ADS GOING LIVE – WHAT DO WE NEED TO SHOW, AND WHICH IS THE BEST SPORTS METAPHOR?

MIKAEL LJUNG AUST, PhD
Senior Technical Leader Collision Avoidance
Volvo Cars Safety Centre

ADS GOING LIVE – WHAT DO WE NEED TO SHOW?

In launching an ADS, Volvo will have to present convincing arguments that this is an ethically responsible action three times:

1. Before we start selling the ADS vehicle (with all ADS hardware inside)
   - To the public/media
   - To Governments + type approval services
   - To the insurance company that will insure these cars
   - To ourselves (Volvo Cars internally)
2. Before we enable the ADS function itself
   - To the public/media
   - To Governments + type approval services
   - To the insurance company that will insure these cars
   - To ourselves (Volvo Cars internally)
3. When the first crash happens (where ADS might have been enabled)
   - To the public/media
   - To Governments + type approval services
   - To the insurance company that will insure these cars
   - To ourselves (Volvo Cars internally)

ADS GOING LIVE – WHAT DO WE NEED TO SHOW?

In an ethical perspective, there are two things that we need to argue:

- That the car will not repeat “old” mistakes that lead to crashes – i.e. the ADS should not repeat, or fall prey, to known human errors in traffic

- That the car will not make “new” types of dangerous mistakes – i.e. the ADS should not make robot errors that lead to new types of crashes

ADS GOING LIVE – WHAT DO WE NEED TO SHOW?

- That the car will not repeat “old” mistakes that lead to crashes – i.e. the AD car should not repeat, or fall prey, to known human errors in traffic

Human errors can be dealt with using our reference driver model

PEDAGOGICAL OUTLINE

Statistics and the public don’t get along - we can’t say it’s just a low probability event if we crash

Instead, we need to know, and be able to explain, that we have done our best to avoid that particular crash in development

We want to be able to say to ourselves: “if it happens it’s extremely unfortunate; however, no one else could have done any better”

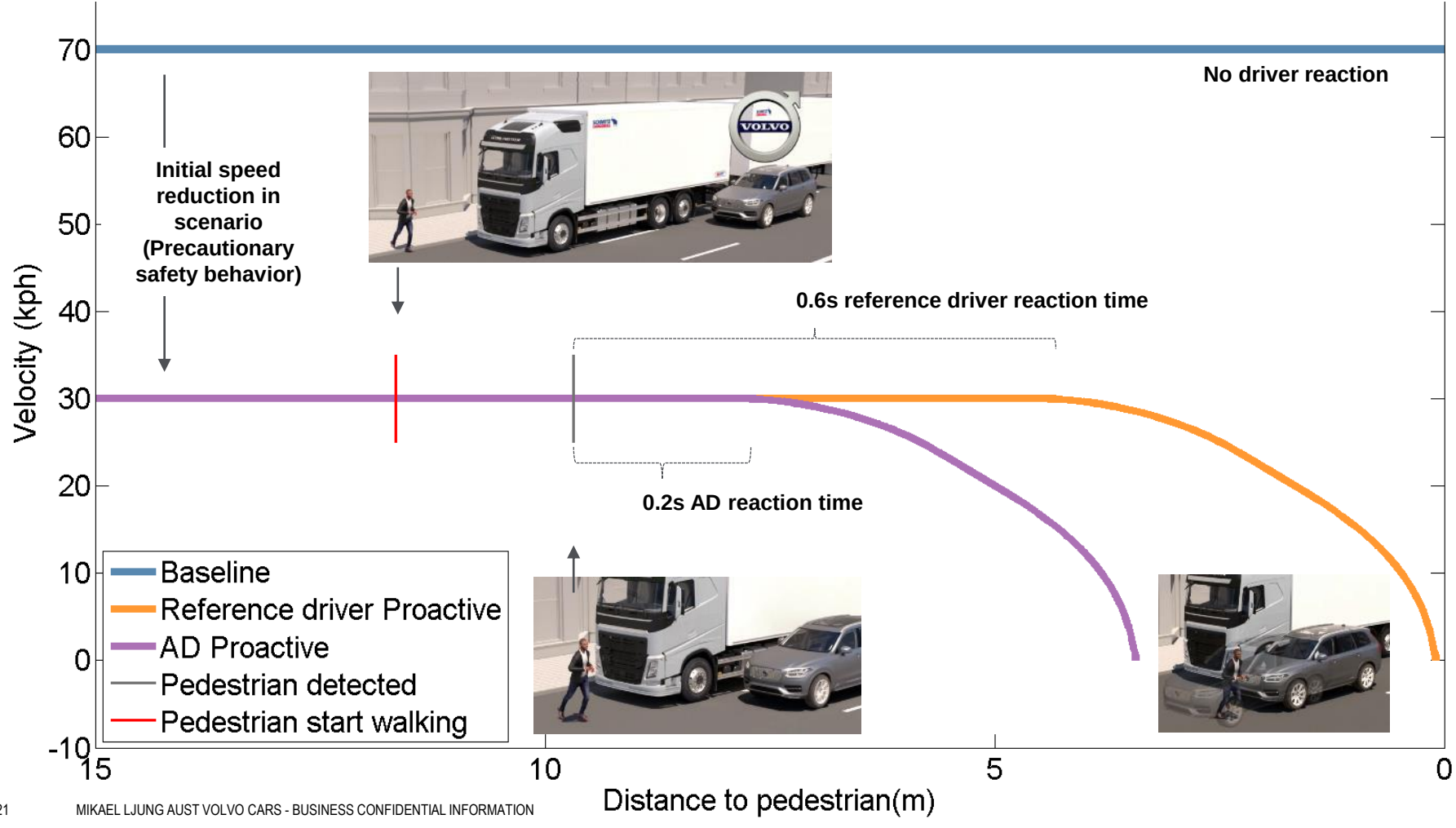
An AD vehicle should not make human errors

Crashes happen when people have a bad day

Let’s take all those crash situations and see what the outcome would be if you had a good day instead (how would the reference driver behave in those situations)

When we know that; lets make sure the AD car performs even better

ATTENTIVE, SKILLED, EXPERIENCED REFERENCE DRIVER MODEL



ADS GOING LIVE – WHAT DO WE NEED TO SHOW?

- Simulating human-error based collisions is not enough to demonstrate safety across all conflict scenarios an ADS may experience
- An ADS cannot perform some human performance degrading activities (hard to distract, possibly impossible to get drunk)
- However, it’s perfectly possible to conceive that it could run a red light
- Given a few minutes and reasonable imagination, one can also think of other, new, ways in which robots may fail

ADS GOING LIVE – WHAT DO WE NEED TO SHOW?

- That the car will not make “new” types of dangerous mistakes – i.e. the ADS should not make robot errors that lead to new types of crashes

”No robot errors” are dealt with by following standards for all relevant problem categories (examples below)

ADS should not break down - Hazard analysis ISO-26262	ADS could do wrong even if everything is working right – SOTIF ISO PAS 21448	Hackers are after your ADS – UN Regulation No. 155, ISO/SAE 21434	ADS should be able to talk to others – SAE 2735, SAE 2945, 11p. in IEEE 802, IEEE 1609 series, etc

AD GOING LIVE – WHAT DO WE NEED TO SHOW?

There are two things we need to show:

1. That the car will not repeat “old” mistakes that could lead to crashes – the AD car should not repeat known human errors in traffic

    - Using the reference driver approach, we can say: “this car was developed to avoid all known human errors (and multiple variants of those), so while we’re truly sorry that the crash occurred, we know that a really good human driver could not have done any better”

ADS GOING LIVE – WHAT DO WE NEED TO SHOW?

There are two things we need to show:

2. The car should not make “new” types of dangerous mistakes – i.e. the AD car should not make robot errors that lead to new types of crashes

   - By fulfilling 26262, SOTIF and other relevant standards in a well documented way, we can say: “this ADS was developed using state-of-the-art requirements for both potential faults as well as problems that may occur even when things are working ok. Therefore, while we’re truly sorry that a crash has occurred, we do not think any other engineering team could have done better”

MY PROPOSAL – DON'T MIX THE TWO

Human errors can be dealt with using our reference driver model

PEDAGOGICAL OUTLINE

Statistics and the public don't get along - we can't say it's just a low probability event if we crash

Instead, we need to know, and be able to explain, that we have done our best to avoid that particular crash in development

We want to be able to say to ourselves: "if it happens it's extremely unfortunate; however, no one else could have done any better"

An AD vehicle should not make human er

Crashes happen when people have a bad

Lets take all those crash situations and se
outcome would be if you had a good day i
would the reference driver behave in thos

When we know that; lets make sure the A
performs even better

For now, these should likely be treated as separate items

–

Both represent unknown quantities, therefore we cannot use potential positive results in one to offset potential negative results in the other

-

both need to stand their own ground in first iterations

"No robot errors" are dealt with in multiple problem categories (examples below)

uld not own alysis 262	Things could go wrong even if everything is working right – SOTIF ISO PAS 21448	Hackers are after your ADS – UN Regulation No. 155, ISO/SAE 21434	ADS should be able to talk to others – SAE 2735, SAE 2945, 11p. in IEEE 802, IEEE 1609 series, etc

GOING LIVE WITH ADS IS ETHICALLY LIKE QUALIFYING FOR AN OLYMPIC DECATHLON - TWICE

- Decathlon combines four runs, three jumps and three throws
  - to qualify, you need Olympic level performance in all events – can't skip any!

- AND we must qualify in both the robot error and the human error versions

1. Before we start selling the ADS vehicle (with all ADS hardware inside)
   - To the public/media
   - To Governments + type approval services
   - To the insurance company that will insure these cars
   - To ourselves (Volvo Cars internally)
2. Before we enable the ADS function itself
   - To the public/media
   - To Governments + type approval services
   - To the insurance company that will insure these cars
   - To ourselves (Volvo Cars internally)
3. When the first crash happens (where ADS might have been enabled)
   - To the public/media
   - To Governments + type approval services
   - To the insurance company that will insure these cars
   - To ourselves (Volvo Cars internally)

Human errors can be dealt with using our reference driver model

PEDAGOGICAL OUTLINE

Statistics and the public don't get along - we can't say it's just a low probability event if we crash

Instead, we need to know, and be able to explain, that we have done our best to avoid that particular crash in development

We want to be able to say to ourselves: "if it happens it's extremely unfortunate; however, no one else could have done any better"

An AD vehicle should not make human errors

Crashes happen when people have a bad day

Lets take all those crash situations and see what the outcome would be if you had a good day instead (how would the reference driver behave in those situations)

When we know that; lets make sure the AD car performs even better

"No robot errors" are dealt with in multiple problem categories (examples below)

Things should not break down	Things could go wrong even if everything is working right	Hackers are after your ADS	ADS should be able to talk to others
- Hazard analysis ISO-26262	– SOTIF ISO PAS 21448	– UN Regulation No. 155, ISO/SAE 21434	– SAE 2735, SAE 2945, 11p. in IEEE 802, IEEE 1609 series, etc