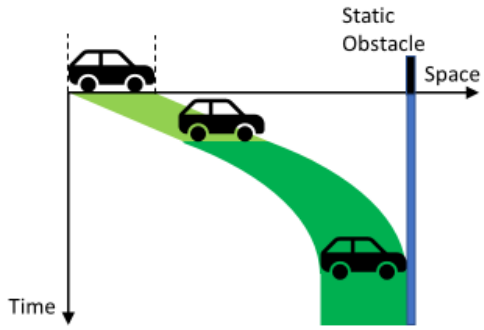


Fig 2. The vehicle applying its safety procedure just in time.

The safety procedure (in dark green) can be calculated given properties of the world and the vehicle state at the current time. Note that guaranteeing obstacle avoidance in a static world can in principle be achieved that easily: while driving around, monitor whether the safety procedure thought of as an area (or volume) in space-time is about to intersect with the vertical shapes of static obstacles, and apply the safety procedure just before that happens, at the latest. If only relying on that constraint, it will not result in a comfortable ride, but we are looking to implement a base constraint that guarantees safety without further limiting the system that operates within those constraints. Intuitively, requiring application of the safety procedure just in time is the minimum constraint that achieves the guarantee of safety. In that sense, this choice seems canonical. In one dimension, the world is simple. We can only drive forwards and backwards. The same safety procedure concept applies for backward driving as illustrated in the figure below. The common concept is to slow down to a stop as soon as possible.

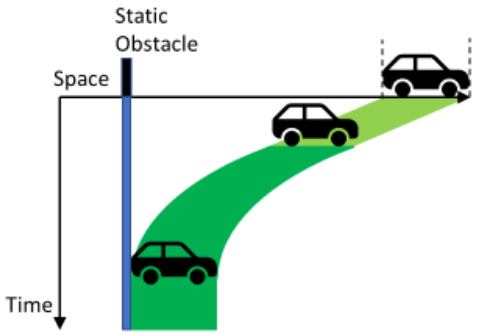


Fig 3. The safety procedure applies also when driving backwards. The common theme is to come to a stop as soon as possible.

It should be immediately clear that if we are in a one-dimensional world between two obstacles, and we apply the safety procedure (stay within the dark green area) whenever it is about to touch one of the obstacles, we never hit any of them.

To provide a margin for practical implementation, the safety procedure is defined as a requirement to decelerate at least as much as a certain schedule. That is, we allow decelerating more than that schedule requires. However, we do put a limit on how much braking is possible (at any time). This limit is expected to be the physically feasible deceleration rate. The reason for this is that we have to understand maximum braking amounts when following another vehicle. Thus, we have two deceleration schedules: the safety procedure schedule and the maximum braking schedule. We refer to the volume in space-time spanned between those two extremes as the claimed set, illustrated in the figure below. The space-time sets are called claimed sets because the actors effectively 'claim' their part of space-time. The Safety Force Field can be thought of as the contention that occurs between different actors if they attempt to claim the same parts of space-time.

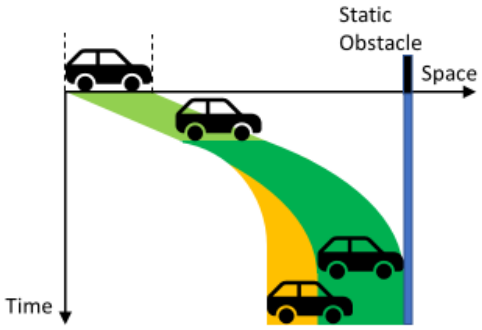


Fig 4. The safety procedure is a requirement to decelerate at least as much as a certain amount (dark green). There is also a maximum braking schedule (orange). We call the set spanned between the two the claimed set (union of the dark green and orange areas).

Let us now consider other moving actors. It is not possible to guarantee absence of collisions regardless of what other actors do, as illustrated in the figure below.

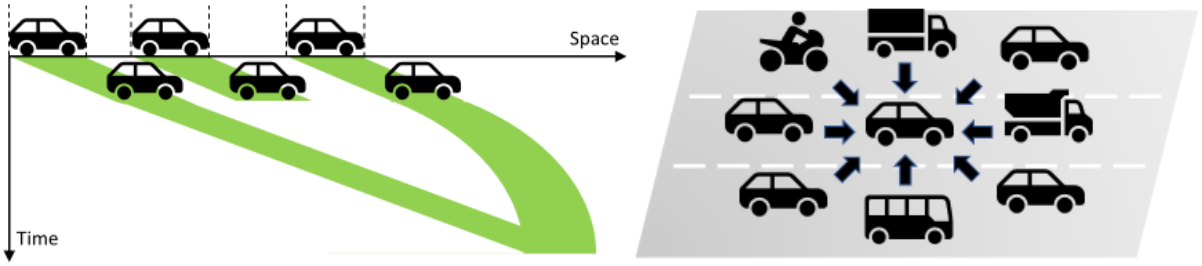


Fig 5. Absolute safety is not possible with adverse other actors. Left: The vehicle in the middle has nowhere to go if its lead vehicle decides to brake and the following vehicle continues to accelerate. Right: The situation is the same in two dimensions since other vehicles may be blocking the sides. We could ask that we be stopped before a collision occurs but would then be unable to drive at speed on a congested highway due to the possibility that the other actors close in on us from all sides.

A more viable approach is to look for the minimal symmetric responsibility that guarantees absence of collisions. While it is reasonable to ask for more than that to accommodate for irresponsible actors, it seems inappropriate to require less since without it, collisions would not be prevented.

The core concept throughout the entire theory is that all actors will be required to help avoid or minimize intersections of claimed sets. For the one-dimensional case of two cars (actors) driving towards each other, there is a critical moment when the safety procedures (claimed sets) of the two actors are just about to start overlapping. This is the moment when the sum of the stopping distances equals the distance between them, as illustrated in the figure below.

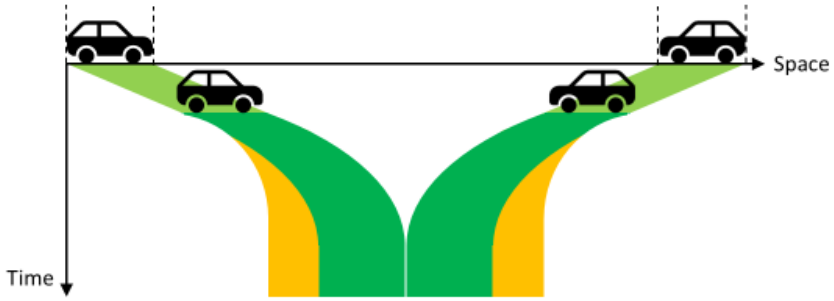


Fig 6. In the case of two oncoming cars, the minimal constraint is that both actors have to apply their safety procedures just before they are about to overlap.

At that moment, both actors will have to act, or a collision will occur. If they do, however, then they are both in the clear (although just barely). Note also that before that moment, there is strictly speaking no constraint on either of them. Thus, we see that avoiding intersection of the claimed sets is the necessary and sufficient constraint to avoid collisions. If we instead consider one car following another, we know that the lead car is required to stay ahead of the maximum deceleration profile, but no more, since common sense implies that if the lead car is hit from behind, it is not at fault, nor can we reasonably require the lead car to not brake. It has to maintain its right to brake, if necessary. The following car thus has the sole responsibility, and it can guarantee no collision by applying its safety procedure just before the front of its claimed set touches the back of the claimed set of the lead car, as illustrated in the figure below.

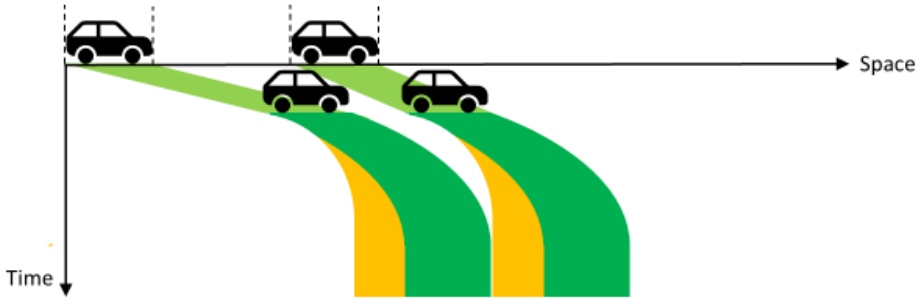


Fig 7. The case of one car following another also becomes critical exactly when the claimed sets intersect. At that moment, the following car has to apply its safety procedure, while the front car has no constraint other than staying ahead of maximum deceleration.

Note again that this is required in the sense that if the following car does not do it, the lead car could brake the maximum allowed and an accident could occur. Note also that it is sufficient in the sense that the lead car is not expected to brake more.

We now identify a common way to state the requirements for all the cases we have considered so far, with the motive that it will generalize. The requirement is that:

All actors are required to apply their safety procedure or take an action that is at least as good before and whenever their claimed sets intersect.

By applying their safety procedure, we mean that they decelerate at least by the safety braking profile. It is also implicit that the actors will not brake harder than the maximum braking profile. Note that this captures all the cases so far, with the understanding that the claimed set of a static obstacle is simply the vertical extension in the space-time plot of the shape of the static obstacle. The caveat 'or take an action that is at least as good' is important. In the precise theory, it amounts to a perturbation analysis that determines whether braking really helps minimize the overlap between the claimed sets. The lead car does not help by decelerating, so it is not required to. This is established by comparing what would happen if following the safety deceleration profile (which establishes a minimum bar) to what would happen with any other control. If another control input does at least as well in terms of minimizing the claimed set overlap, it is allowed.

The main result is absence of collisions between actors that satisfy the requirements. A key property is that the safety braking profile and the maximum braking profile are defined in such a way that the claimed set is non-expanding over time when the safety procedure is applied. The main result then follows from proof by contradiction. If claimed sets intersect, all actors already had started applying their safety procedures before that happened. When they did, the claimed sets did not grow after that, which leads to a contradiction.

Note that we have not yet mentioned reaction time or uncertainty. One way to think of reaction times is that the actors have to account for their own reaction time. Based on the sensing they are now considering, they have to choose actuation commands so that when actuation actually takes effect, the actuation obeys the constraints, even in the face of the inevitable delay between sensing and acting. This, for example, means that additional distance margin has to be required when following. However, it is in some ways cleaner to not dictate those additional margins, as long as each actor can meet the requirements. We will come back to this at the end of the document.

The Safety Force Field in Higher Dimensions

When adding the steering dimension our space-time plot gets two dimensions for space. We may think of it as flat-land—a planar world for sake of imagination—although the real world can have varying altitudes and obstacle heights that are projected to a plane as bounding areas of obstacles. We still plot time as a vertical spatial dimension for illustration purposes, like in the figure below.

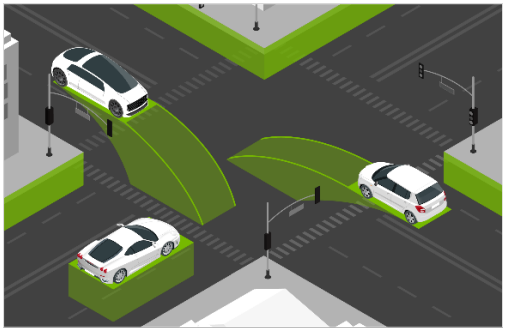
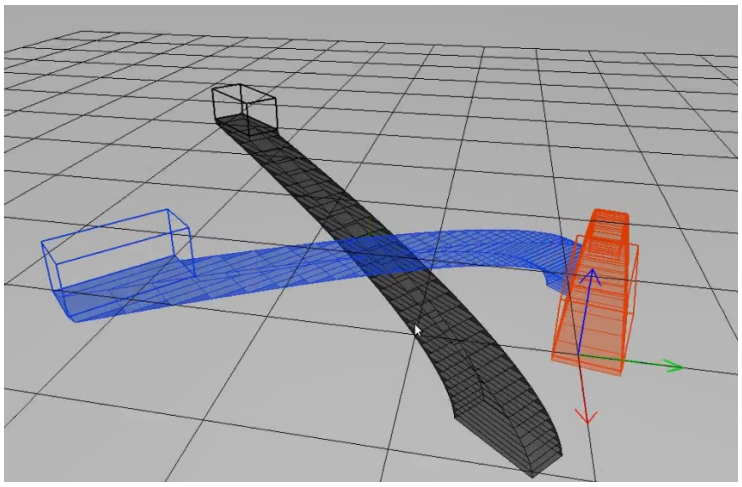


Fig 8. When we add one world dimension as well as the steering control, the claimed sets are tracing out volumes that are approximately braking-parabolas modulated by some steering safety procedure. By rejecting overlap between the volumes at all times, we also prevent actual collisions between the actors.

Here, as before, we get vertical shapes for any static obstacle or actor at rest. Moving actors trace out similar longitudinal trajectories as before, similar to a parabola forward in space-time, except now they sweep out volumes that are modulated by the steering definition of the safety procedure. It is helpful for intuition to think of these volumes as solid geometry with lengths that increase with velocity and the actors as driving around with these volumes attached to them while performing collision analysis for these volumes instead of their actual current bounding shapes. Like in one dimension, the concept is still that before these volumes intersect, the actors will apply their safety procedures, and as they do, the volumes will not change, only play out in time. Again, if they do, the volumes will never intersect, and no actual collisions will occur. In other words, by guaranteeing no collisions of the space-time volumes, we induce a guarantee of no collisions in actual space. The beauty of this is that while it is hard to reject collisions between the actual physical obstacles, it is easy in space-time. The reason is that in actual physical space it requires foresight since actors have inertia, so once a physical overlap takes place it is too late. While in space-time, the volumes can be more or less instantly frozen when danger is about to set in. Now that we have both dimensions, the analysis keeps us safe both laterally and longitudinally, with any geometry between us and contending actors. For example, with crossing geometry as in the figure below.



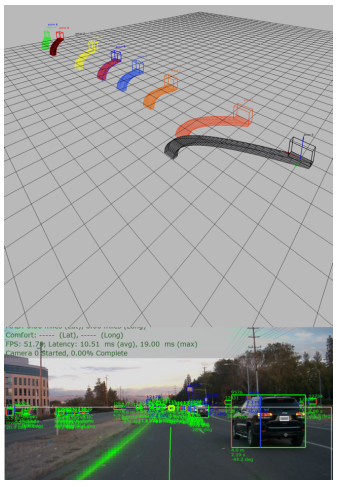


Fig 9. The space-time analysis of the Safety Force Field keeps us safe laterally as well as longitudinally, with any geometry between us and other actors. Left: Synthetic general geometry case. Right: Real case.

or to save us from side-swipes during a lane change as illustrated in the figure below.

Fig 10. The space-time analysis of the Safety Force Field monitoring the safety of a lane change in congested traffic.

Let us now consider the steering aspects of the safety procedure more in detail. The idea is similar laterally as longitudinally, although braking to a stop for the lateral dimension boils down to bringing the lateral rate of change towards zero as expediently as possible. The Safety Force Field formulation allows quite a bit of flexibility in how we define low lateral rate of change. In an unstructured environment or at low speeds, we can define it as driving straight ahead or as to continue on the current steering circle, as illustrated in the figure below.

Fig 11. The lateral part of the safety procedures is about bringing down the lateral rate of change. Unstructured safety procedure choices that do not depend on road structure can be to continue straight or to continue on the current steering circle.

This is important since many environments have unclear path structures and we do not want safety to depend on them in those situations. Paths are also typically perceived with less redundancy than obstacles. For high speed highway driving however, we want to be able to bend our behavior to the road shape to accommodate for curves. Then low rate of lateral change means following the lane shape of the path we are on. If we are already following the road, that

amounts to holding steady in the lane. If we are transitioning laterally like during a lane change, it means lining the vehicle up with the road quickly, as illustrated in the figure below.

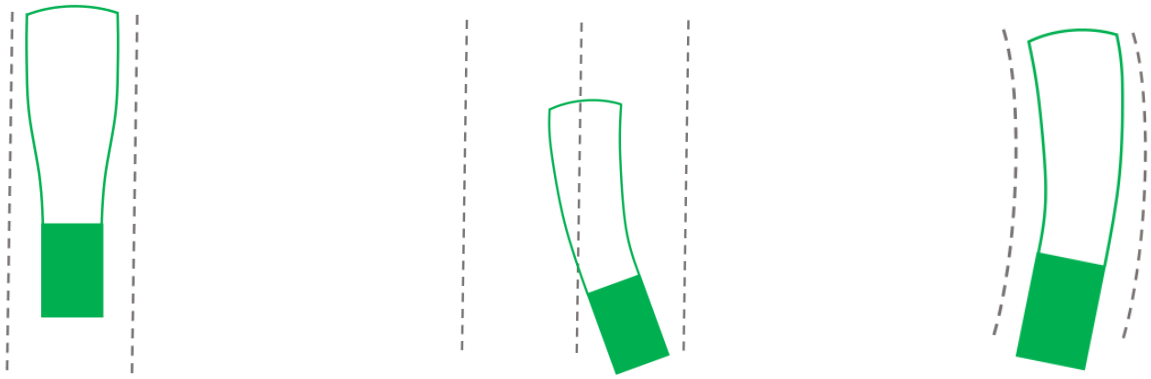


Fig 12. Safety procedures that adapt to the structure of our path, which is useful at high speeds. Left: Holding lane. Middle: During a lane change. Right: Following a curved path.

The Safety Force Field provides easy and seamless transition between unstructured and structured paths since the only change is in the definition of the safety procedure, and all other calculations proceed the same way. It also handles all types of geometries and angles between actors seamlessly without special case rules and caveats that otherwise quickly become unwieldy. It defines what is safe whether following, leading, changing lane, or being involved in arbitrary geometry of crossing paths or unstructured environments, as illustrated in the figure below.

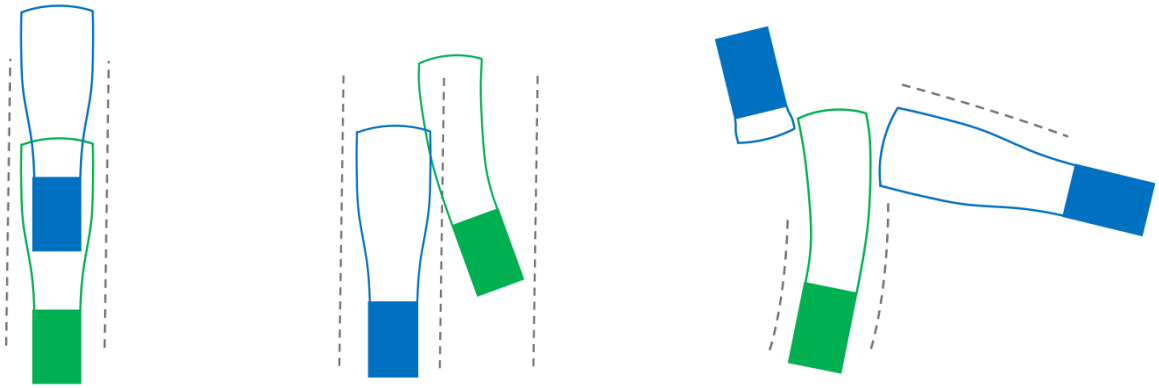


Fig 13. The Safety Force Field seamlessly defines what is safe whether leading or following (left), changing lane (middle) or being involved in complicated scenarios (right).

Other types of actors are also handled seamlessly—they just have different safety procedures. For example, pedestrians and children in particular can change direction erratically, and move in less structured ways. They may also be less vigilant, which we have additional safety checks for, such as the Last Safe Arrival (LSA) analysis, which does not assume pedestrians see us before we are in their path. This can be used for cautious zone driving, in addition to the Safety Force Field. But note that the theory allows what we call safety procedure to actually be 'run in any direction you like as fast as you are able' for a non-vigilant pedestrian. The theory still works and generates a conservative behavior for our vehicle. Safety in this case comes from the reasonable limits on how fast the pedestrian is capable of moving.

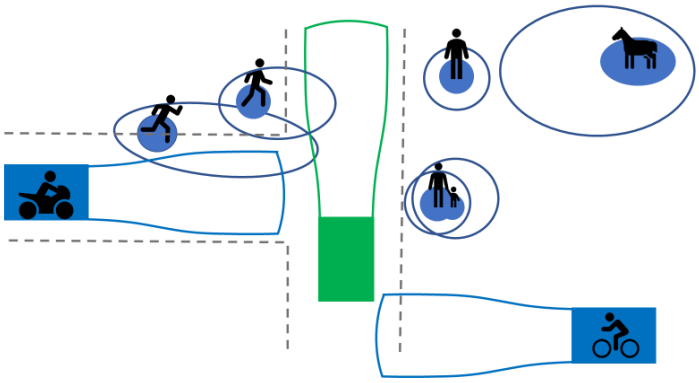


Fig 14. The Safety Force Field handles all types of actors—they just have different safety procedures.

The longitudinal and lateral dimensions are handled jointly. This is important for some cases. An approach that looks at longitudinal and lateral safety margins separately cannot allow the case of pushing diagonally into a lane at low speed. The reason is that at high congestion, we cannot expect to longitudinally clear the vehicle we want to take way from before we are partially in its lateral path, as illustrated in the figure below.

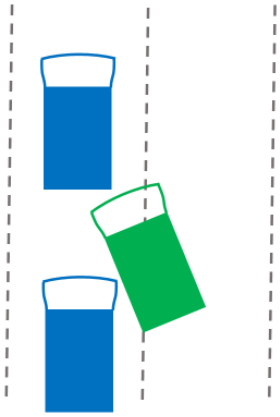


Fig 15. The Safety Force Field naturally allows making way into a congested lane at slow speed as can be required in congested highway situations. This is not possible with a formulation that separates lateral and longitudinal distances and requires at least one of them to be acceptable. Note that in this situation, the ego vehicle (green) is neither laterally nor longitudinally clear from the car behind it to the left.

Visibility is also taken into account. This is done by assuming invisible actors are present at blind corners and behind parked cars where they could reasonably be expected, in states that they could reasonably be in. When we slow down or keep a larger distance before passing by an occluding wall, we can think of it as taking into account the actors that may be around the corner, as illustrated in the figure below.

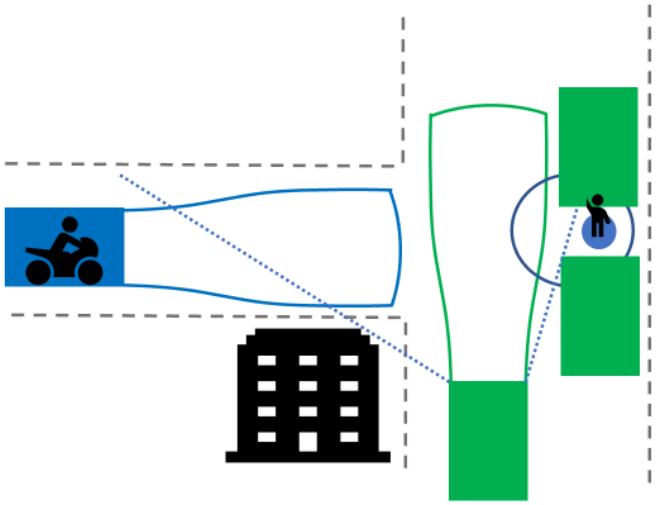


Fig 16. Occlusion is handled by assuming invisible actors are present at blind corners and behind parked vehicles with reasonable velocity limits.

Claiming Parts of Space-Time

We have called the space-time sets 'claimed sets'. The intuition behind this is that actors 'claim' their part of space-time. This seems appropriate because if only one actor has claimed its claimed set, it means that the trajectory played out by its safety procedure is unencumbered, its safety procedure intact. Vice versa, if more than one actor has claimed some portion of space-time, then even if they do their best and conduct their safety procedures, there will still be an issue. One could in principle hope that evasive maneuvers other than the safety procedure would avoid an accident, but this leads to dangerous territory. Such maneuvers are not guaranteed, and worse, could interfere with yet other actors, with knock-on effects that are hard to predict or guarantee. The Safety Force Field can be thought of as the contention that occurs between different actors if they attempt to claim the same parts of space-time, as illustrated in the figure below.

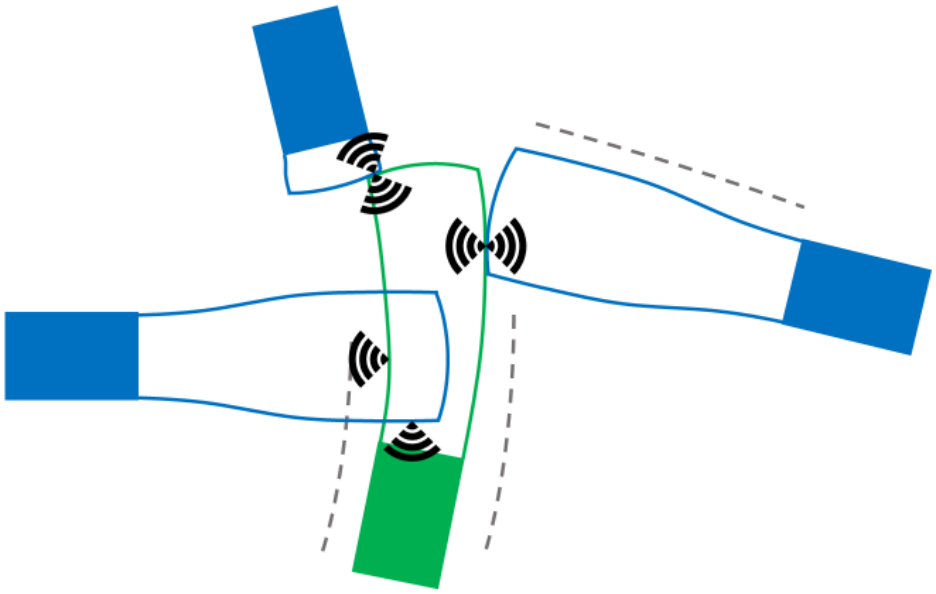


Fig 17. The Safety Force Field can be thought of as the contention that occurs between different actors if they attempt to claim the same parts of space-time.

When they do not claim the same parts, there is no force. When they do, it usually imposes a constraint on both actors, which is that they both have to contribute to reducing the overlap at least as much as their safety procedure would. Consider the example of one vehicle about to perform a close cut-in lane change in front of another vehicle, as illustrated in the figure below.

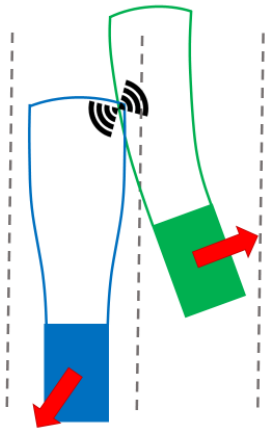


Fig 18. The contention between actors induces a constraint on the control actions the actors are allowed to take. Here is an example of a close cut-in lane change. We have drawn the constraint arrows at the vehicles to emphasize that the constraint is worked back to a constraint on the control of the vehicles. The constraints indicate that the left vehicle should brake and not steer right, although steering left is acceptable. The right vehicle should not steer left.

If there is a close vehicle from both sides however, the freedom is further limited, as illustrated in the figure below.

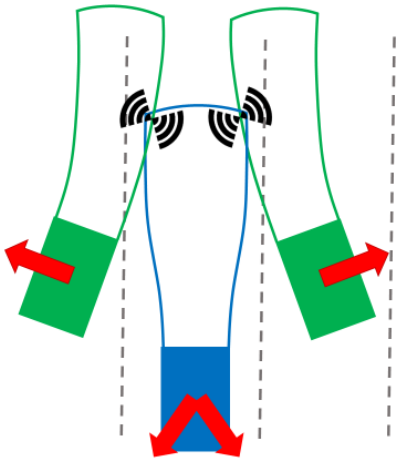


Fig 19. The Safety Force Field does not always require that the safety procedure be executed exactly. The safety procedure is there as a baseline and other control actions that are better are allowed. For example, when another actor pushes us dangerously from the side. With constraints on only one side, the Safety Force Field allows the useful flexibility of swerving to avoid an accident. With constraining vehicles to both sides, however, the blue vehicle now has to apply its safety procedure (brake hard and follow the road).

In this case, the blue vehicle only has the option left of braking hard and following the lateral course its safety procedure dictates (note that if the safety procedure is structured by the road, the road shape breaks the tie for the choice of lateral course). Further examples of contention geometries are shown in the figures below. Crossing geometry where no vehicle has appropriately given way typically ends up constraining both actors to brake hard. With more complex geometry and many constraints, the safety procedure always remains as an allowed action.

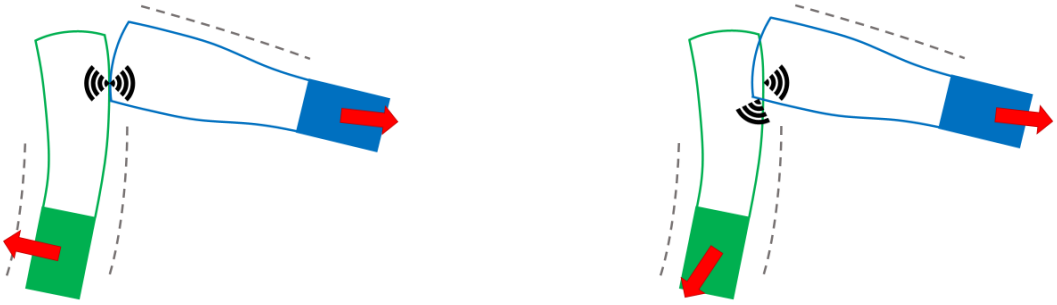


Fig 20. Crossing contention geometry. Left: A crossing contention may for a split second start as a braking requirement for one vehicle and a lateral constraint for the other. Right: However, it will quickly turn into a requirement for both vehicles to brake hard, unless one of them appropriately gives way before that happens.

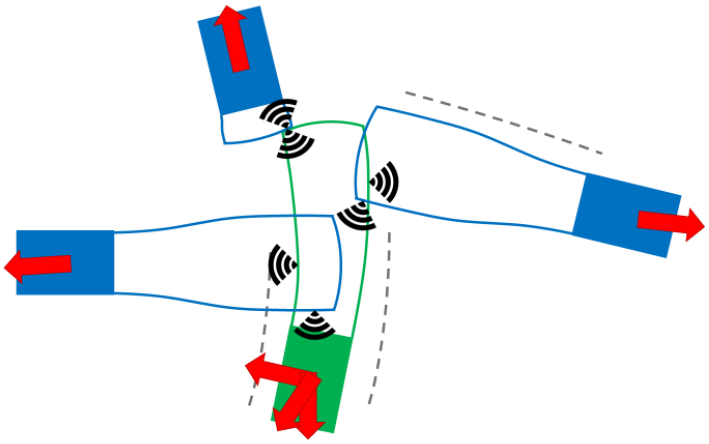


Fig 21. With many constraints, the safety procedure may be the only allowed action left (when no other action can improve upon the safety procedure), but that action is always allowed.

Perturbation Analysis

To provide the seamless flexibility to handle all spatial contention geometries, the Safety Force Field considers the geometry to determine what is helpful and what is not. More precisely, it calculates how good control actions are at rejecting overlap between the claimed sets traced out by the safety procedures. Remember that in one dimension, a following vehicle has to brake, while a lead vehicle can proceed unimpededly. That is a special case that just falls out of the general analysis without making any special provisions or applying special case rules. We start by illustrating in the figure below how this simple example follows from the general principle using perturbation analysis.

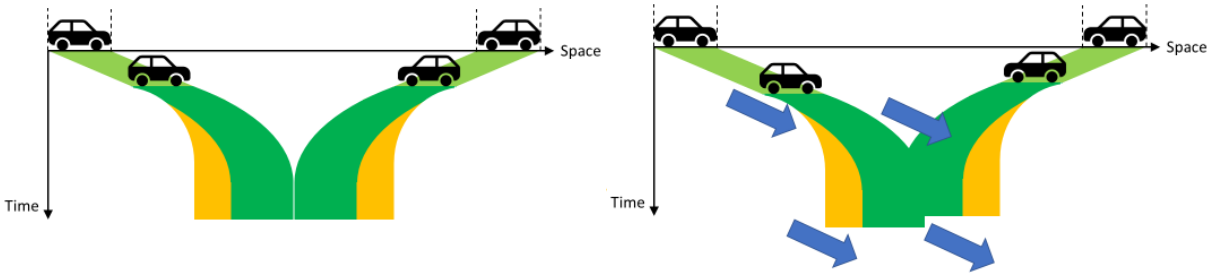


Fig 22. We illustrate how the basic one-dimensional cases follow from the general principle without adding additional rules. Left: The base case of two oncoming vehicles at the critical time. Right: Perturbation analysis of the left vehicle continuing forward at the same speed instead of applying its safety procedure. This results in the claimed set 'pushing into' the other. This is not a better choice than the safety procedure, so is not allowed.

Perturbation analysis takes the current state and considers different control actions, looking at how they contribute to changing the situation. A control action is allowed if and only if it contributes to improving the situation at least as much as the safety procedure. In the oncoming case, braking any less than the safety procedure is worse, so braking is required. The vehicle following case is illustrated below.

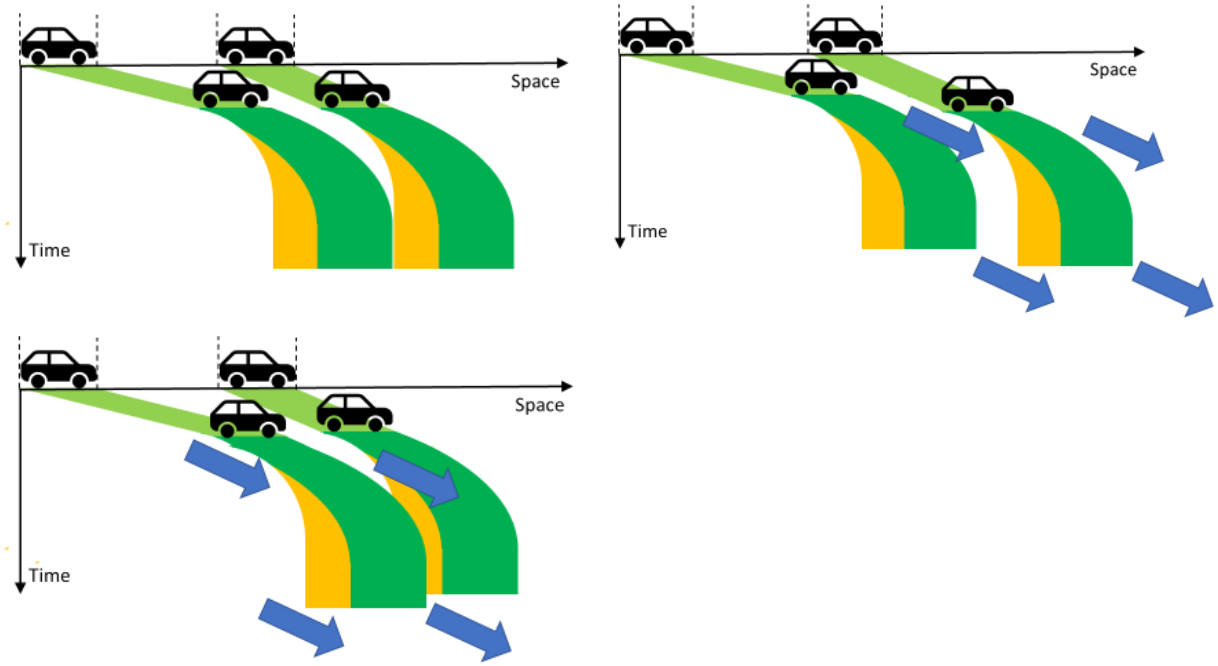


Fig 23. Perturbation analysis for the car following case. Top Left: Base case. Top Right: If the lead vehicle continues with the same speed, it is better than the safety procedure (contributes to opening a margin), so it is allowed (as is any action since the safety procedure is the worst action). Bottom Left: If the following vehicle continues at the same velocity it contributes to pushing the claimed set into the other, and it is worse than the safety procedure, so is not allowed.

We see that the natural obligations for the actors follow from our single basic principle: do at least as well as the safety procedure whenever claimed sets intersect. We can also apply the perturbation analysis to the close lane change case, see figure below.

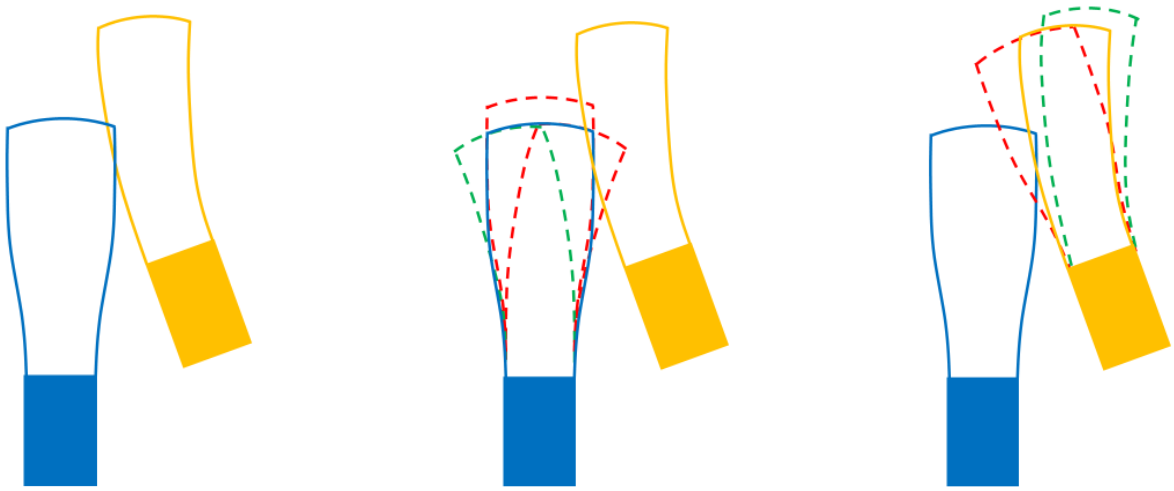


Fig 24. Perturbation analysis for the cut in case. Left: Base case. Middle: For the blue car, swerving left and braking is better than just braking, but swerving right or accelerating is worse. Right: For the yellow car, swerving left is worse, but swerving right and accelerating is better.

In this case, the perturbation analysis yields the following. The car that is about to cut in should swerve back. Accelerating is better so braking is not required. For the car that is experiencing the cut in, braking and swerving away is better. We also show an example of perturbation analysis with crossing geometry in the figure below. Again, results are reasonable. We elaborate on the perturbation analysis in Appendix A.

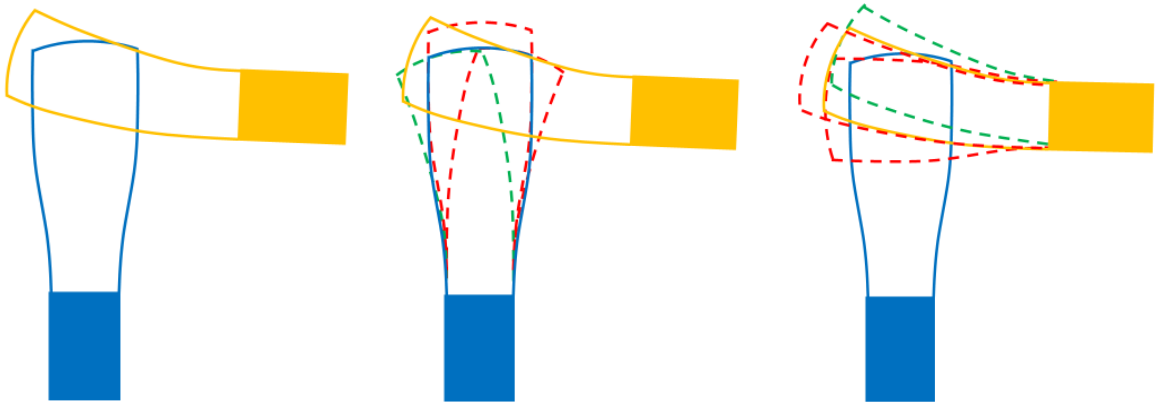


Fig 25. Perturbation analysis for crossing geometry. Anything not involving hard braking is forbidden. Swerving does not change much, but is allowed in some directions (while not required since straight safety procedure is always allowed).

Yielding: Right of Way is Given, Not Taken

Let us now consider yielding and right of way. Yielding and right of way are handled in another layer above the core obstacle avoidance layer that includes the Safety Force Field. We believe that the statement 'Right of way is given, not taken' is correct. Our implementation honors this precisely in a clear way. In our interpretation it means: If the Safety Force Field indicates that there is a constraint due to basic obstacle avoidance, all actors have to act on it. No exceptions. No caveats. In other words, if another actor has failed to give way, we cannot try to take it.

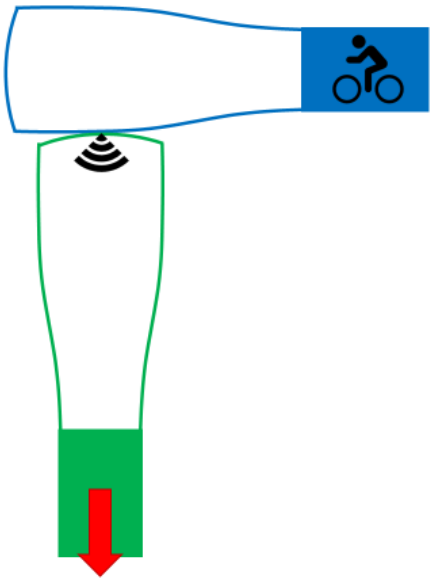


Fig 26. Right of way is given, not taken. Regardless of whether the green car has right of way due to rule or convention, if the blue bicyclist was supposed to yield but failed to do so, the green car becomes constrained. The Safety Force Field has no exceptions or caveats, providing essentially the first practical actual implementation of a 'safety cocoon' that is always on and always watches out for basic obstacle safety.

This arises both from common sense and from the desire to have separation in layers of safety. Common sense: if you are about to crash into someone who has already obviously failed to yield, you would rather brake very hard to minimize damage than insist on your right to priority. Separation: You do not want primal obstacle avoidance safety in a critical situation to be at the risk of being invalidated based on complex decisions regarding things such as priority, traffic rules, the existence of a solid dividing line and its implication in different jurisdictions, a traffic sign with 'no U-turns on Thursdays between 4-6PM', or even the state of a traffic light. A collision risk is a collision risk no matter who is right, and it is desirable to have a separately validated system to handle basic collision avoidance. It is similar to the desire to not have an emergency braking system depend on traffic sign or light perception, or a map. Other interpretations have caveats, but they are dangerous since it forces a dependency in validation and the risk that the caveats are incorrectly determined.

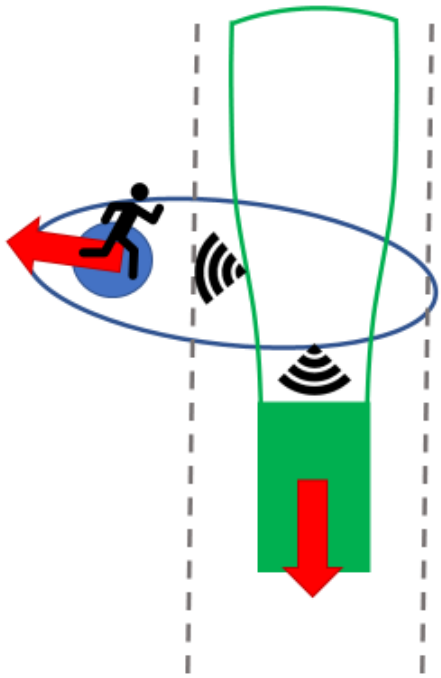


Fig 27. Pedestrian stepping out in front of traffic while we think we have priority (in a sudden way and such that a collision is inevitable). The Safety Force Field will always require braking regardless of who caused this situation or whether we think we should have had priority (such as for example due to a red pedestrian traffic light).

A common case is when an intersection is so congested that the vehicles that went into the intersection while the light was green end up stuck in the intersection, leaving them in the path of crossing traffic that now has a green light. While this is in principle something that they should have anticipated ('don't block the box'), this is not always easy or followed in practice, and in those cases, we can certainly not proceed straight at the vehicles stuck in the intersection. Note that this type of situation could arise for many other reasons, such as a broken down vehicle, a prior collision, road work, or a drunk driver.

Note that the Safety Force Field alone does not force actors to yield appropriately. It is perfectly possible to obey the Safety Force Field and still fail to yield correctly according to the rules and conventions of the road. For example, a vehicle can go into an intersection and stop there, after which it obeys all Safety Force Field constraints (since sitting still is always allowed). But this does not prevent it from annoying crossing traffic or violating traffic rules. This is a necessary result of wanting to separate the core obstacle avoidance from the long unwieldy list of traffic rules. If we required them all to be followed already in the core obstacle avoidance layer, they would all get pulled into it, which as explained above is very undesirable. The vehicle that fails to yield, however, does cause a Safety Force Field constraint that is experienced by cross traffic in a case where it should not due to the rules of the road. This leads us to our definition of yielding. Yielding in a sense is giving way, since other actors cannot take it:

Yielding to another actor is to behave in such a way as to not induce a constraint on that actor, and to make it clear to that actor, by giving sufficient margin, that we will not induce a constraint.

This means, for example, that we can traverse an intersection with crossing traffic or perform an unprotected left turn, provided we are very well clear with large margin from inducing a Safety Force Field constraint. But if we do not have a large margin when going first, we have to stay back and do so with a clear enough margin that it does not make crossing or oncoming traffic worried that we may not yield.

Note that in this definition we have captured what it means to implement yielding. We need to combine this with a determination of who should yield, answering the question of right of way. This is a complex matter that requires the full battery of wait perception (traffic lights, traffic signs, stop lines, yield lines). It includes analyzing who stopped first at a multi-way stop, who is approaching from the right, and understanding of rules that vary by country and region and detecting situations where they apply. This also benefits from map information or training data that provides such rules and allows us to understand the local rules and conventions as keenly as a local.

Note also that yielding seems harder to make as canonically precise as obstacle avoidance. Take, for example, when turning left into a T-crossing and merging into crossing traffic. If we are to yield, we should not put a constraint on the crossing traffic. But after turning, we may face a stopped vehicle and be unable to proceed quickly away from the crossing. Now the crossing traffic that comes after us will eventually experience a constraint from us being there. At some point in this process, our responsibility to yield turns into the common sense that if we are ahead, it is the responsibility of the other traffic to slow down. The Safety Force Field clearly constrains us from performing a downright dangerous cut-in, but yielding was supposed to be more than that, to even avoid inducing the constraint on the other traffic. Yielding by going above and beyond the Safety Force Field constraints is more of a polite best effort to not constrain those to whom we are supposed to yield. This is yet another reason to handle it as a separate layer.

Uncertainty and Errors

Any practical perception system will have uncertainty and errors. The way uncertainty is handled is to provide confidence intervals for all the metrics needed to calculate the Safety Force Field constraints, such as shape, position (including distance), velocity, and acceleration. The calculations are then performed at the edge of the confidence intervals (such as after dilation of shape and distance to the closest point, and velocity to the highest towards us). Rounding and quantization in the computations also have to be taken into account. Errors, on the other hand (by which we mean false detections, failures to detect, cases outside the confidence intervals, or mis-classification of actor type), are handled by redundant perception systems. The very latest form of fusion is to compute the Safety Force Field constraints for each of three redundant perception systems (or more) and allow the control actions allowed by the majority of the perception inputs.

Reaction Time

Taking latency into account is critical for any practical implementation and for satisfying the constraints. We have detailed the Safety Force Field constraints in pure form. The reason is to keep the constraints as close to canonical as possible, in the sense that they are necessary and sufficient. There is some useful flexibility in how to model the state space of actors and their shape, how they are affected by control, and how to define the safety procedure and the safety potential. But once that is done, the constraints follow naturally. For purposes of standardization, it seems purer to require adherence to constraints that can be measured and observed in the actor behavior. Specifically, we have required that actors obey constraints on when they start braking and steering. We also have to define what is reasonable to expect regarding acceleration of any actor. Without such limits, the theoretical actor capable of near infinite acceleration could be anywhere to a practical observer with a finite reaction time. But whether the observing actor has a delay of a hundred milliseconds or one second between sensing and actuation, we want the external behavior of the actor to be constrained in the same way. An actor with high latency will have to send their braking commands sooner relative to the sensor data they observe compared to an actor that can sense and turn around ultra-quickly to actuate braking. We can, and should, build in extra margins, but we argue that the standardized part of those margins should be spatial margin in the bounding shape, margin in the expected acceleration of actors, and control margins in the definition of the safety procedure. In any practical system to satisfy the standard, implementors will then have to add margins for their reaction time by accounting for that obstacles can be closer and have accelerated by the time the actuation takes place. This is similar to how uncertainty is handled. From one point of view that is what reaction latency causes, namely additional uncertainty in the world state by the time actuation will take place.

References

[1] The Safety Force Field, NVIDIA, 2017.

Appendix A: From Perturbation Analysis to the Control Constraint Image

To perform the perturbation analysis and make it precise, we define a safety potential that is a measure of pairwise overlap between the claimed sets. This enables us to consider change in the safety potential over time with respect to change of our state and with respect to the state of the contender (given some specific choice for our control action and the control action of the contender). There is some flexibility in how to define the safety potential. We just need it to reflect a measure of set intersection in the sense that it should be positive where there is set intersection and zero elsewhere. One choice that leads to practical calculations is to use a function (such as the sum) of the time intervals between when the first intersection occurs and when the actors are fully stopped when following the safety procedure. This choice has strong correlation with the impact velocities. The time between the first intersection and fully stopped is how long a vehicle would have continued to move after the collision time if the contender had not been there. It is possible to make the safety potential smooth and to make sure that it rejects intersections before they actually happen by making a smooth but tight bump function that rises right before overlap occurs, as illustrated in the figure below.

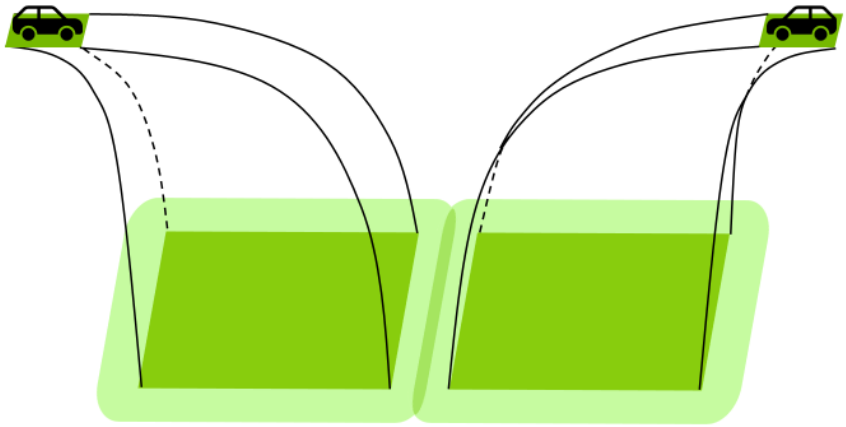


Fig 28. The safety potential can be a smooth and tight function that rises right before overlap occurs. Abstractly, the dark shaded sliver near the overlap illustrates the safety potential.

We illustrate the safety potential as a cost landscape in the figure below, where the state of our vehicle and the state of the contender are imagined as two axes of the function domain. That is, the safety potential changes due to our motion, and due to the contender's motion. We can write out the change in the safety potential with the chain rule, which results in one term due to our motion plus one term due to the contender's motion.

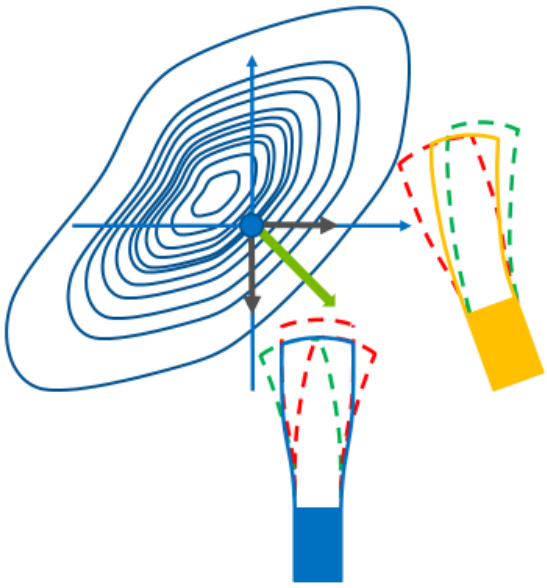


Fig 29. An abstract illustration of the safety potential, as it depends on our motion and the motion of the contender. The safety procedure always makes an actor's contribution to the safety potential 'downhill' (or not uphill to be precise). Our main principle is that each actor should do at least as well as the safety procedure. Thus, if both actors follow the rules, they both contribute 'downhill'. By virtue of the chain rule, that means the total change is not uphill. Thus, no collisions can occur.

This is the beauty of the chain rule, in that it disentangles the effect of our action from the effect of the contender's action. In a multi-agent game with discrete time steps, we have the game tree. In a continuous-time multi-agent scenario, we have the chain rule. We have no control over the other actor, so we can concentrate on our contribution and doing our part. We can now state our core principle in a more complete and accurate form:

All actors are required to apply control actions that contribute at least as much as the safety procedure to improving the safety potential.

Everything follows from this single principle. A key property is that the safety braking profile and the maximum braking profile are defined in such a way that the claimed set is non-expanding over time when the safety procedure is applied. That is, if a vehicle stays between the maximum braking profile and the safety braking profile, and we re-calculate those profiles at some time later, the bounded set of trajectories at the new time instance are contained within the bounds from the earlier time instance, as illustrated in the figure below. Since the safety procedure does not grow our claimed set, our contribution to the safety potential is at most zero if we choose the safety procedure. Since the safety procedure is always allowed, we are always guaranteed to be able to not increase the safety potential relative to any other actor. Since our main principle is to always do as well as the safety procedure, our main principle guarantees that we do not contribute to deteriorating the safety potential (with respect to any other actor). In other words, we guarantee that we do not contribute to increasing danger. If all actors follow these requirements, the chain rule shows that the safety potential cannot increase. That is how the main result is achieved that if the definitions are followed by all, there will be no collisions.

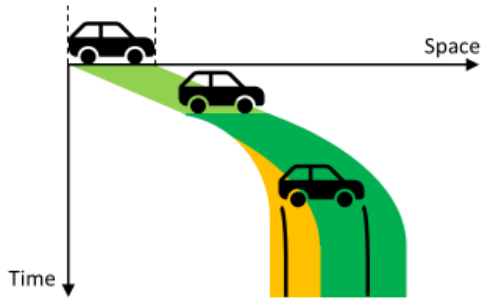


Fig 30. A key property is that the claimed set does not grow over time when applying the safety procedure. Since all actors are required to apply the safety procedure before an intersection of claimed sets happen, it follows by contradiction that an intersection can never occur.

If we analyze how our choices of control action affect the change in the safety potential (via our state), we can implicitly or explicitly calculate the value of each choice of control action. We can then compare them to the safety procedure. While the theory extends to any number of control dimensions, in the case of a normal vehicle, we only have two control dimensions and can display this as an image where the horizontal axis selects a steering action and the vertical axis selects a braking/acceleration action. We can binarize this image to indicate which control actions are acceptable and which are not. We call the result the control constraint image, illustrated in the figure below.

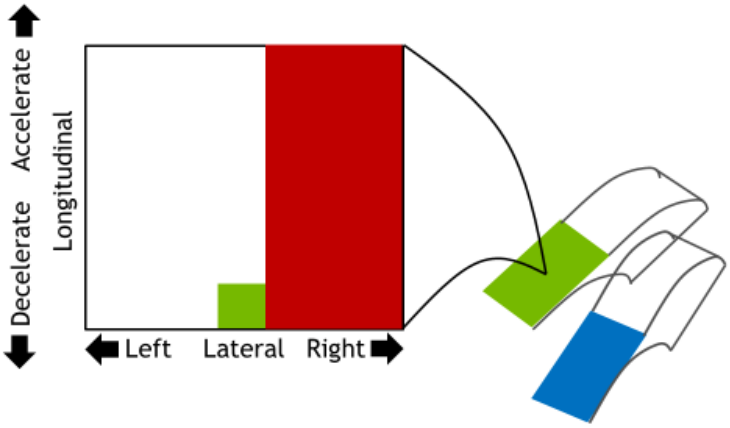


Fig 31. The control constraint image for the green car on the left while being limited by another car from the right. The image shows the acceptable control actions (in white and green) and unacceptable control actions (in red), with steering going from left to right and acceleration towards the top, deceleration towards the bottom. The safety procedure is shown as a small green square at the bottom.

We can link these variable spaces (the control parameters, the vehicle state, the safety potential value) together via the chain rule. A choice of control results in a change of state over time. Change of state over time results in a change in the safety potential. The derivative of the safety potential with respect to state is what we actually call the Safety Force Field. The link is illustrated in the figure below.

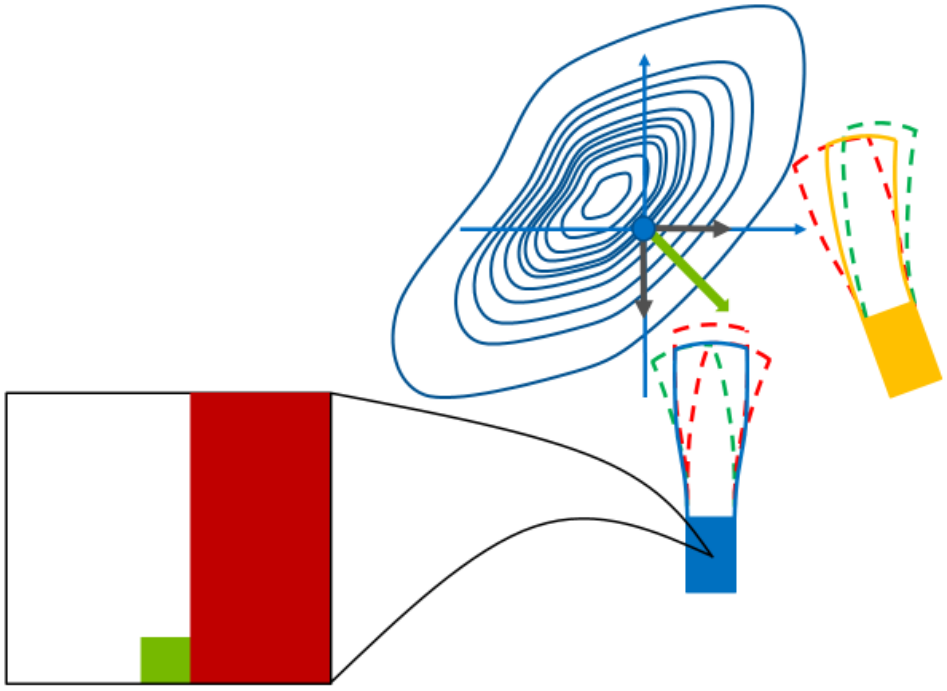


Fig 32. The chain rule links the control to the vehicle state change over time and in turn the state change to the safety potential. What we actually call the Safety Force Field is the gradient of the safety potential with respect to the state (the derivative of the safety potential with respect to state change). Thus, it indicates the direction in state space that is the safest direction, and provides downhill pull.

Appendix B: Visualizing the Safety Force Field

It is always useful to be able to visualize software and what it is doing. In this case, we would like to visualize the constraints that the Safety Force Field is monitoring. One way to do that is to draw the claimed sets as a volumetric visualization. Another approach that allows us both to relate the constraints to sensor data such as a camera projection, and to easily spot intersections, all in a single glance, is the following. Consider each time-slice in space-time. Since we would like to monitor that there is no interference with our claimed set in any of the time-slices, we can align all the time-slices by applying a warp to each time-slice so that our own bounding shape remains in place and occupies the same area across all time slices. This makes our claimed set a vertical column in the transformed volume. Then we can essentially 'look down the time-axis' and see if anything comes close to us. We can further produce an image that contains that as an orthographic projection along the time-axis, and we can think of it as an image in the ground plane. That image can be viewed from a birds-eye perspective. We can also go further and perspective-project it as an overlay into our camera views. If we do, we

get a visualization that smears claimed sets starting at current obstacle positions (so that it is easy to understand what causes them) since the current time slice is included and positioned without a transformation. If we tailgate someone, their shape stretches towards us. If we get tailgated or someone cuts us off too closely, the shape of the offender stretches towards us. If a lane change is not safe because a vehicle is coming up from behind at high speed, their shape stretches out and claims the lane next to us. The key thing to watch for is smeared sets that reach our own bounding shape. When that happens is when the Safety Force Field kicks in. We call this the Safety Force Field relative view.

Note that in general our claimed set can include rotation and expansion of our bounding shape from the first time slice. In this case, the warp of each time-slice is more than just a translation. A good choice is to translate and rotate our shape in each time-slice to center it, and then to subtract a fixed radial amount from each ray starting at the origin, so that our bounding point on that ray aligns between the warped time slice and time zero. By doing this, we push a small part of each ray into the origin and take away a small amount. This is exact as long as our shape is convex, in the sense that overlaps happen if and only if the claimed sets intersect.

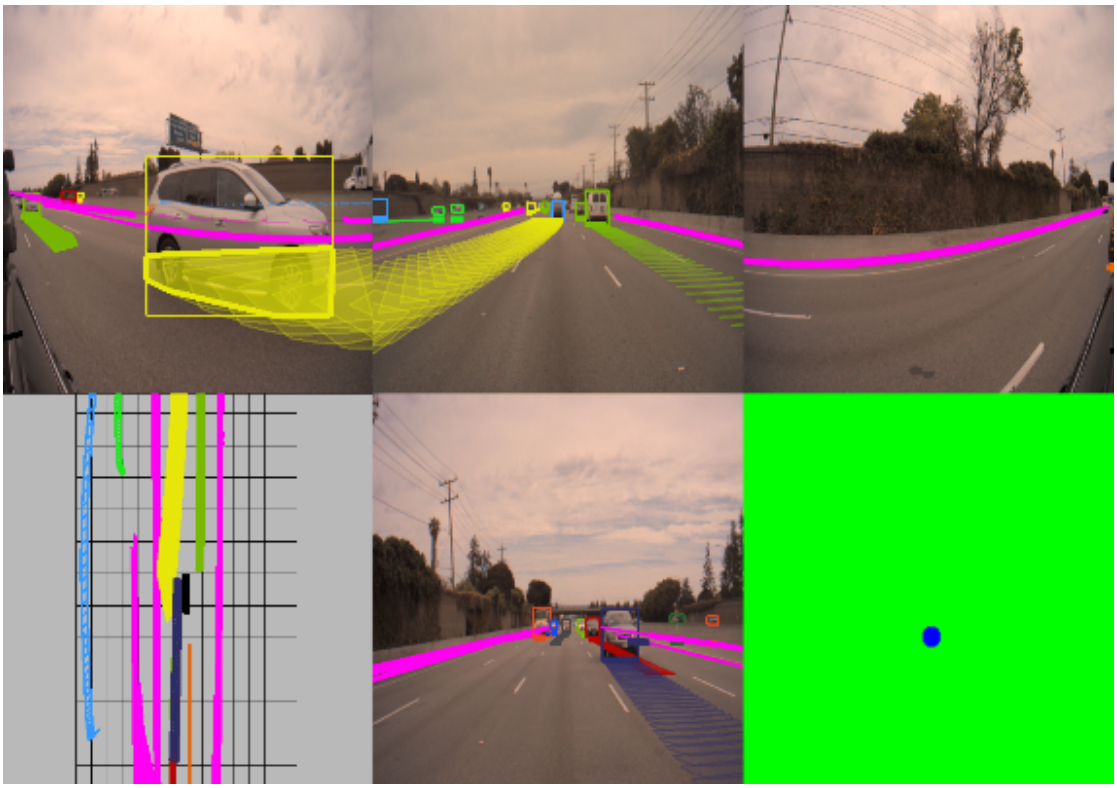


Fig 33. The Safety Force Field relative view. Claimed sets are smeared from their current obstacle positions in a way that we can monitor for ones that come close to us. This provides a very direct visualization of the Safety Force Field constraints that is easily read in a single glance. Note constraints from moving and static obstacles. Lower Left: Birds-eye view. Lower Right: Control constraint image, which is currently green since no claimed sets come too close to us.

For a mathematical deep dive, read The Safety Force Field.