Designing Trustworthy Driver Assistance Systems

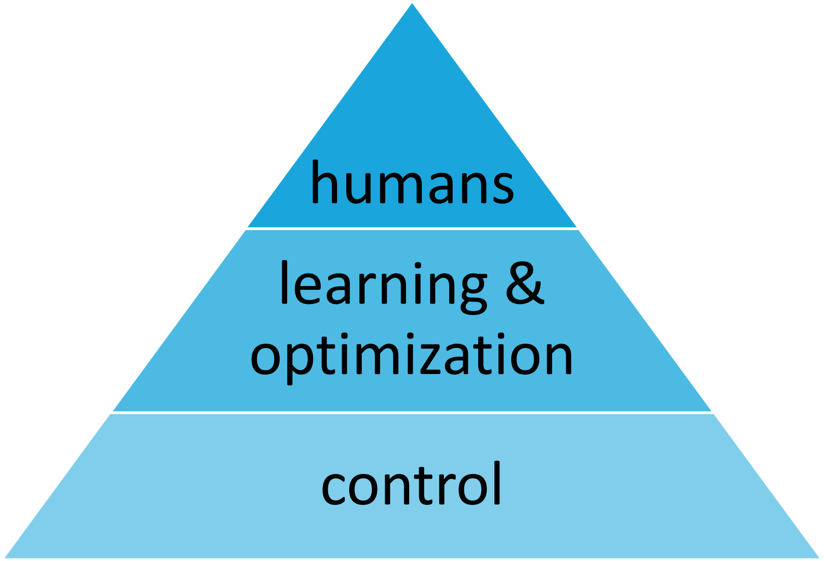
Katie Driggs-Campbell
Assistant Professor, ECE
University of Illinois at Urbana-Champaign

IPAM Tutorials: Human Factors I
September 22, 2020

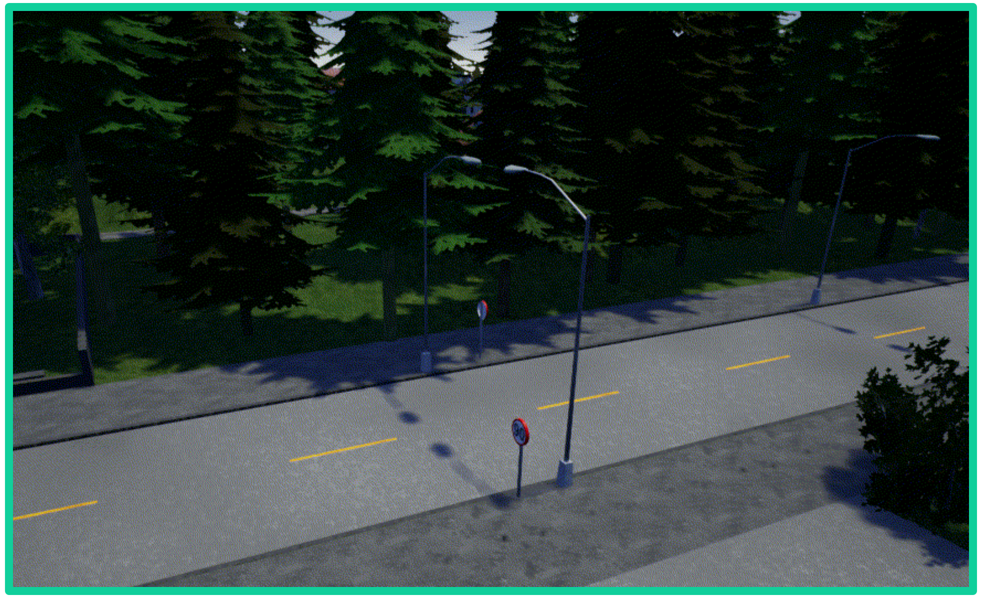
Ye-Ji Mun

HUMAN-CENTERED
AUTONOMY LAB

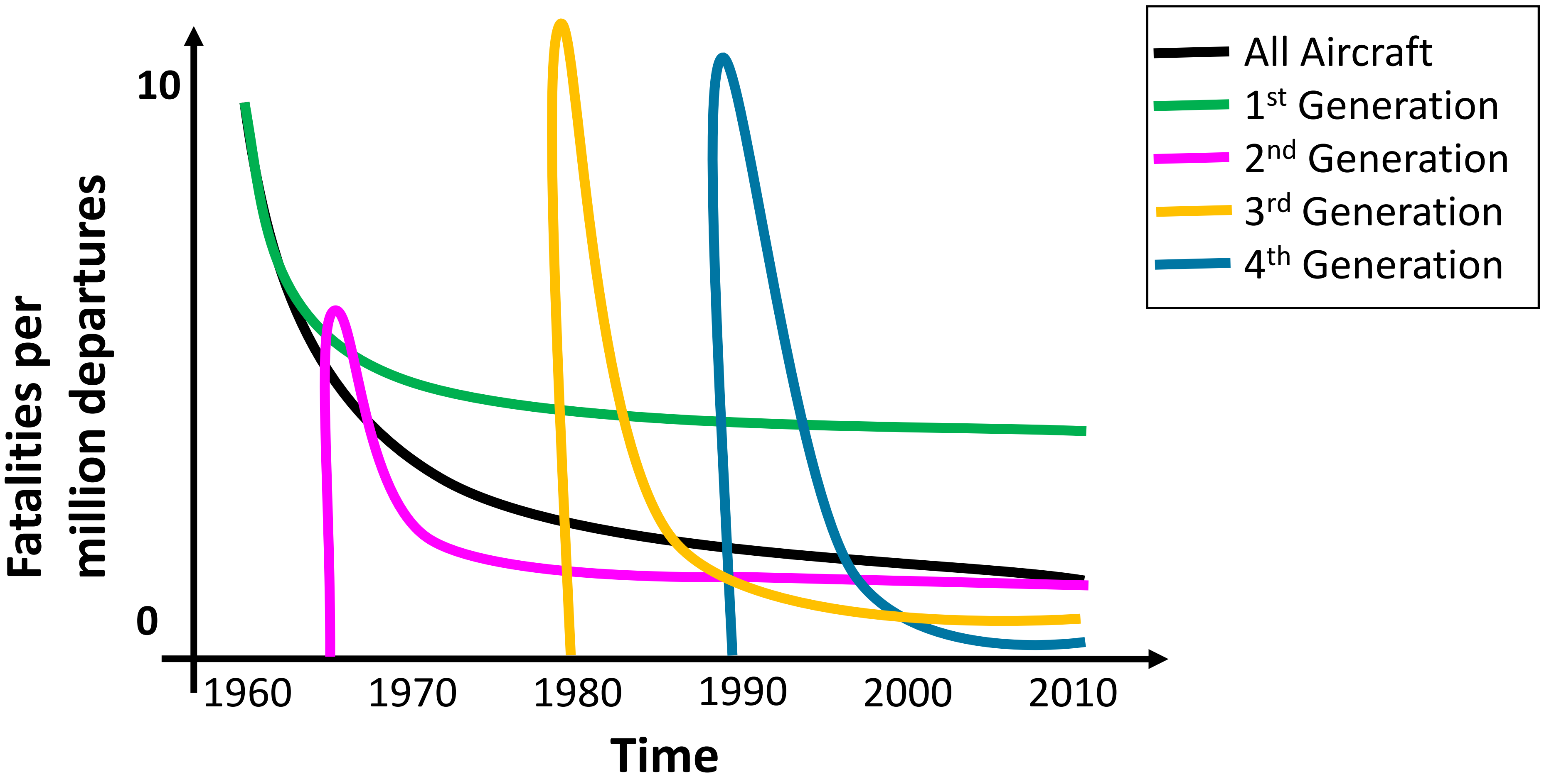
How can we ensure safety in autonomous systems that operate with people in the real-world?



How can we ensure safety in autonomous systems that operate with people in the real-world?



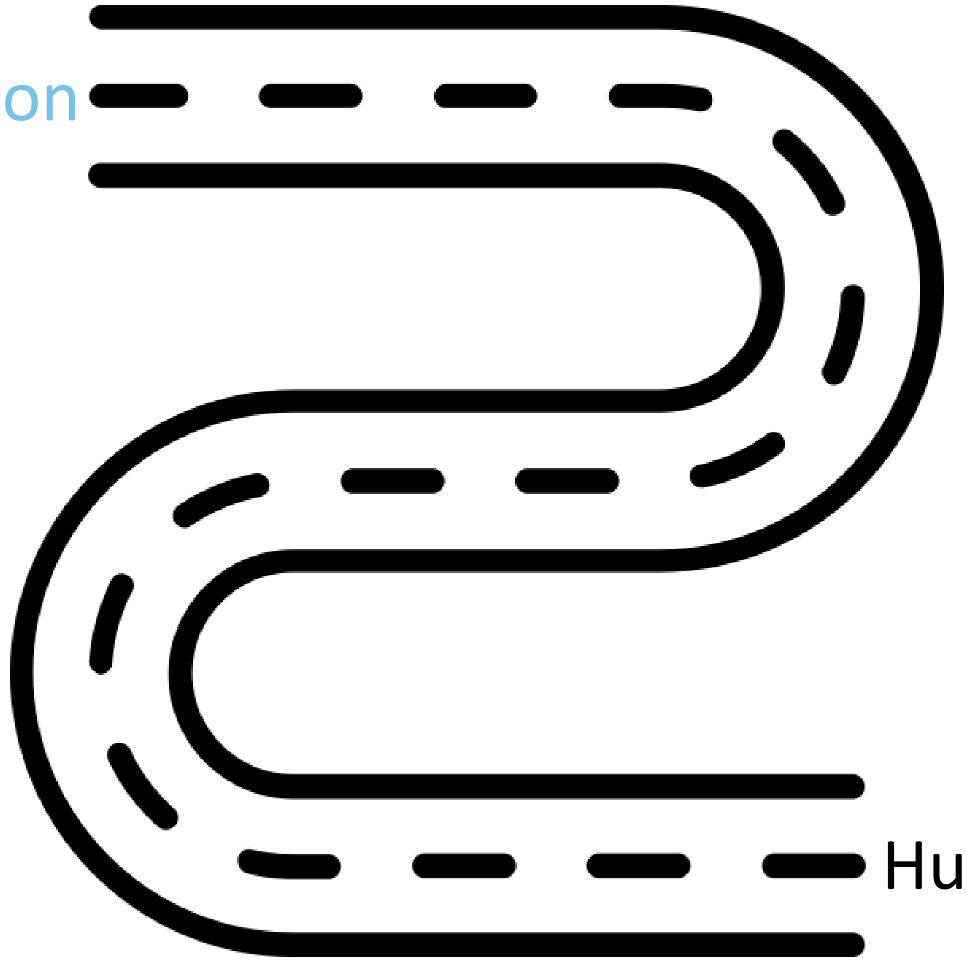
Emergence of Autonomy in Planes



Dominique Chatrenet, *Air Transport Safety Technology & Training,* ETP 2010.

Today’s Roadmap

Introduction



Levels of Autonomy and ADAS

Value Sensitive Design and Empirical Evaluation

Human Metrics

SAE Levels for Autonomy

Level 0	Provide warnings with no interventions	Driving
Level 1	Partial or shared control between the driver and the vehicle	Hands on
Level 2	Performs all driving functions, but may require quick intervention from the driver	Hands off
Level 3	Performs all driving functions, but may call on the driver for assistance with timed handoffs	Eyes off
Level 4	Fully autonomous, but restricted to certain areas or conditions with prescribed handoffs	Mind off
Level 5	Fully autonomous with no restrictions	Steering wheel optional

https://www.sae.org/standards/content/j3016_201806/

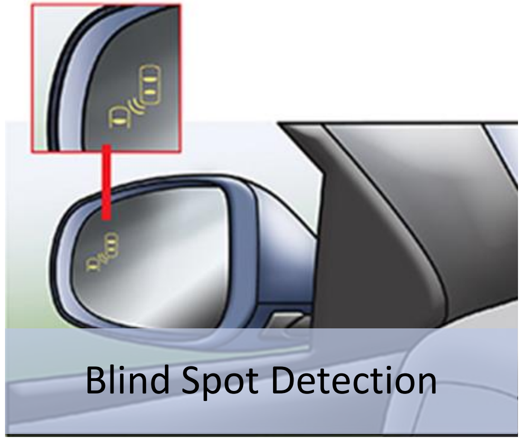
Advanced Driver Assistance Systems

Forward Collision Mitigation

Warning: ↓27%
Autobrake: ↓50%
Rear-end rate: ↑20%*

Lane Keeping Assistance

Warning: ↓11%



Blind Spot Detection

Warning: ↓14%

Rearview Assistance

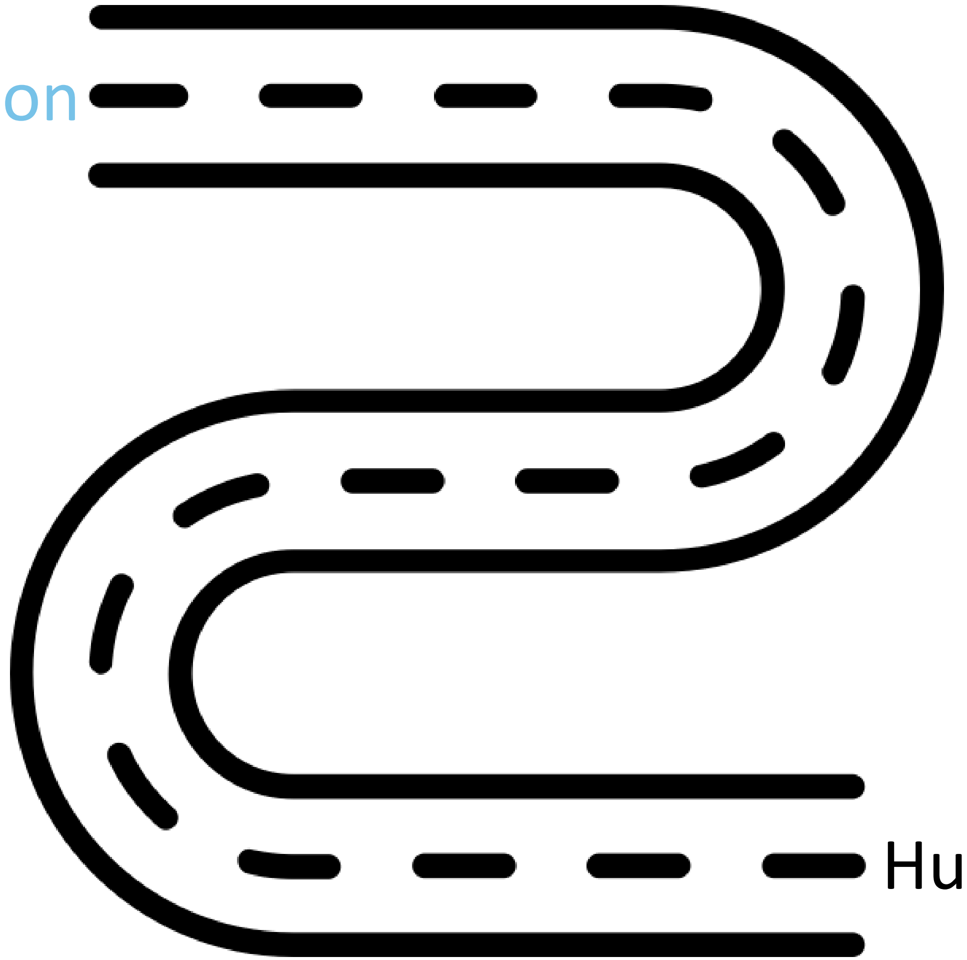
Camera: ↓17%
Warning: ↓22%
Autobrake: ↓78%

Image Credit: Google Image Search

https://www.iihs.org/topics/advanced-driver-assistance

Today’s Roadmap

Introduction



Levels of Autonomy and ADAS

Value Sensitive Design and Empirical Evaluation

Human Metrics

Value Sensitive Design

- How do designers connect human values to engineering specification?
  - Explicit encoding of safety and legal constraints
  - Ex: safety and efficiency / comfort → how to handle conflicting values?

Value Sensitive Design: Ethical Considerations

- Philosophy formally characterizes ethical entanglements
- Deontology: follow set of rules with no exception
- Consequentialism (utilitarianism): cost-based framework that evaluates choices based on consequences (considers outcome of chosen action)
- How to incorporate into engineering design?
  - Actions bounded by deontological constraints
  - Cost functions realized via consequentialism

$$
\left.\begin{array}{ll}
\underset{\textit{action}}{\text{minimize}} & \textit{path deviation} \\
 & \textit{steering rate} \\
 & \textit{traffic law violation*}
\end{array}\right\} \quad \text{consequential costs}
$$

$$
\left.\begin{array}{ll}
\text{subject to} & \textit{actuator limits} \\
 & \textit{obstacle avoidance} \\
 & \textit{traffic law violation*}
\end{array}\right\} \quad \text{deontological constraints}
$$

S. Thornton, et al. "Incorporating ethical considerations into automated vehicle control." *IEEE Transactions on Intelligent Transportation Systems*, 2016.

Value Sensitive Design

- How do designers connect human values to engineering specification?
  - Explicit encoding of safety and legal constraints
  - Ex: safety and efficiency / comfort → how to handle conflicting values?
- VSD iterates over conceptual, technical, and empirical investigations by connecting with stakeholders
  - **Conceptualization:** Who are the direct and indirect stakeholders? What are the human values?
  - **Technical Implementation:** What are the assumptions? What are the inputs and goals? What is the formal approach? How does it encode human values?
  - **Empirical Analysis:** How will you design the experiment? What is the formal hypothesis? What is the baseline?
  - **Iterate:** Creates documentation and justification for design decisions

Case Study: Pedestrian Collision Avoidance

Direct and Indirect Stakeholders
- Vehicle Occupants
- Traffic Authorities
- Pedestrians
- Surrounding vehicles
- Bystanders
- Community and society

Human Values
- Safety, Care / Respect for others → safety
- Legality, Respect for authority → legality
- Mobility, Autonomy → efficiency
- Mobility, Trust and transparency → smoothness
- Fairness and reciprocity → stakeholder

Technical Considerations
- Safety constraints: Requires distance, velocity, detection
- Legality: Vehicle codes
- Time efficiency: Requires vehicle speed
- Smoothness: Requires acceleration

Problem Statement
Given: Distance to crosswalk
Vehicle speed
Detection of pedestrian crossing
Control input: Longitudinal acceleration
Goal: Drive safely, efficiently, and smoothly

S. Thornton, et al. "Value sensitive design for autonomous vehicle motion planning." *IEEE Intelligent Vehicles Symposium (IV)*, 2018.

Empirical Evaluation: Experimental Design

Experimental Components

Conditions (Treatments)

- Drug vs placebo
- Random vs optimal
- RRT* vs RL

Measures (Responses)

- Symptom progression
- Comfort
- Cost

Experimental Units (Subjects)

- Patients
- Users
- *Domains / Problems*

Assignment Method (Allocation)

- Random
- Comparative (subjects “see” both)

Conditions, Measures, and Hypotheses

- **Independent Variables:** conditions
  - What are you manipulating? (e.g., drug type, controller type, detection method)
  - If there are *n* possible values (levels) for one IV, then we run *n* experiment settings
  - If there are two IVs with *n* and *m* levels, we run *nm* experiment settings
- **Dependent Variables:** measured outcomes
  - What are you measuring? (e.g., symptoms, trust, accuracy)
  - Check for construct validity (e.g., does IQ measure intelligence? Do user ratings of predictability indicate predictability?)
- **Hypothesis:** Independent variable *x* (positively) affects dependent variable *y*
- **Type of Experiment**
  - Observational, prevalence (cross-sectional), and controlled
  - *Exploratory Data Analysis:* What the data can tell us beyond the formal modeling or hypothesis testing?

Case Study: Kidney Stone Treatments

- A study was performed to compare the effectiveness of two classes of treatments for kidney stones

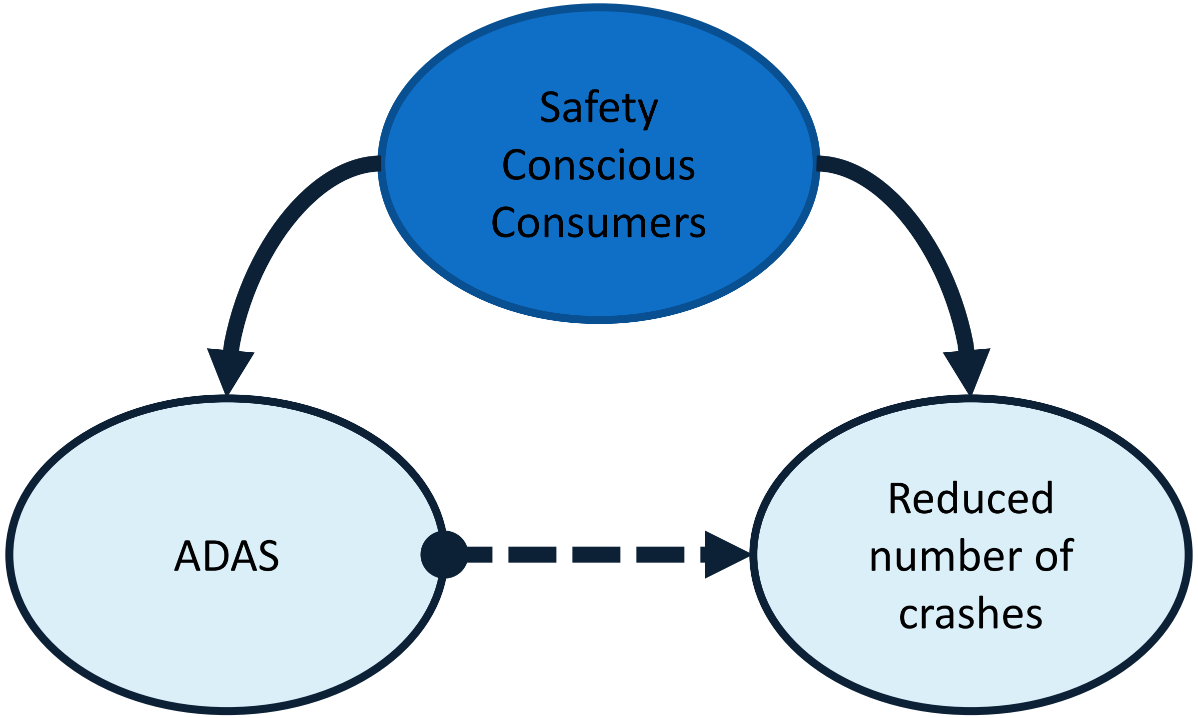
Stone Size	Open Surgery	Percutaneous Nephrolithotomy (PN)
< 2cm	93% (81/87)	87% (234/270)
>= 2cm	73% (192/263)	69% (55/80)
Overall	78% (273/350)	83% (289/350)

- PN has a higher success rate overall, but open surgery has a higher success rate in both group
- The assumption is that the decision for the treatment is independent on the size of the stone – which is incorrect!

→ **Simpson's paradox // "correlation does not equal causation"**

http://courses.ieor.berkeley.edu/ieor165/lecture_notes/ieor165_lec15.pdf

Causality and Confounds



Confounds: Variable whose effect cannot be distinguished from the effect of the tested conditions (independent variable).

Tools to eliminate confounds?

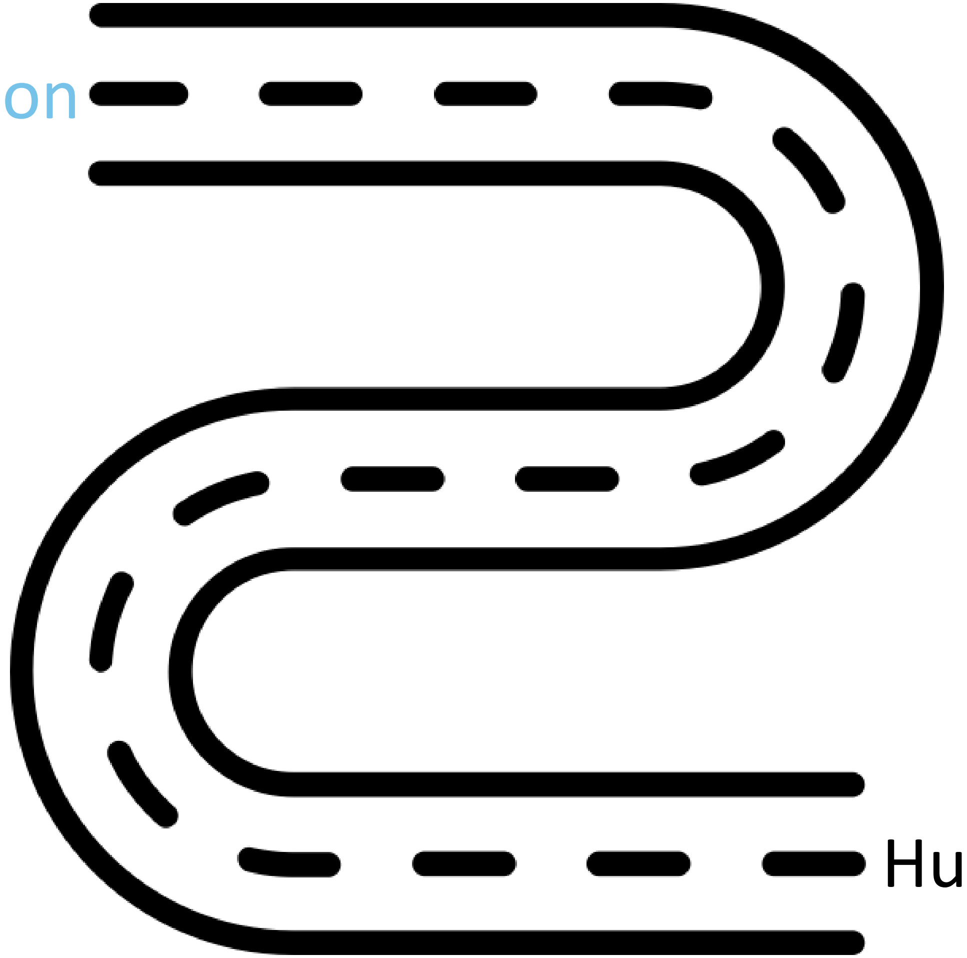
- Randomization
- Within subjects / counter-balanced
- Block strategies
- Be the devil's advocate
    - Optimize the baseline
    - Use measures that do due justice to the baseline

Value Sensitive Design

- How do designers connect human values to engineering specification?
  - Explicit encoding of safety and legal constraints
  - Ex: safety and efficiency / comfort → how to handle conflicting values?
- VSD iterates over conceptual, technical, and empirical investigations by connecting with stakeholders
  - **Conceptualization:** Who are the direct and indirect stakeholders? What are the human values?
  - **Technical Implementation:** What are the assumptions? What are the inputs and goals? What is the formal approach? How does it encode human values?
  - **Empirical Analysis:** How will you design the experiment? What is the formal hypothesis? What is the baseline?
  - **Iterate:** Creates documentation and justification for design decisions

Today's Roadmap

Introduction



Levels of Autonomy and ADAS

Value Sensitive Design and Empirical Evaluation

Human Metrics

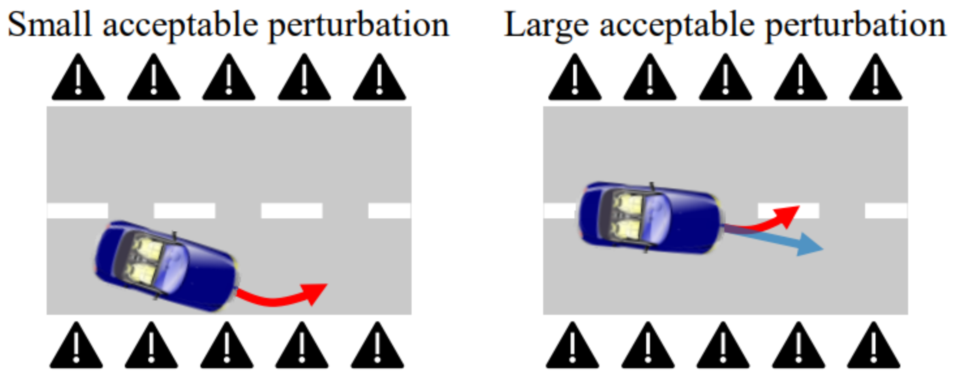
Common Metrics for HRI: Biasing Effects

- Communication: delay, jitter, bandwidth
- Robot Response: timing factors, reliability
- User / Human Factors:
  - Performance shaping factors
    - Operational: tactics, time constraints
    - Equipment: workspace layout
    - Task: complexity, repetitiveness
    - Personnel: training, motivation, stress
    - External: visibility
  - Interface (design factors, clutter, input mechanism)
  - Role (supervisor, operator, mechanic, peer, bystander)

A. Steinfeld, et al., “Common Metrics for Human-Robot Interaction.” ACM HRI, 2006.

Common Metrics for HRI: (Shared) Metrics

- System Performance
  - Quantitative (effectiveness, efficiency)
  - Subjective Ratings (incorporate stakeholders for feedback)
  - Utilization of Human-Robot teaming (requests for assistanc
- Operator Performance
  - Situational Awareness → [tool] Situational Awareness Glob Technique (SAGAT)
  - Workload → [tool] NASA-Task Load IndeX (NASA-TLX)
  - Accuracy of mental models (conceptual, movement, spatial, and modality compatibility)
- Robot Performance
  - Self-awareness (assessment, monitoring, understanding of intrinsic limitations)
  - Human Awareness (ability to model and predict, user-oriented perception, adaptation)
  - Autonomy (functionality when encountering anomalies, no human help (neglect tolerance))



Increased technology doesn't increase situational awareness.

A. Steinfeld, et al., "Common Metrics for Human-Robot Interaction." ACM HRI, 2006.

Sample Trust Subjective Measures

The following statements were used to assess general propensity to trust:

1. I am suspicious of the [system name]'s intent, action, or outputs.
2. The [system name]'s actions will have a harmful or injurious outcome.
3. The [system name] is dependable.
4. The [system name] is reliable.
5. I can trust the [system name].

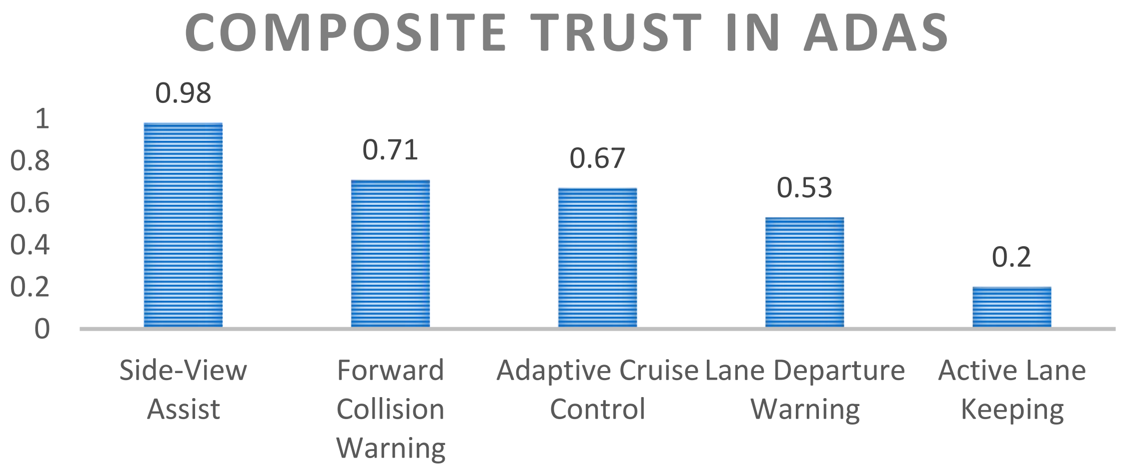
Questions from Muir's Questionnaire include:

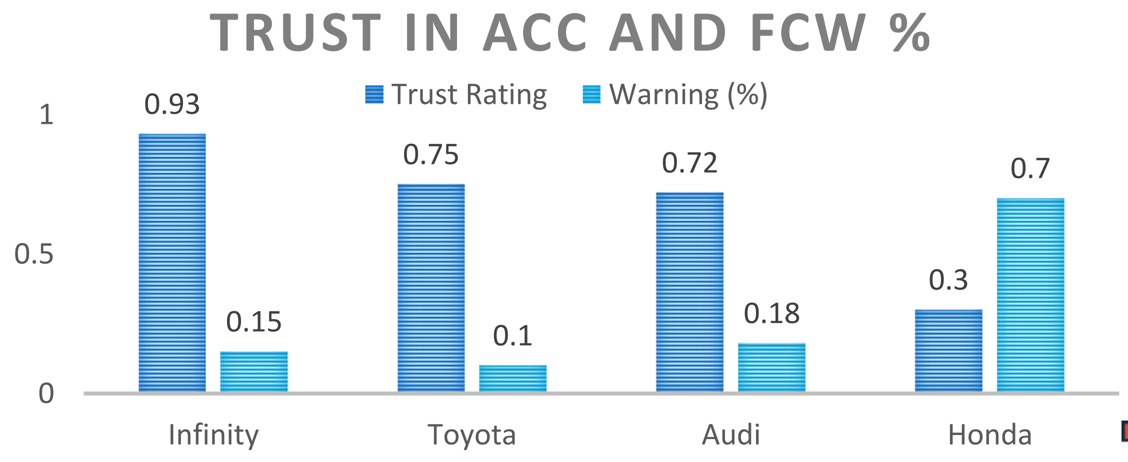
1. To what extent can the [system name] behavior be predicted?
2. To what extent can you count on the [system name] to do its job?
3. What degree of faith do you have that the robot will be able to cope with similar situations in the future?
4. Overall, how much do you trust the [system name]?

D. Kidd, et al., "Driver trust in five driver assistance technologies following real-world use in four production vehicles." Traffic Injury Prevention (TIP), 2017.
B. Muir. "Operators' trust in and use of automatic controllers in a supervisory process control task." 2002.

Case Study: Trust in ADAS

- Participants drove a 2016 Toyota Prius, 2016 Honda Civic, 2017 Audi Q7, or 2016 Infiniti QX60 for several weeks
  - Equipped with forward collision warning, adaptive cruise control, active lane keeping, side-view assist, and lane departure warning
- Subjects reported mileage and warnings daily, and trust survey post-use
  - Responses were averaged to create a composite measure of trust ranging from -2 (strongly disagree) to +2 (strongly agree)





D. Kidd, et al., “Driver trust in five driver assistance technologies following real-world use in four production vehicles.” Traffic Injury Prevention (TIP), 2017.

What does "trust" mean?

Trust is a willingness to take risk and to be vulnerable to another party.

One proxy for assessing trust is to gauge
whether users rely on their active safety systems.

D. Rousseau, et al. "Not so different after all: A cross-discipline view of trust." Academy of Management Review, 1998.

Is ADAS used?

Observations at dealerships of seven automakers found:

- 93% of forward collision systems were activated
- 99% of the blind spot detection and alert systems were activated
- 90% of driver monitoring alerts were activated
- 52% of lane departure warning and lane-keep assists were activated

Which systems do you think people *trusted*?

I. Reagan, et al. "Crash avoidance and driver assistance technologies–Are they used?" *Transportation research part F: Traffic Psychology and Behaviour*, 2018.

Is ADAS used?

Data shows that 40% and 27% of Volvo and Toyota owners, respectively, have experienced unnecessary alerts. 17% and 25% of these owners reported missed warnings.

Recall that lane departure systems are much less activated by drivers. Users note that lane departure alerts are "more annoying" than forward collision alerts. Why?

Is it the alert itself?
Mode of communication and interaction is a common study in Human Factors and Human-Computer Interaction
→ generally vibration warnings are often "less annoying" than audio

Is it a functional aspect of the system?
1. Drivers who do not use a signal deem warnings as false alarms
2. Intentional deviations may be required, leads to needless alerts
3. Cameras may miss-detect lanes, leading to false alarms or missed warnings

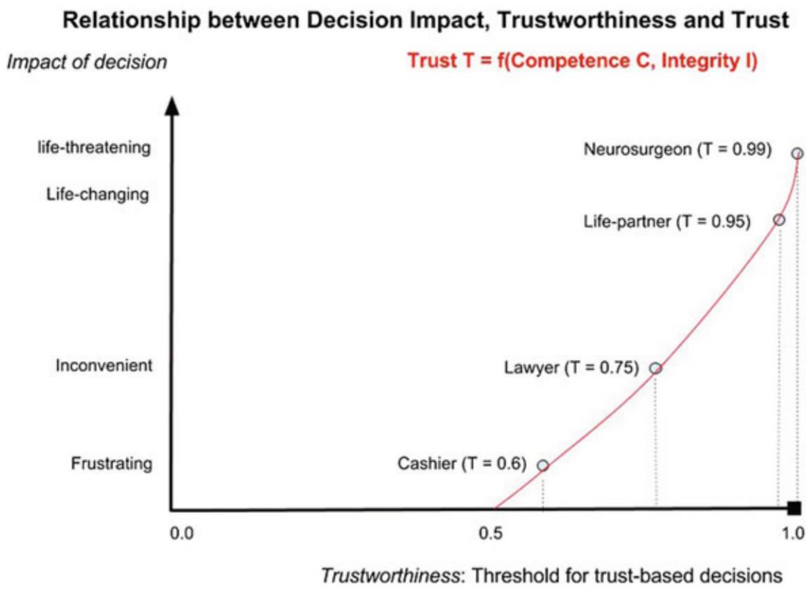
→ False or unnecessary alerts decrease user acceptance

Perceived **usefulness** and **ease of use** are factors that largely determine user acceptance and trust

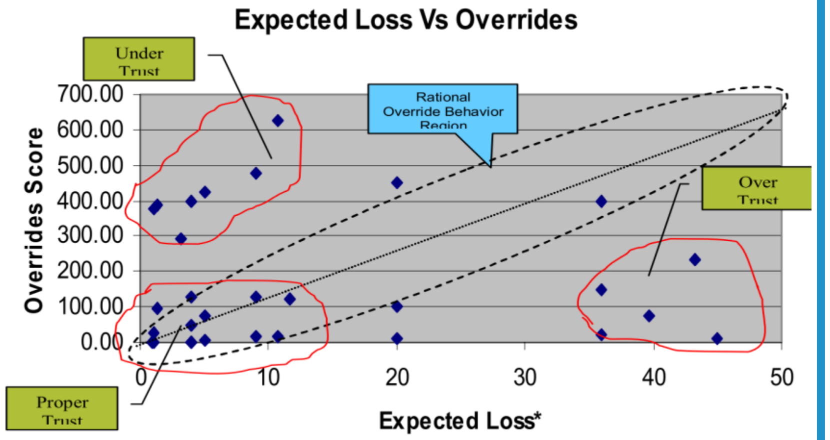
I. Reagan, et al. "Crash avoidance and driver assistance technologies–Are they used?" *Transportation research part F: Traffic Psychology and Behaviour*, 2018.

Many Facets of and Considerations for Trust

Context and Impact



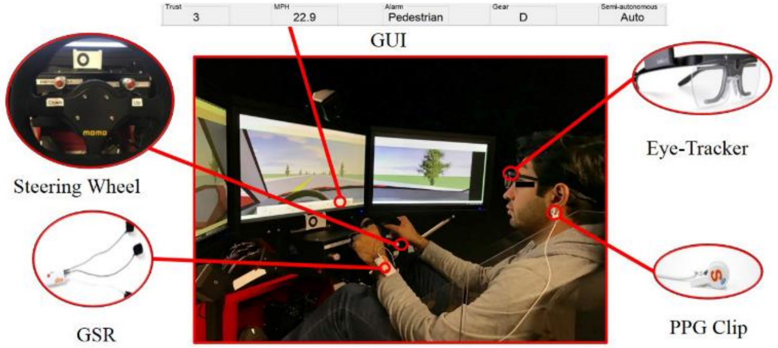
Under- and Over- Trust



Expected Loss:

$$EL = \sum_{i=1}^{n} \frac{Expected\ Loss\ at\ trial}{Number\ of\ Overrides}$$

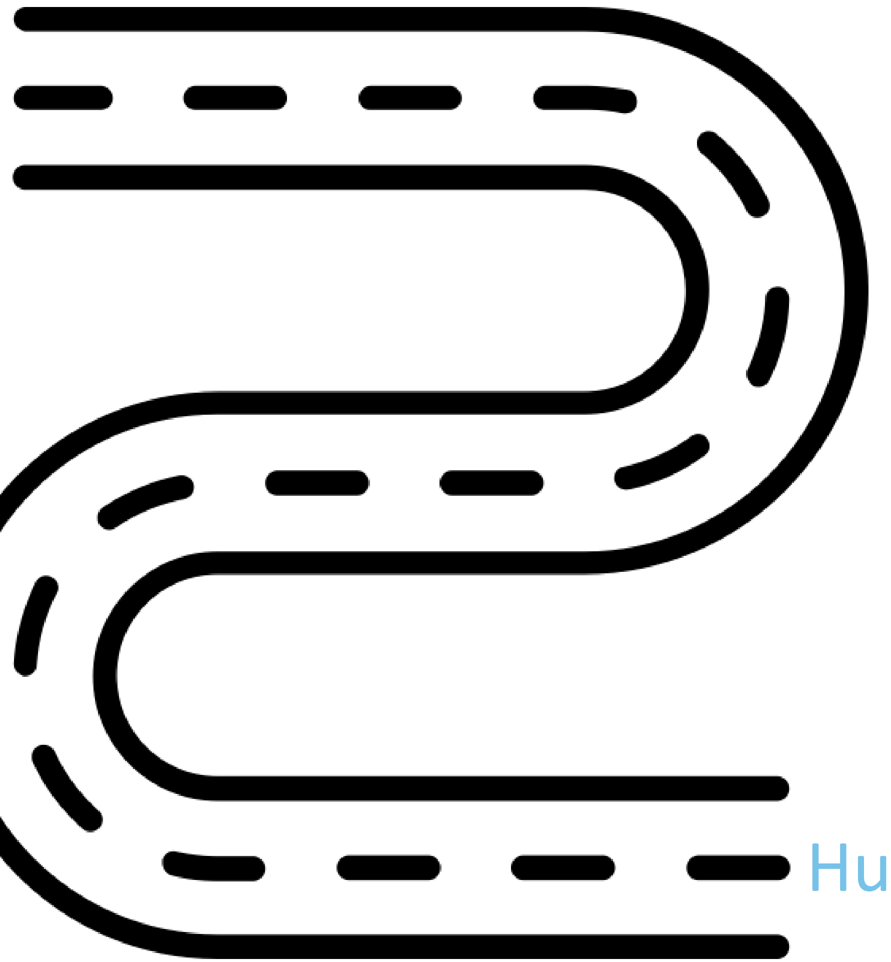
Physiological Response



H. A. Abbass, et al., “Foundations of trusted autonomy.” Springer Nature, 2018.
A. Freedy, A., et al. “Measurement of trust in human-robot collaboration.” IEEE International Symposium on Collaborative Technologies and Systems, 2007.
S. Sheng, et al. “A case study of trust on autonomous driving.” ITSC, 2019.

Today's Roadmap

Introduction



Levels of Autonomy and ADAS

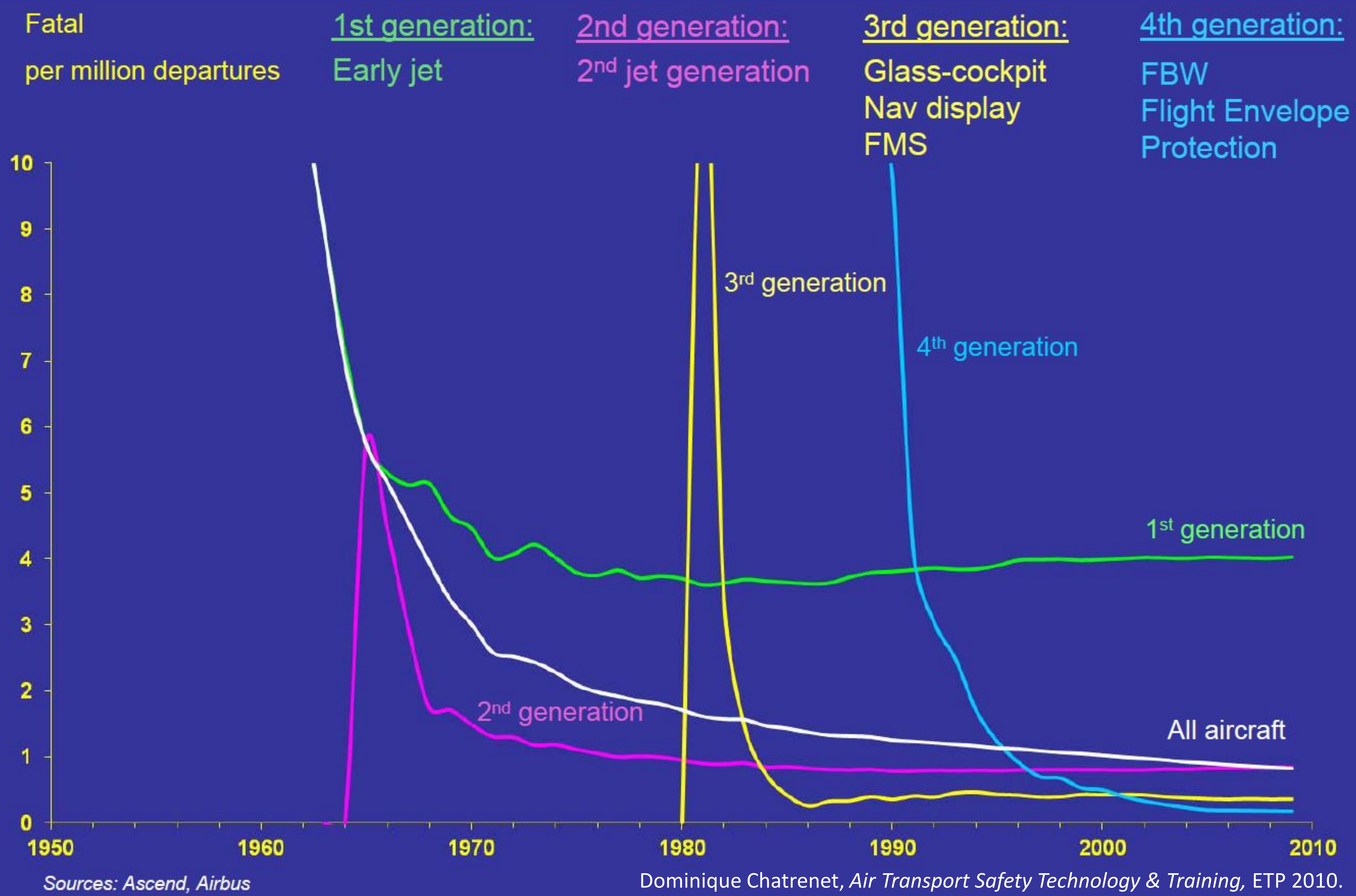
Value Sensitive Design and Empirical Evaluation

Human Metrics

References and Other Resources

- **Survey on ADAS:** K. Bengler, et al. "Three decades of driver assistance systems: Review and future perspectives." IEEE Intelligent Transportation Systems Magazine 6.4, 2014.
- **Value Sensitive Design:**
  - B. Friedman, P. Kahn, and A. Borning, "Value Sensitive Design and Information Systems," in Human-Computer Interaction and Management Information Systems, 2006.
  - B. Friedman, P. Kahn, and A. Borning. "Value sensitive design: Theory and methods." University of Washington Technical Report 2-12, 2002.
- **Introduction to Human Factors:** [textbook] Handbook of human factors and ergonomics by G. Salvendy
- **Causality:** [textbook] Causality: Models, Reasoning, and Inference by J. Pearl
- **Models and Techniques for Assessing Trust:**
  - **Automation Acceptance Model**: M. Ghazizadeh, J.D. Lee, and L. Ng Boyle. "Extending the Technology Acceptance Model to assess automation." Cognition, Technology & Work, 2012.
  - **Calibrating Trust and Appropriate Reliance:** J.D. Lee and K.A. See. "Trust in automation: Designing for appropriate reliance." Human factors 46.1, 2004.

Fatal rate by year - valid end 2009



Dominique Chatrenet, *Air Transport Safety Technology & Training,* ETP 2010.

Case Study: Antilock Braking

- Antilock brakes allow for shorter stopping distances and the ability to steer / maintain control while hard braking, especially on wet and slippery surfaces
- Owners of late model cars equipped with antilock brakes were surveyed in North Carolina and Wisconsin regarding their experiences with antilocks
- The survey results indicated that more than 50% of the drivers in North Carolina and 40% in Wisconsin incorrectly indicated how to brake a car in an emergency situation on wet and slippery pavements in a way that will effectively activate the antilock feature
- More drivers in Wisconsin than in North Carolina reported that their cars' antilock feature had been used, but more than 33% of the Wisconsin drivers and 62% of North Carolina drivers said they had never used the antilock feature of their cars' brakes
- Conclusion: people did not understand the automation!

https://www.iihs.org/topics/advanced-driver-assistance