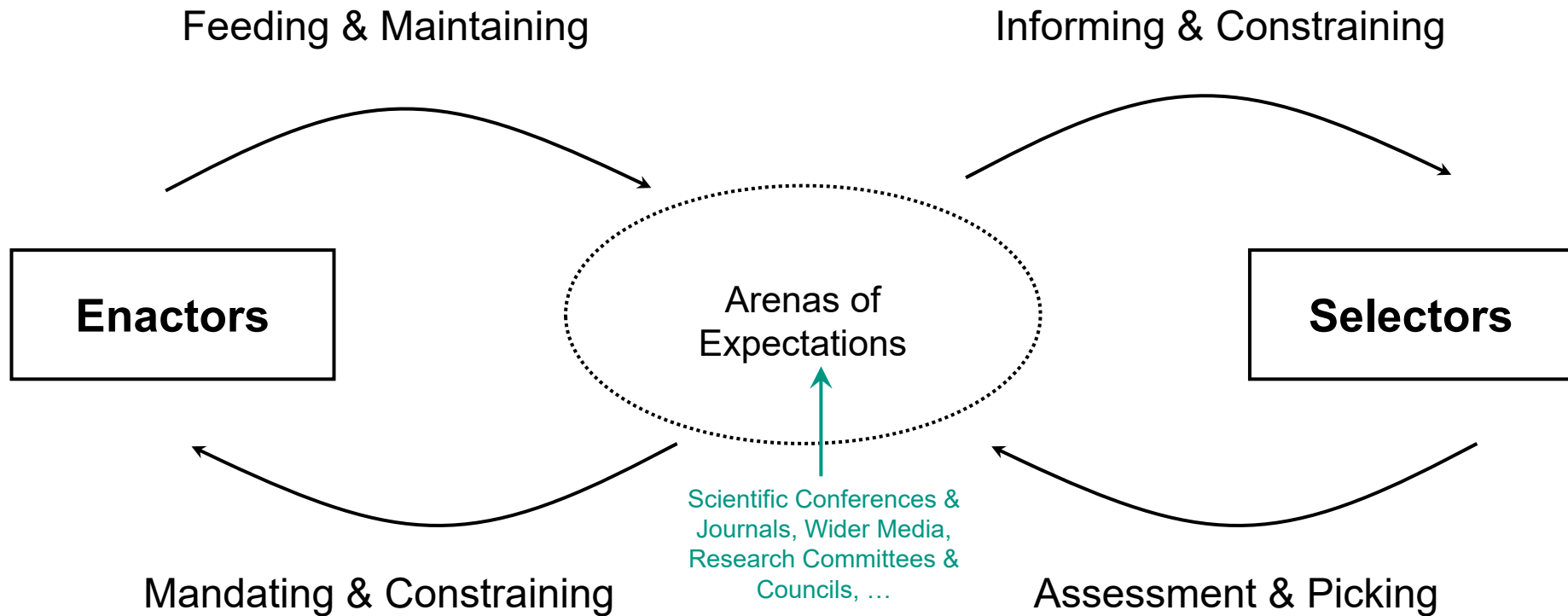


# The Troubles with ‚Safety‘ & ‚Acceptance‘ Observations from the Sideline – and a Proposal

Torsten Fleischer



# Starting Points (1): Selector-Enactor Games



Source: Garud/Ahlstrom 1997, Rip/te Kulve 2008, Bakker et al. 2011

## Starting Points (2)

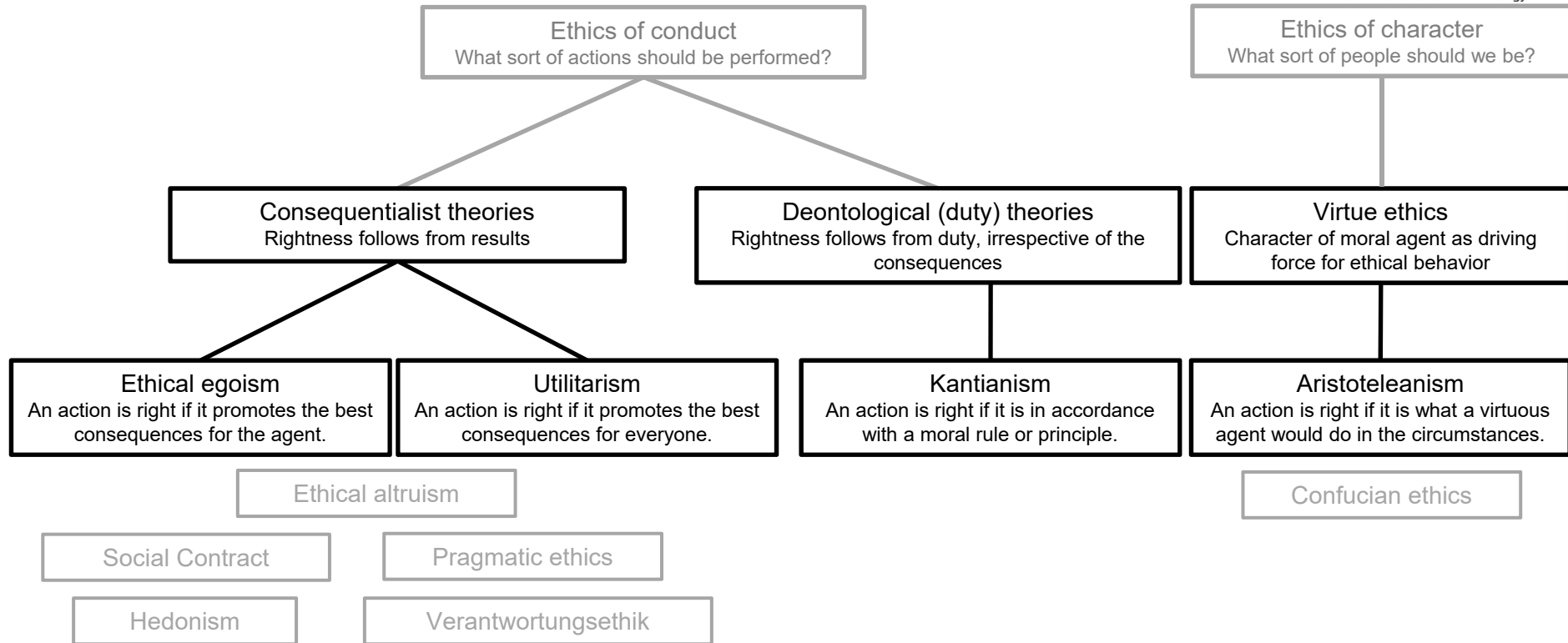
- Definitions are neither true nor false. They can be useful or useless.
- Definitions serve purposes.
- **Safety** plays a role in CAD discourses in various meanings:
  - (Rather metaphorical) as a (likely the most important) ‘accepted promise’ for actor coordination in innovation processes / justification for development and deployment
  - Perceived safety as an antecedent for behavioral intentions to use / buy AVs (‘predictor for adoption behavior’)
  - Design criterion for developers and manufacturers of AVs
  - Assessment criterion for regulators (e.g. within the type certification processes for AVs)
  - Collective safety gains / losses as well as redistributions of individual risks as outcomes of the wider diffusion of CAD

# ISO 26262 / 21448 Vocabulary: Safety, risk et al.



- **functional safety**: absence of *unreasonable risk* due to *hazards* caused by *malfunctioning behaviour* of *E/E systems* (ISO 26262)
- **safety of the intended functionality (SOTIF)**: absence of *unreasonable risk* due to *hazards* resulting from *functional insufficiencies* of the intended functionality or its implementation (ISO DIS 21448)
- **malfunctioning behaviour**: *failure* or unintended behaviour of an *item* with respect to its design intent
- **functional insufficiencies**: insufficiency of specification (e.g., incompleteness) or performance limitation (e.g., of technical capabilities) of the intended functionality (i.e. specified function on *vehicle level*)
- **hazard**: potential source of *harm* caused by *malfunctioning behaviour* of the *item / hazardous behavior* of the *system*
- **harm**: physical injury or damage to the health of persons
- **unreasonable risk**: *risk* judged to be **unacceptable** in a certain context according to **valid societal moral concepts**
- **risk**: combination of the probability of *occurrence* of *harm* and the *severity* of that *harm*

# A (simplified) taxonomy of ethical theories



# Acceptance and Acceptability



**Acceptance** as an *empirical* phenomenon. (What is accepted? / What will be accepted?):  
Different ways to conceptualize / measure acceptance: actual use, (behavioral) intention to use, considered appropriate, tolerated, absence of conflict,...

Challenges: Measurement concept, predictability, scalability, extrapolation, temporal stability, plurality of individual preferences and values vs. collective benefit, ...

**Acceptability** as a *normative* approach (What should be accepted?)

a) Derived from current risk (taking) behavior, using rationality and consistency criteria (inconsistency is seen as an indicator for non-rationality)

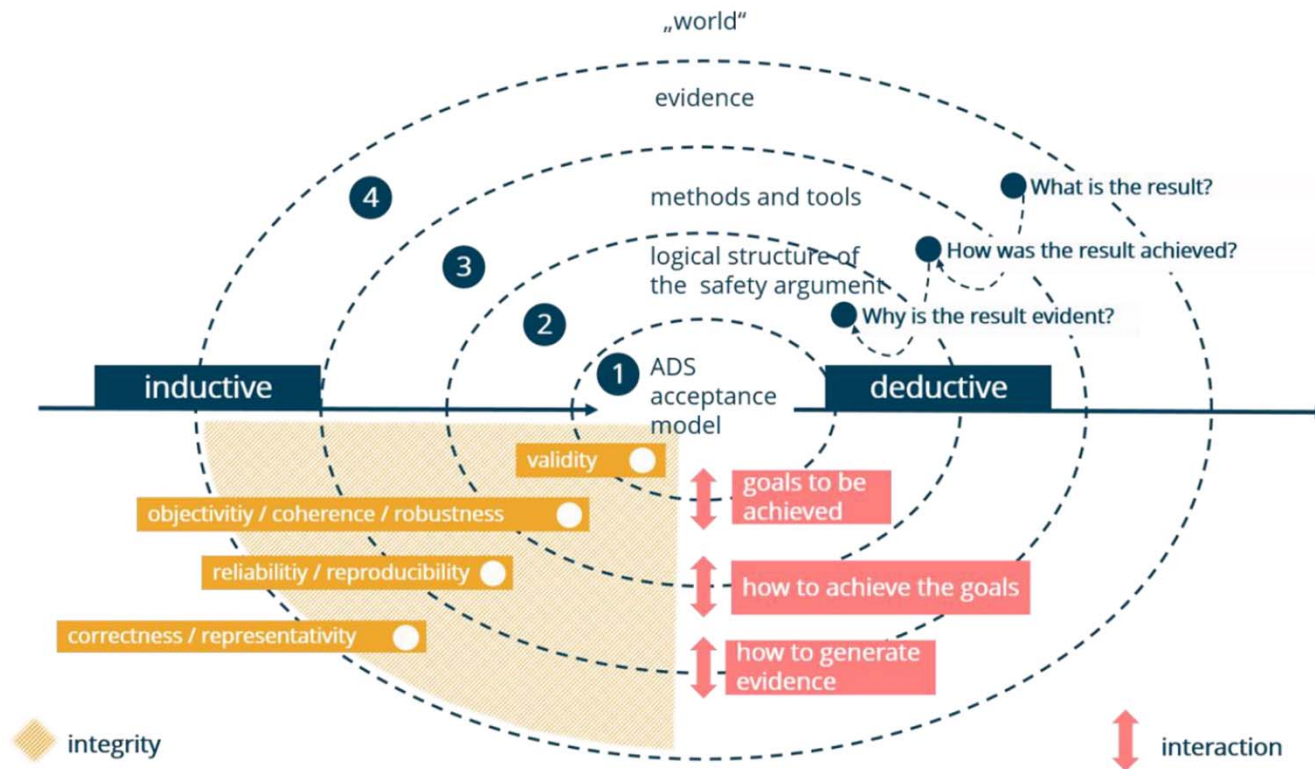
Challenges: Quantifying and comparing different risks (unified scales presuppose decontextualization), rationality and consistency are no prerequisites for social interaction and political participation

b) Broadly accepted procedures of decision-making, applied by democratically legitimated institutions, lead to commonly binding (risk taking) decisions

Challenges: (perceived) erosion of democratic standards

Sources: partially based on Gethmann/Sander 1999, Petermann/Scherz 2005, Grunwald 2005

# From the 'PEGASUS Safety Argumentation'



Source: PEGASUS Safety Argumentation. <https://www.pegasusprojekt.de/files/tmpl/pdf/PEGASUS%20Safety%20Argumentation.pdf> (12.05.2022)

# From the ‘PEGASUS Safety Argumentation’



“Proposal for a framework to support an approval recommendation particularly aimed at highly automated driving functions.”

## *Layer 1 – ADS acceptance model*

PEGASUS embeds the first layer of the ADS acceptance model in a large context. The specifics are not the focus of PEGASUS. *The key element of this layer is a scientific model for describing the dependence of the social acceptance for Automated Driving Systems from several factors. A key premise here is that individual or social acceptance cannot be explained with a single cause. This premise is in line with established models on technology acceptance such as the Technology Acceptance Model (TAM), the Theory of Planned Behaviour (TPB), and the Unified Theory of Acceptance and Use of Technology (UTAUT).*

Depending on the model, different factors are postulated: Attitude, Perceived Usefulness, Perceived Ease of Use, Subjective Norms, Perceived Behavioural Control, Performance Expectancy, Effort Expectancy, Social Influence, Perceived Enjoyment. [...]

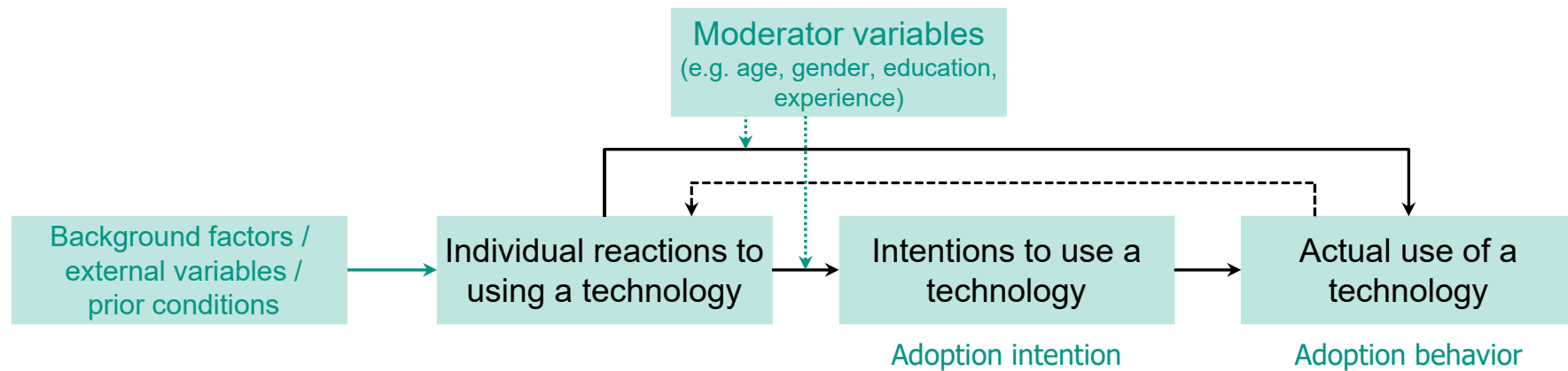
Since in PEGASUS the focus is on the verification of safety and reliability of highly automated driving functions, it is proposed that these be subsumed under the factor of *Performance Expectancy*<sup>1</sup>. This allows a connection to be made between the first layer and the second layer, the presentation of the logical structure of the safety argument. As part of the PEGASUS Safety Argumentation, layers 2, 3 and especially 4 are to be understood as an operationalisation of the Performance Expectancy factor (in particular here safety and reliability).

Fn1: Since none of the existing models are further developed and no new model is proposed as part of PEGASUS, we are attempting here to define an interface to the existing research in this field.

Source: PEGASUS Safety Argumentation. <https://www.pegasusprojekt.de/files/tmpl/pdf/PEGASUS%20Safety%20Argumentation.pdf> (12.05.2022)



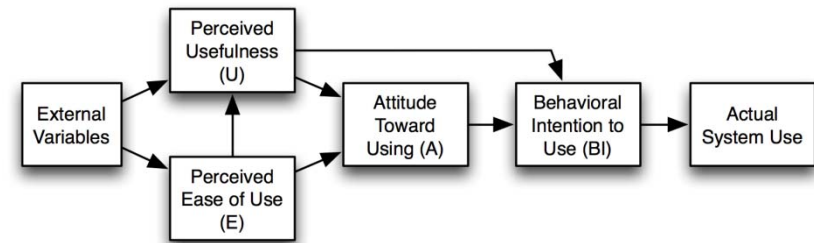
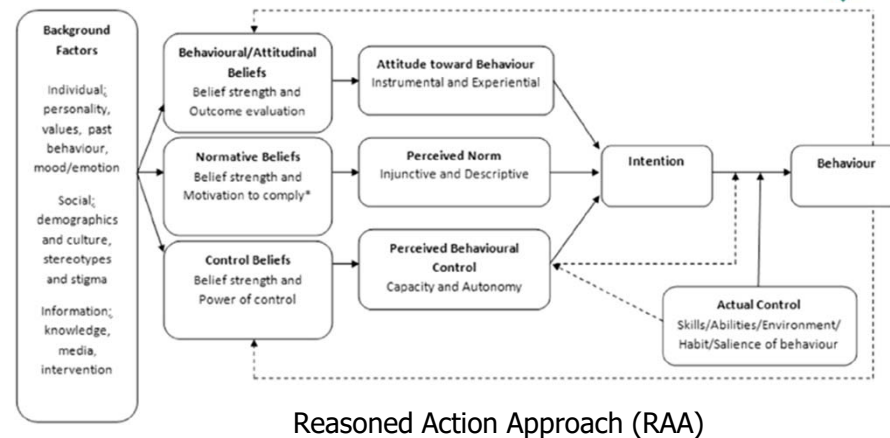
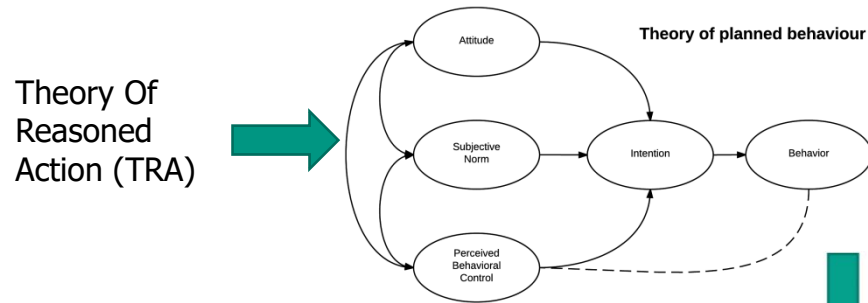
# Basic Concept of User Acceptance Models



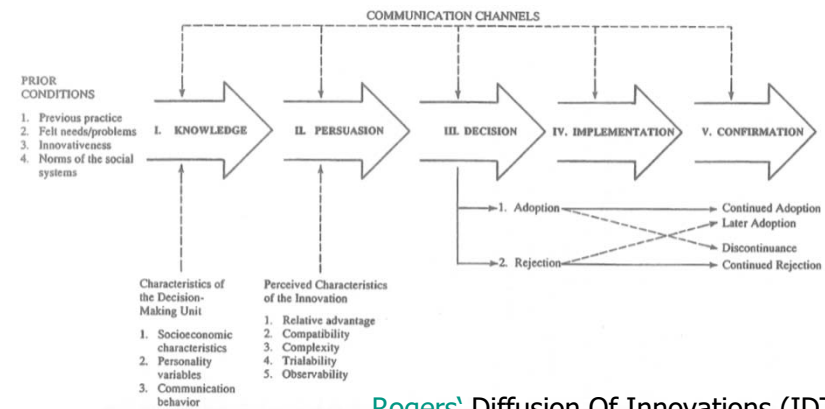
# Influential Models



## The Ajzen-Fishbein B-I Models



Davis' Technology Acceptance Model (TAM)



Rogers' Diffusion Of Innovations (IDT)

# User Acceptance Models



Diffusion Of Innovations Theory (DOI/IDT) (Rogers 1962/ 2003)

Social Cognitive Theory (Bandura 1963/1977)

Theory of Reasoned Action (TRA) (Fishbein & Ajzen 1967/1975)

Theory of Interpersonal Behaviour (TIB) (Triandis 1979)

Theory of Planned Behaviour (TPB) (Ajzen 1985/1991)

Technology Acceptance Model (TAM) (Davis, 1989) / TAM2 (Ventakesh & Davis 2000) / TAM3 (Ventakesh & Bala 2008)

Model of PC Utilization (MPCU) (Thompson et al. 1991)

Motivational Model (MM) (Davis, Bagozzi & Warshaw 1992)

Igbaria's Model (IM) (Igbaria, Schiffman & Wieckowski 1994)

Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh et al. 2003) / UTAUT2 (Ventakesh et al. 2012)

Car Technology Acceptance Model (CTAM) (Osswald et al. 2012)

4P Acceptance Model (Nordhoff et al. 2016)

Theory for the Acceptance and Use of Smart Mobility (TAUSM) (Wieker+Kauschke 2018)

Autonomous Vehicle Acceptance Model (AVAM) (Hewitt et al. 2019)

Multi-level Model on Automated Vehicle Acceptance (MAVA) (Nordhoff et al. 2019)

AV Acceptance Meta-framework (AVAM-F) (Keszey 2020)

# Unified Theory of Acceptance and Use (UTAUT)



Focus on Acceptance and Use of *Information Technologies* in Complex Organizational Settings at the *Workplace*

Integrates concepts and findings from eight acceptance models (TRA, TAM, TPB, MM, C-TAM/TPB, MPCU, IDT, SCT)

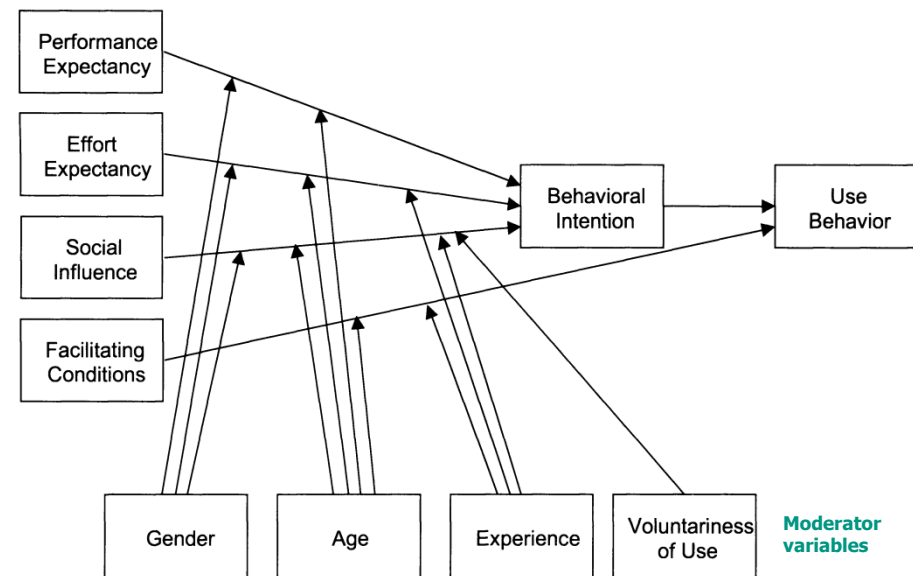
**Performance Expectancy:** degree to which using a technology will help him or her to attain gains in job performance

**Effort Expectancy:** degree of ease associated with the use of the system

**Social Influence:** the degree to which an individual perceives that important others (e.g., family and friends) believe he or she should use the new system

**Facilitating Conditions:** the degree to which an individual believes that an organizational and technical infrastructure exists to support use of the system

Direct determinants of user acceptance and usage behavior



Source: V. Venkatesh, M.G. Morris, Gordon B. Davis, Fred D. Davis: *User Acceptance of Information Technology: Toward A Unified View*. MIS Quarterly, Vol.27 (2003), No.3, pp.425-478

# Unified Theory of Acceptance and Use (UTAUT2)

**Performance Expectancy:** degree to which using a technology will provide benefits to consumers in performing certain activities

**Effort Expectancy:** degree of ease associated with consumers' use of technology

**Social Influence:** extent to which consumers perceive that important others (e.g., family and friends) believe they should use a particular technology

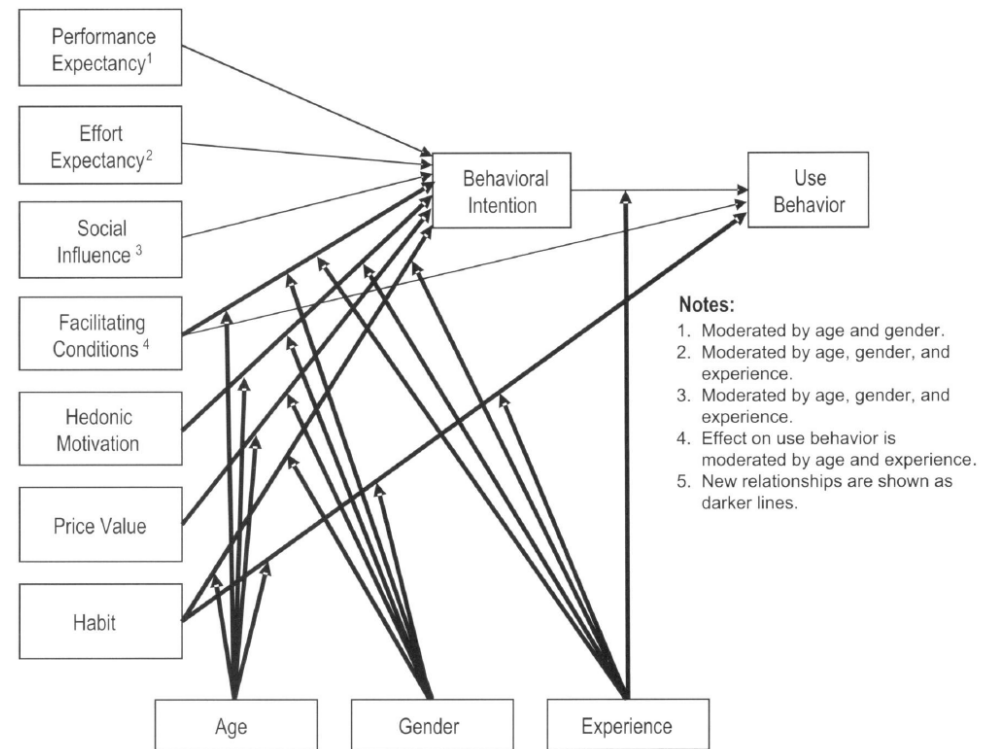
**Facilitating Conditions:** refer to consumers' perceptions of the resources and support available to perform a behavior

**Hedonic Motivation:** the fun or pleasure derived from using a technology

**Price Value:** consumers' cognitive tradeoff between the perceived benefits of the applications and the monetary cost for using them

**Experience:** reflects an opportunity to use a target technology and is typically operationalized as the passage of time from the initial use of a technology

**Habit:** extent to which an individual believes the behavior to be automatic

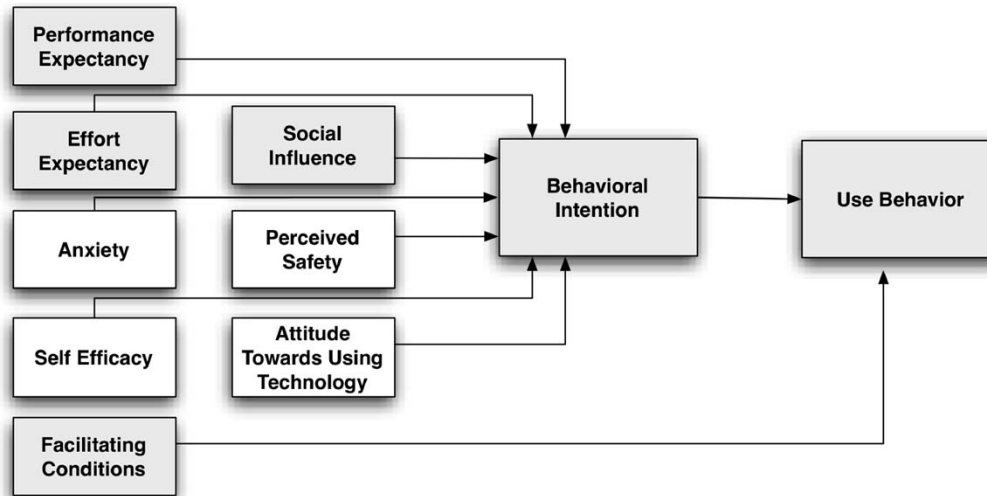


Source: V. Venkatesh, J.Y.L.Thong, X. Xu: *Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology*. MIS Quarterly, Vol.36 (2012), No.1, pp.157-178

# Car Technology Acceptance Model (CTAM)



Expansion of UTAUT



“We define **perceived safety** as the degree to which an individual believes that using a system will affect his or her well-being.

We named the construct *perceived safety* considering the self-reflective character of perceiving a situation hazardous.

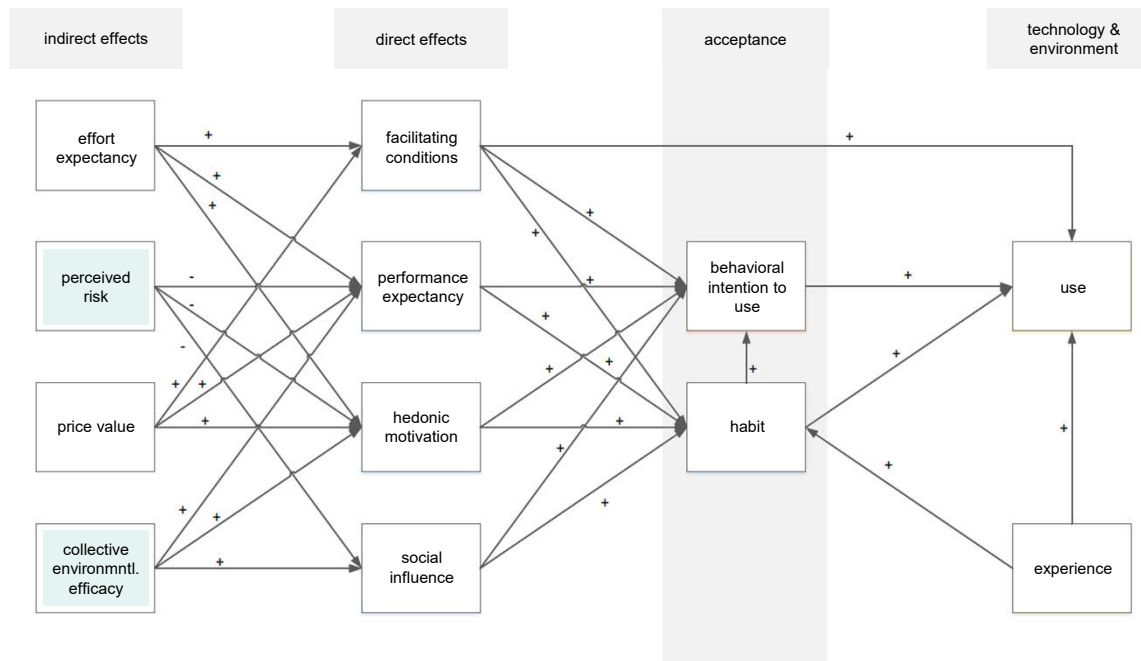
Within the car, this also comprises the judgment of one's own driving skills and safety feeling in relation to other drivers.

The impact of perceived safety is assumed as *critical in the process of predicting the behavioral intention to use*, as the user will estimate the potential effect of safety-related consequences through using an information technology while driving.”

**Source:** Osswald et al. Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '12), October 17–19, 2012, Portsmouth, NH, USA

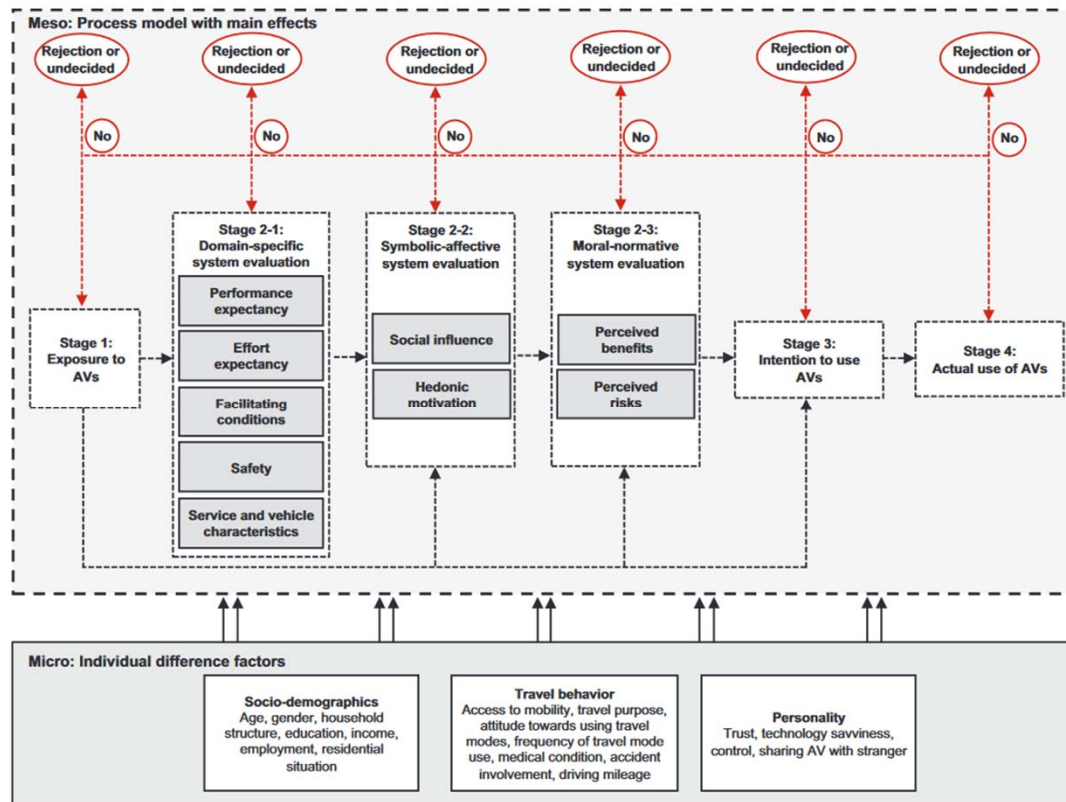
# Theory for the Acceptance and Use of Smart Mobility (TAUSM)

Expansion of UTAUT2



Source: Wieker et al. 2020, modified

# Multi-level Model on Automated Vehicle Acceptance (MAVA)

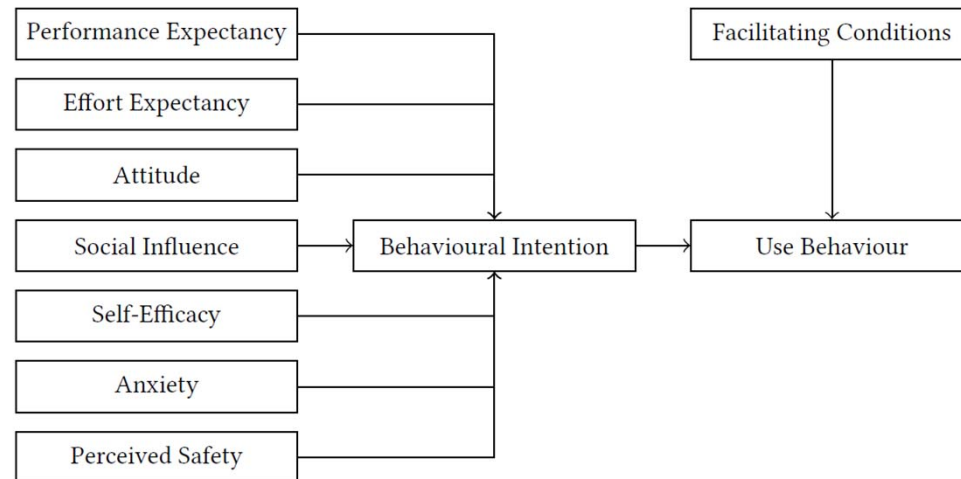


Source: Nordhoff et al. TheoretSsErgonSci Vol.20(2019) No.6, pp.682-710



# Autonomous Vehicle Acceptance Model (AVAM)

Expansion of UTAUT, similar to CTAM



Source: Hewitt et al. Proceedings IUI '19, March 17–20, 2019, Marina del Rey, CA, USA

# Perceived safety in a nutshell



- Feeling safe, attaining a state of *perceived safety* has high importance for humans. It is usually achieved when basic human needs are met. These include, but are not limited to, the **absence of threats** to personal security and health as well as **sufficient predictability, reliability, and order**.
- Perceived safety is especially relevant in situations that require judgments or decisions. These are made in social, environmental and situational **contexts** which may evoke certain thoughts, feelings, or behaviors and are **influenced by social others** (imitation, group pressure).
- When making decisions, humans are not able to consciously consider all available relevant information (*bounded rationality*), especially under time pressure and with increasing complexity of a task. In many cases, they instead use **heuristics**.
- Many heuristics work through *attribute substitution* (substituting a complicated question with a simpler one). These simplifications usually work sufficiently well but may in some instances result in distorted judgments, known as **biases**. (e.g. *zero-risk bias, optimistic bias, overconfidence*)
- Especially in decision making under uncertainty, **feelings** that occur in a certain situation become a source of information (*affect heuristics, feelings-as-information hypothesis*).

Source: based on Raue et al. 2019

# Examples for OVs used to measure Perceived Safety



Observed variables for PS used in [CTAM \(Osswald et al. 2012\)](#)

- PS1 I believe that using the system is dangerous.
- PS2 Using the system requires increased attention.
- PS3 The system distracts me from driving.
- PS4 I feel safe while using the system.
- PS5 Using the system decreases the accident risk.
- PS6 I can use the system without looking at it.

Observed variables for *perceived risk* used in [TAUSM \(Wieker et al. 2020\)](#)

- SM08\_01 Die Verwendung von Smart Mobility ist riskant.
- SM08\_02 Ich vertraue SM-Technologien nicht.
- SM08\_03 SM funktioniert möglicherweise nicht so gut wie herkömmliche Mobilität und verursacht Probleme.
- SM08\_04 Es gibt zu viele offene Fragen rund um SM.
- SM08\_05 Ich habe gewisse Angst vor SM.

Observed variables for PS used in [AVAM \(Hewitt et al. 2019\)](#)

- 24 I believe that using the vehicle would be dangerous.
- 25 I would feel safe while using the vehicle.
- 26 I would trust the vehicle.

Observed variables for PS used in [Nordhoff et al. 2021](#)

- PS1: I feel safe most of the time.
- PS2: I feel relaxed most of the time.
- PS3: I feel anxious most of the time.
- PS4: I feel bored most of the time.
- PS5: I am concerned about my general safety most of the time
- PS6: I entrust the safety of a close relative to my partly automated car.

Observed variables for PS used in [Montoro et al. 2019](#)

- PS1: Overall, AVs would help make my journeys safer than they are when I use conventional cars.
- PS2: AVs would act better than myself in a complicated traffic situation.
- PS3: A driverless/automated vehicle may not be 'smart' enough for guaranteeing my safety during the journey.
- PS4: AV-related systems could easily break down, or be hacked, thus compromising my safety.
- PS5: AVs would respond adequately to unexpected situations that commonly require rapid responses from drivers.

Observed variables for PS used in [Koul&Eydgahi 2020](#)

1. I would trust that a computer in an AV could get me to my destination safely with no assistance from me.
2. I believe an AV would be safer to drive on populated streets when compared to the average human driver.
3. I would be comfortable entrusting the safety of a close family member riding in an AV.
4. I believe an AV would be safer to drive on expressways and highways compared to the average human driver

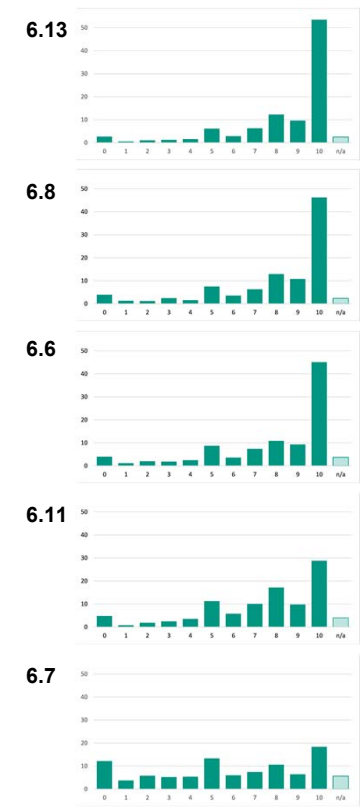
# Safety Design Expectations (Fleischer et al. 2022)

From a Representative Survey of German Population, Nov 2021, n=2001, Automation Level ~SAE L4

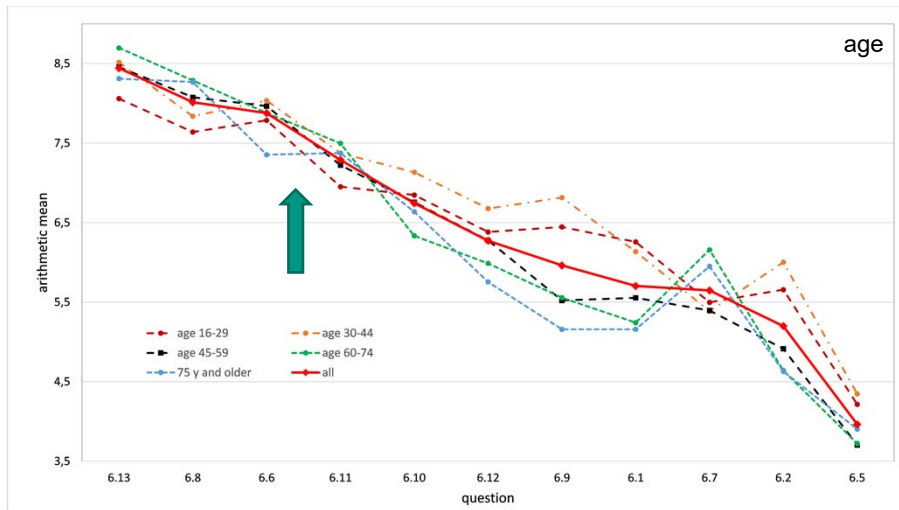
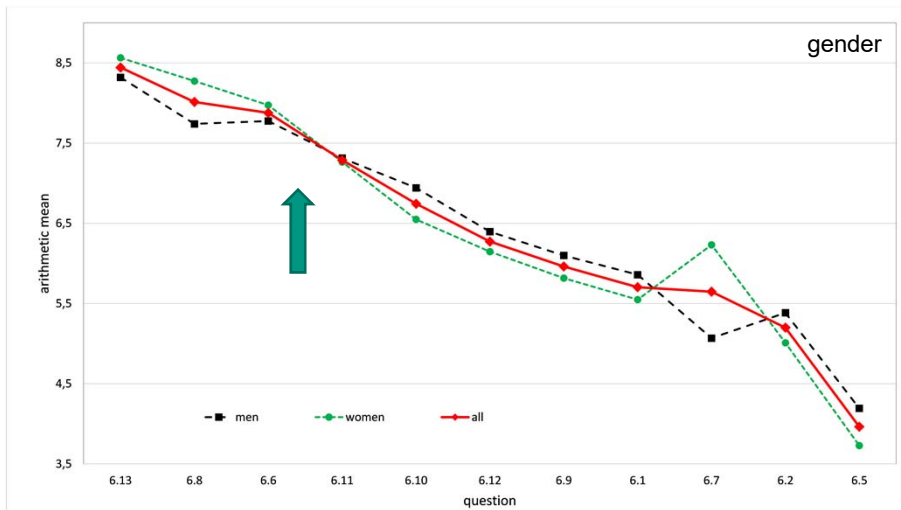
Q6: To make such a development toward autonomous driving possible, some framework conditions of today's traffic might have to be *changed*. Assuming that would include the following changes: *Would you be more likely to welcome or more likely to oppose them?*

- 6.1 financial support for private individuals
- 6.2 existing regulatory framework should be relaxed
- 6.3 AV manufacturers should assume liability for damages.
- 6.4 AV owners should assume liability for damages.
- 6.5 Data protection regulations should be relaxed.
- 6.6 Users should be able to intervene if accidents are imminent.
- 6.7 AV only in their own lanes
- 6.8 every road user can recognize AVs at all times
- 6.9 AV can violate traffic rules if this prevents accidents.
- 6.10 AV to be tested in transparent field trials on public roads.
- 6.11 Citizens should be involved in field trials
- 6.12 Private mobility providers should be given generous testing opportunities
- 6.13 AVs should drive carefully when they perceive vulnerable road users

DE	Averages			Top2-Box	
	ArMean	StdDev	Med	Top	Bottom
6.1	5,70	3,503	6	26%	17%
6.2	5,20	3,147	5	15%	16%
6.3	7,33	2,852	8	41%	5%
6.4	5,55	3,618	5	27%	19%
6.5	3,96	3,390	4	11%	32%
6.6	7,88	2,801	9	54%	5%
6.7	5,65	3,407	6	25%	16%
6.8	8,01	2,730	9	57%	5%
6.9	5,96	3,354	7	25%	15%
6.10	6,74	3,044	7	34%	9%
6.11	7,29	2,767	8	39%	6%
6.12	6,27	2,912	7	22%	10%
6.13	8,44	2,405	10	63%	3%



# Q6 Changing regulations and institutions



6.13 AVs should drive carefully when they perceive vulnerable road users  
 6.8 Every road user can recognize AVs at all times  
 6.6 Users should be able to intervene if accidents are imminent.

6.11 Citizens should be involved in field trials  
 6.10 AV to be tested in transparent field trials on public roads.  
 6.12 Private mobility providers should be given generous testing opportunities  
 6.9 AV can violate traffic rules if this prevents accidents.

6.1 financial support for private individuals  
 6.7 AV drive only in their own lanes  
 6.2 existing regulatory framework for vehicle certification should be relaxed  
 6.5 Data protection regulations should be relaxed.

Source: Fleischer et al. IEEE Intelligent Vehicles Symposium 2022

# A proposal



- Full knowledge about the safety implications of CAD (measured as number and severity of accidents, number of persons affected, redistribution of accident risk, etc.) will be impossible to achieve before deployment.
- Perceived safety and safety design expectations of users and other relevant stakeholders might serve as a substitute, particularly in early deployment. This knowledge is incomplete, partly uncertain, somewhat fuzzy, and could need to be corrected over time.
- Building on PS and SDE does not “automatically” provide for an uncontested way forward. Product and service design as well as strategies for deployment would need to be negotiated. They should be “learning strategies”, adapted over time in close coordination between stakeholders as new knowledge is obtained.
- Limited spaces for experimentation with CAD vehicles / services and the applicable rules and “learning strategies” should be legitimated, ideally by an act of parliament.
- There still will be misjudgments, errors, and – as a consequence – accidents. Be perfectly clear about this in your communication with policymakers, the media and the general public. Please do not overpromise.



# Vielen Dank

[Torsten.Fleischer@kit.edu](mailto:Torsten.Fleischer@kit.edu)

 (0721) 608-24571

[www.itas.kit.edu](http://www.itas.kit.edu)

 [ITAS@KIT](mailto:ITAS@KIT)

Karlstrasse 11

D-76133 Karlsruhe