

Assuring Individual, Social, and Cultural Embeddedness of Autonomous Cyber- Physical Systems (ISCE-ACPS)

Werner Damm

Carl von Ossietzky Universität Oldenburg

Alexander Pretschner

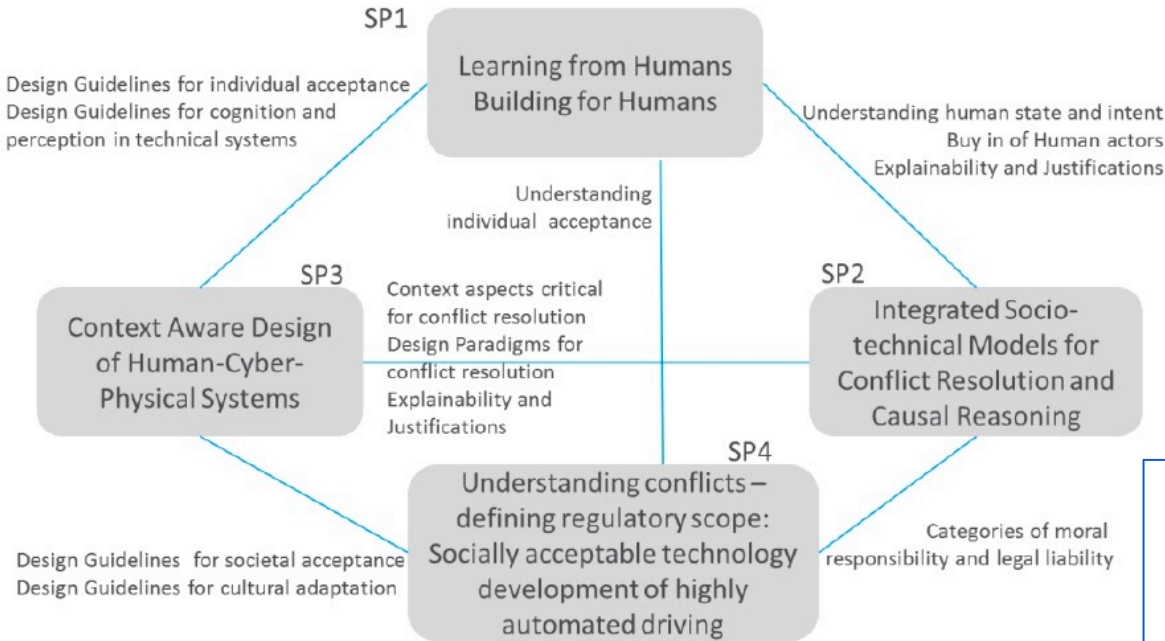
Technische Universität München

Coordinators DFG Package Project PAK 987/1
for 2nd phase DFG co-financing of NSF PIRE Project 1743772

Structure of Presentation

Structure and results of ISCE-ACPS
Strategic Impact of ISCE-ACPS – Research
Strategic Impact of ISCE-ACPS - Teaching
Key Findings

Structure and Participating Organizations



Name	Affiliation	Denomination
Prof. Dr. Klaus Bengler	Technische Universität München	Ergonomie
Prof. Dr. Werner Damm	Carl von Ossietzky Universität Oldenburg	Sicherheitskritische eingebettete Systeme
Prof. Dr. Martin Fränzle	Carl von Ossietzky Universität Oldenburg	Hybride Systeme
Priv. Doz. Dr. Meike Jipp	DLR Institut für Transportsysteme	Abteilungsleiterin Human Factors
Prof. Dr. Frank Köster	DLR Institut für Transportsysteme und Carl von Ossietzky Universität Oldenburg	Bereichsleiter Automotive DLR Professur Intelligente Transportsysteme
Prof. Dr. Sebastian Lehnhoff	Carl von Ossietzky Universität Oldenburg	Energieinformatik
Dr. Andreas Lüdtkke	OFFIS	Senior Principle Scientist and Group Leader Human Centred Design
Prof. Dr. Alexander Pretschner	Technische Universität München	Software- and Systems-Engineering
Prof. Dr. Jochem Rieger	Carl von Ossietzky Universität Oldenburg	Angewandte Neurokognitive Psychologie
Prof. Dr. Mark Schweda	Carl von Ossietzky Universität Oldenburg	Ethik in der Medizin
Prof. Dr. Jürgen Taeger	Carl von Ossietzky Universität Oldenburg	Bürgerliches Recht, Handels- und Wirtschaftsrecht sowie Rechtsinformatik
Prof. Dr. Markus Tepe	Carl von Ossietzky Universität Oldenburg	Politisches System Deutschlands

**Assuring
Individual, Social, and Cultural
Embeddedness of
Autonomous Cyber-Physical Systems**

Selected Results

Mixed Traffic

Demonstrated complete flow from in-the-wild [fNIRS based and video camera based measurements](#) to infer frustration level to building [cognitive models of frustration-levels](#) of drivers to using car2X communication to [guide strategy selection of autonomous cars](#) possibly risking to collide with “standard” driver controlled ego-vehicle

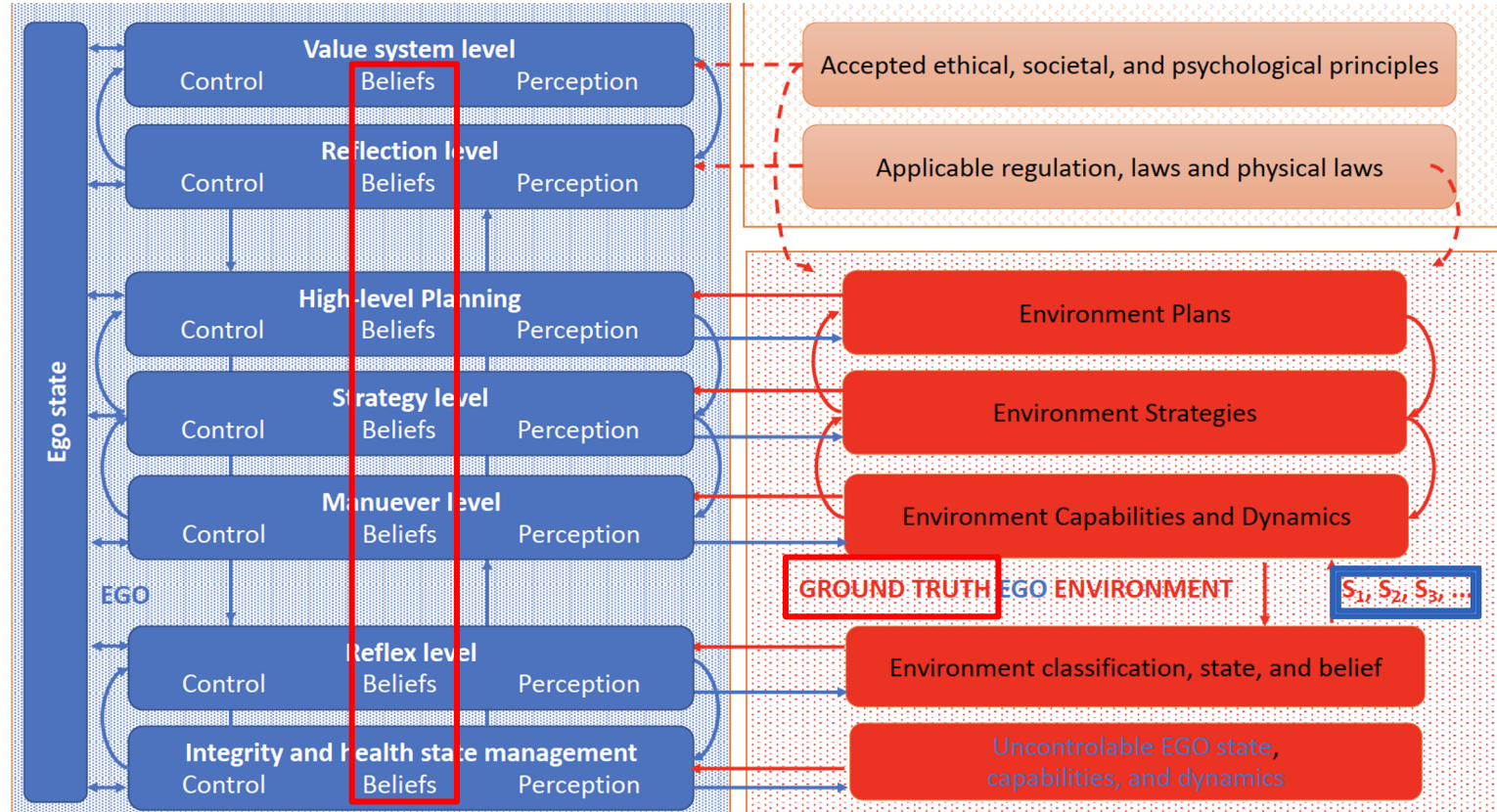
Causal Analysis

Demonstrated complete flow for [causal analysis for highly autonomous driver assistance](#) systems in cooperation with Vanderbilt

Energy

Demonstrated [adaptability](#) of virtual expert for detecting and mitigating cyber-attacks on smart grids [to current level of situational awareness of control-room operators](#)

Integrated Socio Technical Models for Safety Critical Systems



Ego-System: Human or CPS or system of HCPS

Environment of Ego-System

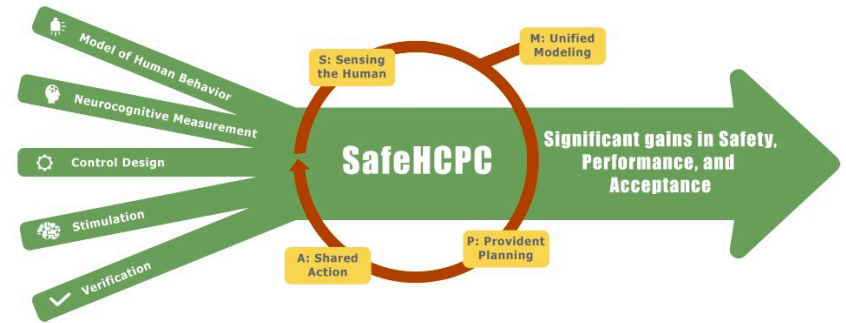
The meta-model is capable of ...

- Capturing all classes of system-level **interactions** between **CPS** and **professional or semi-professional humans**
- Capturing all relevant aspects of **environments** of such systems;
- Capturing, expressing, assessing, and adapting **beliefs about its environment** and their confidence levels;
- **Reasoning about the discrepancies between beliefs and ground-truth**;
- Capturing strategies for **self-adaption** and **self-evolution** in dynamically changing contexts;
- Supporting **ex-ante, on-line, and ex-post analysis** of all incidents/states potentially **impacting safety, trust, compliance to ethical and/or societal principles or regulations**;
- Supporting **seamless cooperation** of **mixed teams of humans and CPS** to reach shared objectives within given time-frames on multiple levels of interaction;
- Supporting **analysis of realizability** of system objectives and **synthesis of strategies** to achieve such objectives, possibly involving coalition partners;
- Supporting the **assessment of trustworthiness** of Human-Cyber Physical Systems;
- Supporting **justifications** for actions chosen by CPS subsystems;
- Serving as a **blue-print** for building such systems in multiple application domains through specialization.

Strategic Impact of ISCE-ACPS

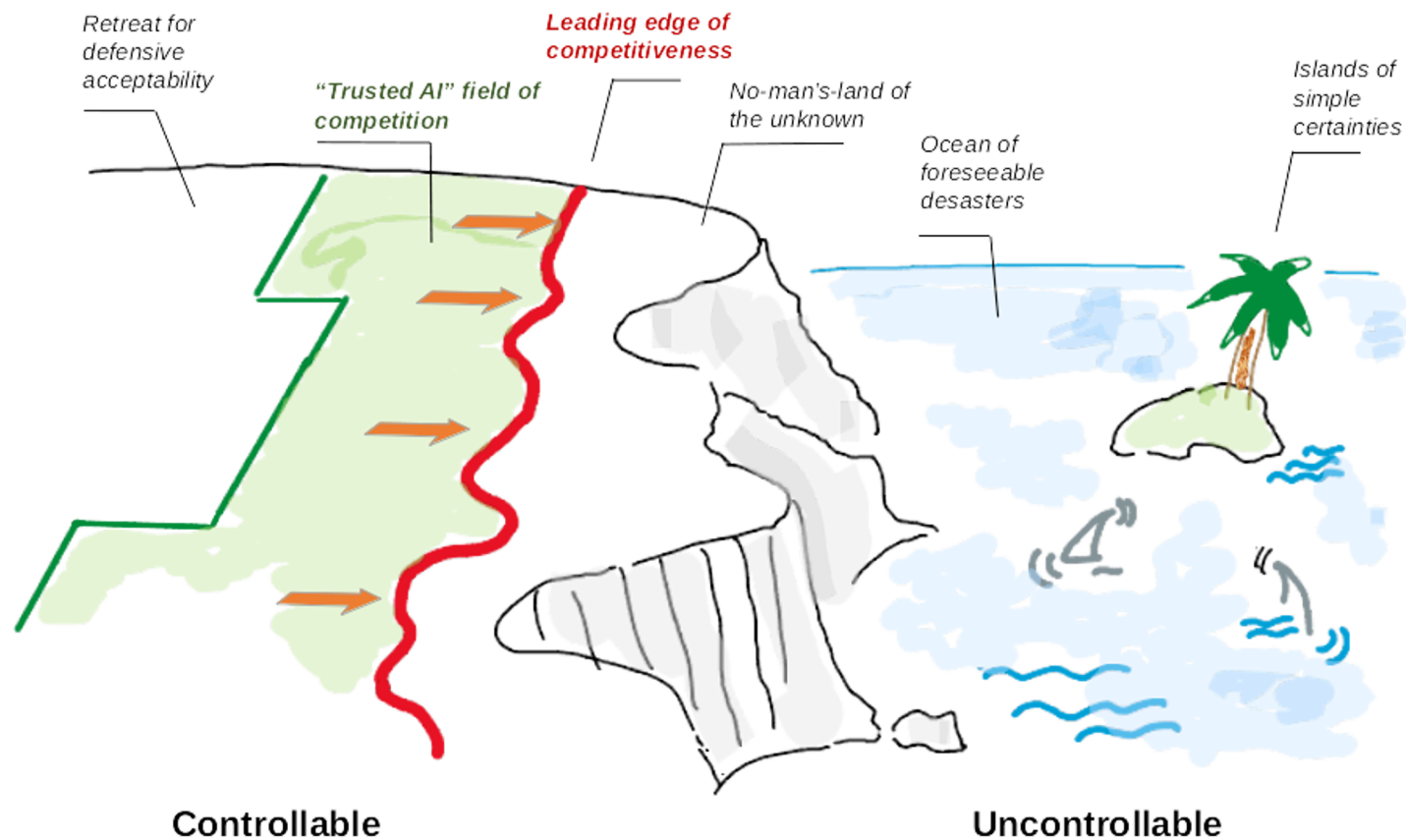
Safe Human Machine Cooperation
The Trusted AI Initiative
Ethical Software Engineering

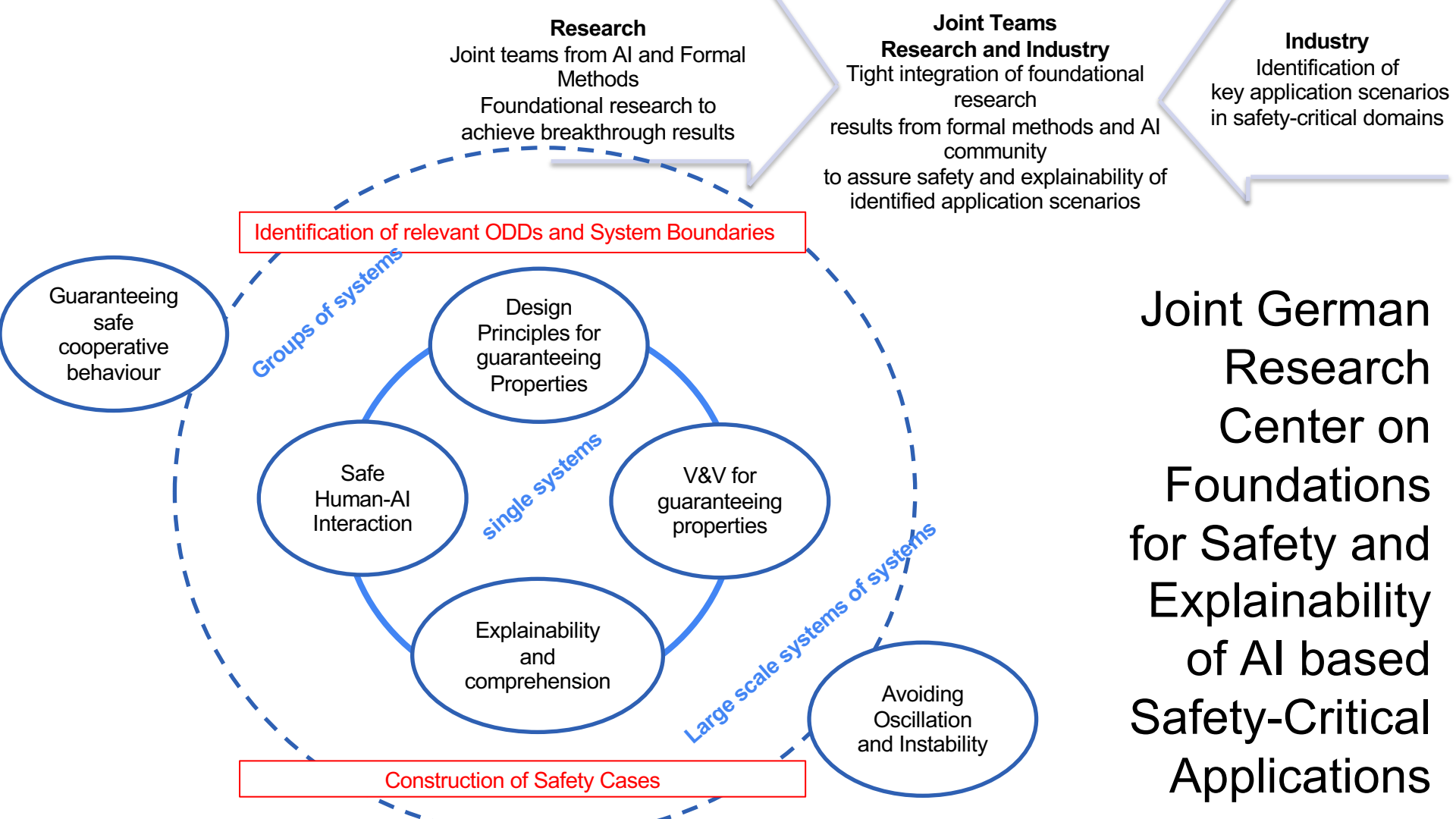
Vision SafeHCPS



- Human-cyber-physical systems which immediately and sensibly react to – and vice versa influence – human assessment of the situation and impending human behavior, even if these have not been expressed explicitly.
- Components of such HCPSes maintain a domain-appropriate ‘mental’ image of the states and plans of involved humans that is of comparable quality to a human’s mental image of the state and future behavior of the technology and the fellow humans she interacts with.
- It exploits in its domain of operation that image for comfortable, predictable, and provably safe human-cyber-physical cooperation, thus being able to bridge the dividing line between humans and machines in sensing the environment, in behavioral planning and control, and in shared action.

The Trusted AI initiative



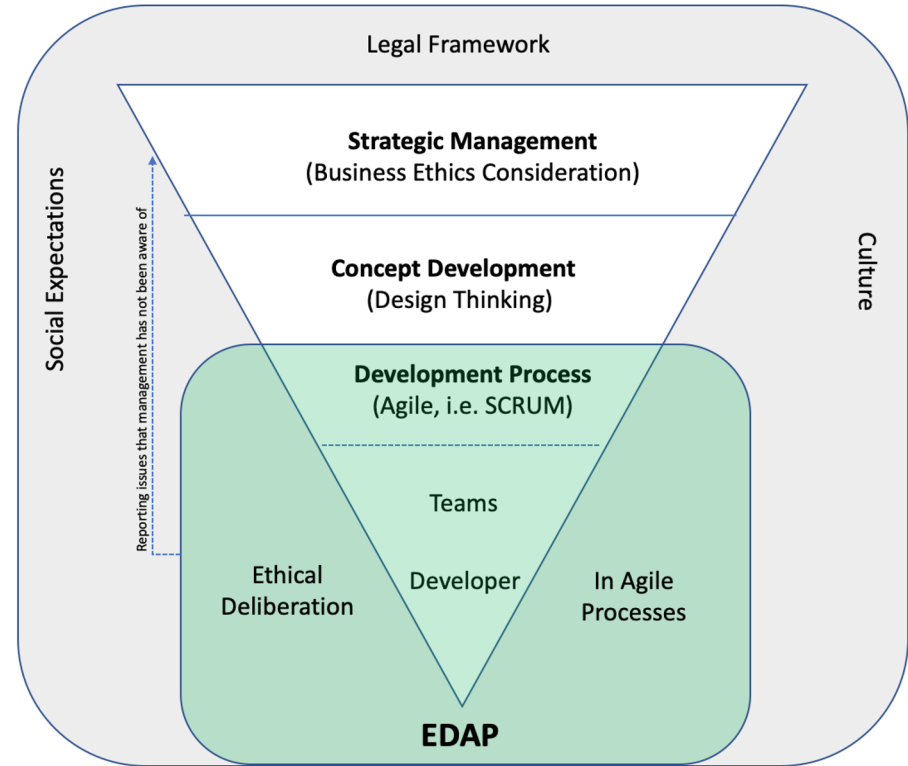
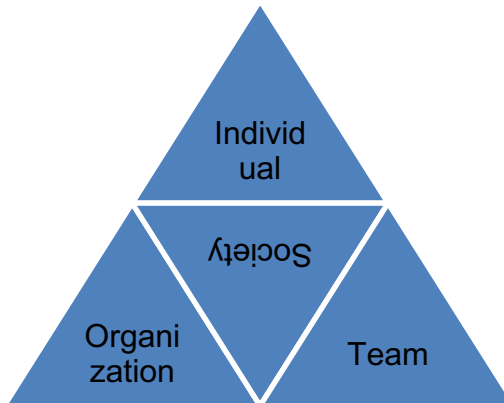


Ethical Software Engineering

Research Question: How do we build software systems in an ethical way?

Collaboration with David Hess

A method to embed ethical deliberation into agile development processes.



Codes of Conducts and Trade-Offs

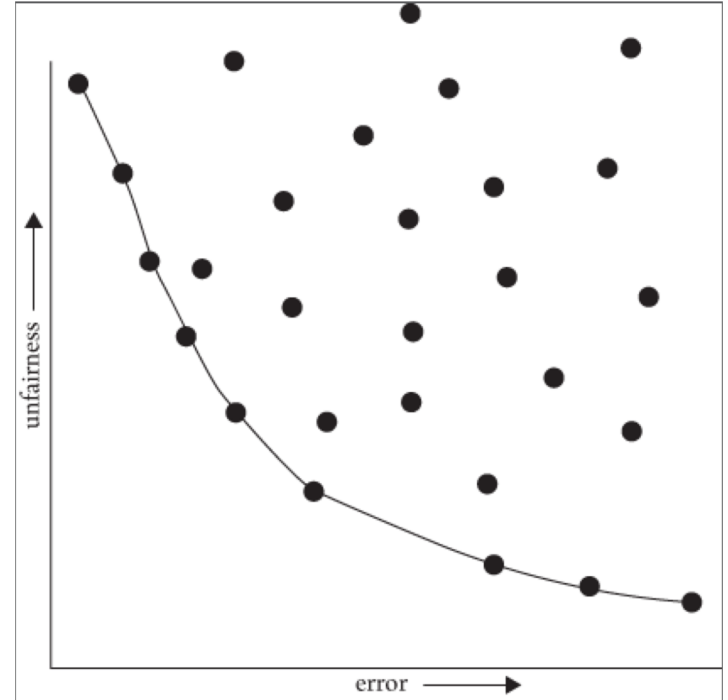
CoCs do not tell us what to do when values interfere with each other, like privacy vs. transparency, freedom vs. safety, etc.

When we try to optimise a model/software that relies on two (or more) values a good shot is Pareto optimality

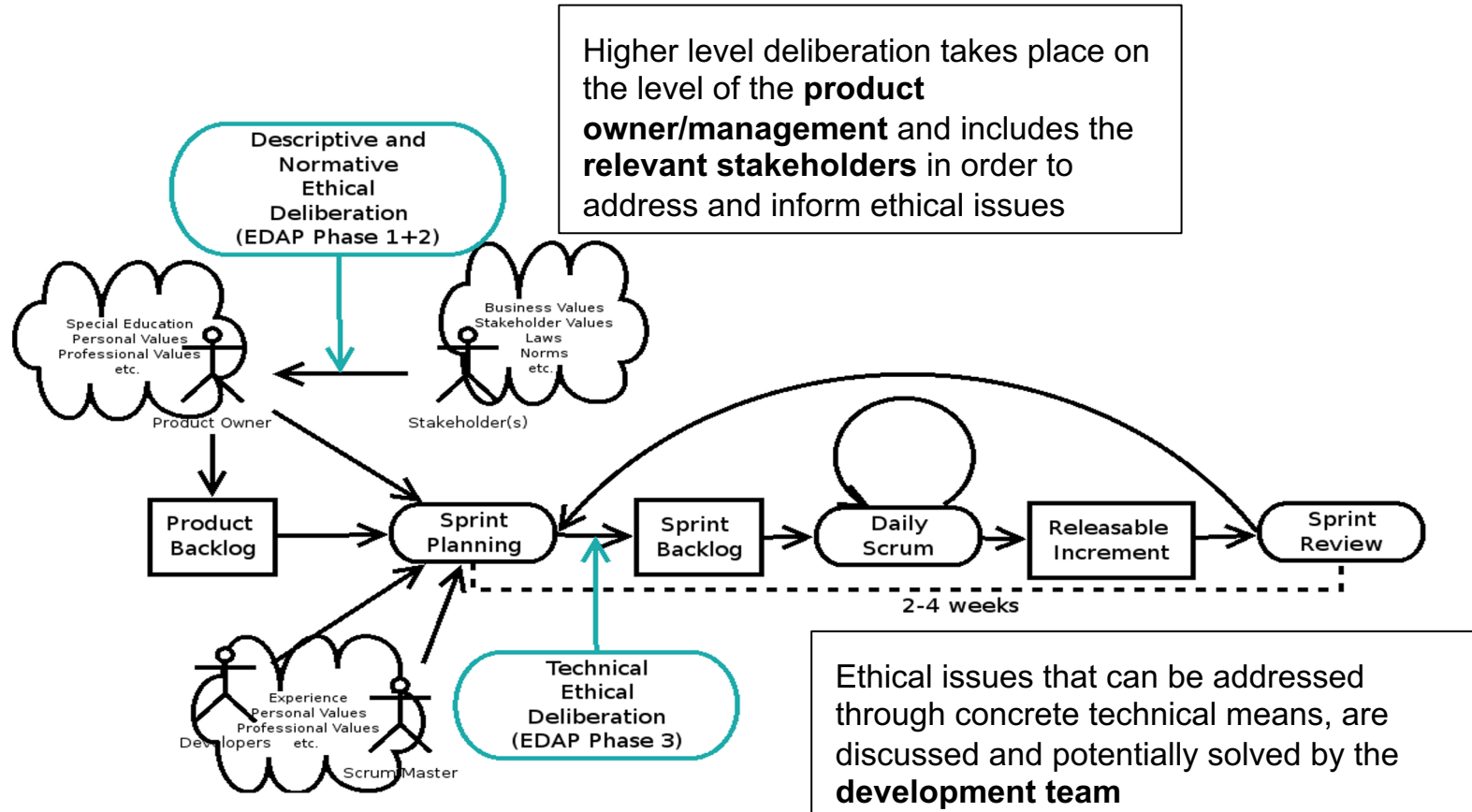
Only solutions on the frontier are reasonable

But ... **where on the curve?**

(and Pareto optimality is problematic as such)



Extending Agile Framework with Ethical Deliberation



Impact on Teaching

Oldenburg

- **Engineering of Socio-Technical Systems - Master's Programme**

This English-language Master's degree programme takes an interdisciplinary approach to the development of safety-critical, computer-based interactive systems, and particularly focusses on *the interaction between humans and technology*. Courses combine content from neuroscience with methods used by engineers to develop information systems.

- **Research Training Group SEAS - "Social Embeddedness of Autonomous Cyber Physical Systems" – PHD program**

SEAS analyses the social embeddedness of ACPS and the potential effect of self-explaining ACPS from an interdisciplinary perspective, focusing on the application domains *Mobility (autonomous cars), Energy (Smart Grid) and Health Service Research (highly automated Patient Care)*. We distinguish three analytical dimensions, which we label *Acceptance, Cooperation and Governance of ACPS*. Ph.D projects are supervised by two professors of complementary disciplines each, and focus on how to realize the opportunities promised by these new technologies in a way that is compatible with the normative and ethical principles of liberal civil societies.

Munich

- **bidt Colloquium "Digital Transformation in Society"**

The mission of the interdisciplinary Bavarian Research Institute for Digital Transformation, founded by one of the PIRE co-PIs and now at >30FTEs and ~20 projects financed across Bavaria, is to study and help design digital transformation in society. As one activity, we organize a monthly colloquium for the >50 PhD students across Bavaria whose research is funded by the institute. Held by distinguished international speakers, the colloquium is meant to foster the dialogue across research areas and across disciplines.

- **TUM lecture series "Digital Transformation"**

The TUM department of informatics, represented by one of the PIRE co-PIs has organized an interdisciplinary lecture series on digital transformation since 2017. This international lecture series, open to the general public, is heavily inspired by discussions taking place in the PIRE consortium. By approaching different facets of digital transformation from the perspectives of individual disciplines, we aim at providing a transdisciplinary and holistic perspective at the interconnections between digital transformation, technology, governance, and society.

Key Findings

- Established an interdisciplinary team working on highly relevant topics within the scope of the project achieving key results
- The two transatlantic workshops on highly automated vehicles and unmanned aerial vehicles have been highly successful in showing the relevance of the research and approaches developed in the projects for future societal scale CPS, as well as highlighting substantial differences in regulations and certification procedures between Europe and the US, thus showing the benefits of US-DE cooperation in this field.
- A generic meta-model for societal scale CPS is currently under development. In the next project phase, research on this model, currently only done in the German part of the project, will be extended and finalized in cooperation with the US partners. Especially we will test the hypothesis that this model can serve as a basis for modelling and analysing societal scale CPS that are parametrized with region-specific societal and regulatory contexts.
- The cooperation within PIRE lays the foundation for a series of large scale high impact follow up projects both in the US and in Germany/Europe. In the next phase we will establish instruments for the continued cooperation between the partners and between these follow up projects.