Carl von Ossietzky
Universität
Oldenburg

Intelligence Complicates Matters:

On Modelling and Analysing Human-in-the-Loop Behaviour

Martin Fränzle (C. v. Ossietzky Universität Oldenburg, Germany)

Grand Séminaire MACS
Paris, October 7, 2026

High Integrity, Safety Criticality: Human-Cyber-Physical Systems







Societal Need:

Significant gains in **safety**, **performance**, **acceptance** of human-cyber-physical co-operation

High Integrity, Safety Criticality: Human-Cyber-Physical Systems





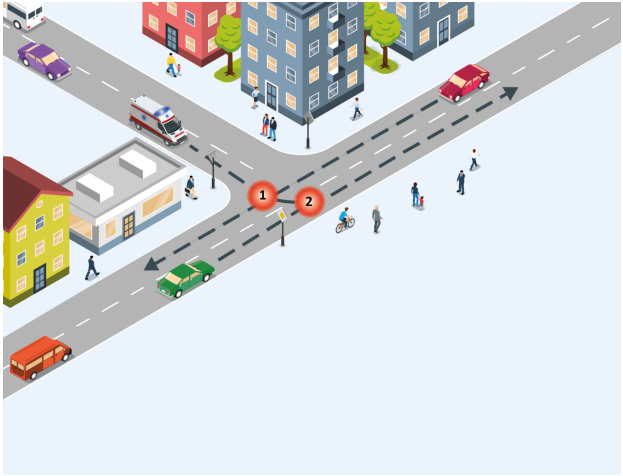


Societal Need:

Significant gains in **safety**, **performance**, **acceptance** of human-cyber-physical co-operation when

- operating in open context,
- taking critical decisions under uncertainty: in perception, in human state and intent, …

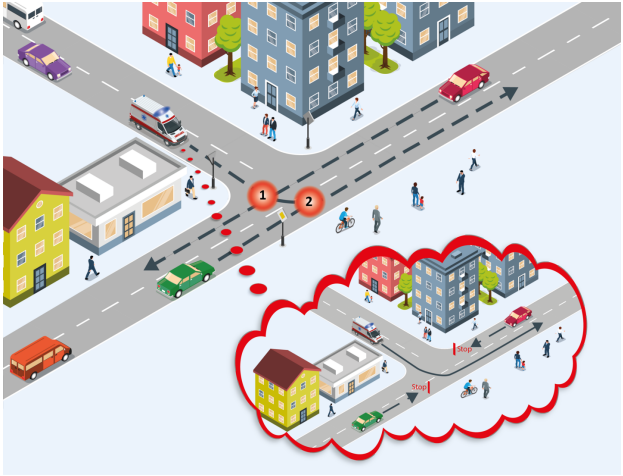
The Technological and Scientific Challenges — A Virtually Endless List



Distributed Sensing

- uncertainty
- diverse sensor variants & types
- common-cause faults
- aggregation of perspectives
- inconsistent world views
- privacy vs. informedness
- identification of malicious data
- …

The Technological and Scientific Challenges — A Virtually Endless List

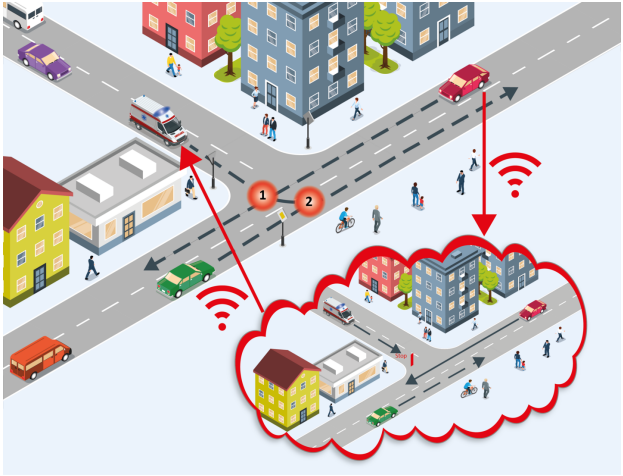


Distributed Sensing

Distributed Planning & Control

- fluctuating system boundaries
- diverse capabilities
- diverse local control algorithms
- controllability fluctuating due to bandwidth variation and jitter
- scalability & interoperability
- fault isolation
- emergence
- …

The Technological and Scientific Challenges — A Virtually Endless List



Distributed Sensing

Distributed Planning & Control

Communication

- ad-hoc networks
- rapidly changing topology
- robustness under disturbances
- intrusion & attack detection
- scalability
- self-adaption to environmental conditions
- …

The Technological and Scientific Challenges — A Virtually Endless List



Distributed Sensing

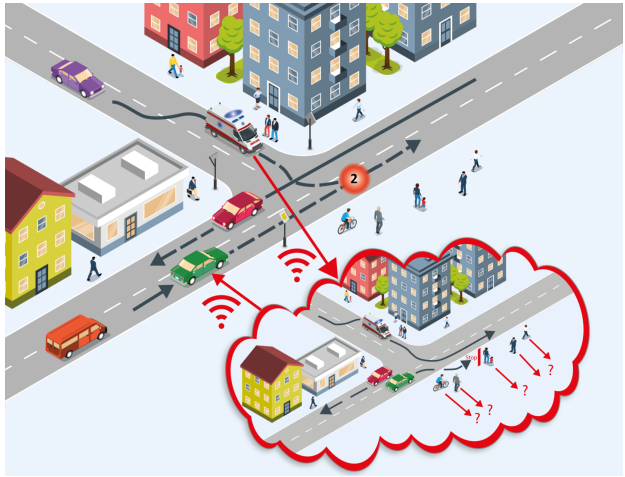
Distributed Planning & Control

Communication

Interoperability & Syst. Evolution

- smart/dumb/legacy cars
- physical limits of individual cars
- in-field monitoring and diagnosis
- in-field updates
- …

The Technological and Scientific Challenges — A Virtually Endless List



Distributed Sensing

Distributed Planning & Control

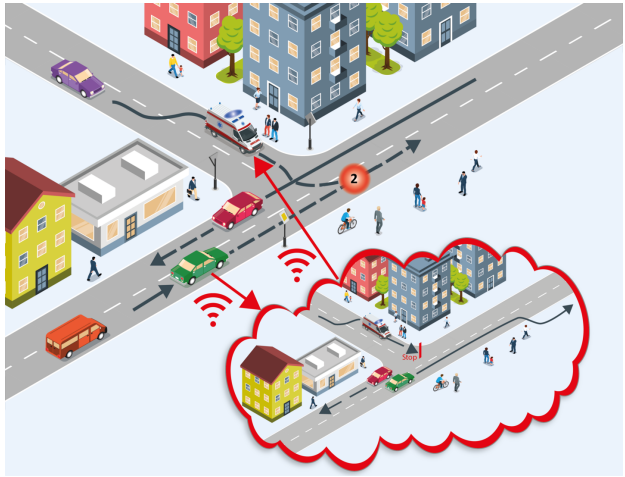
Communication

Interoperability & Syst. Evolution

Humans in the Loop

- human state
- emotions
- workload
- variation in task performance
- individual skills sets & capabilities
- situation awareness
- intent & plan
- adequacy of reaction
- …

The Technological and Scientific Challenges — A Virtually Endless List



Distributed Sensing

Distributed Planning & Control

Communication

Interoperability & Syst. Evolution

Humans in the Loop

- human state
- emotions
- workload
- variation in task performance
- individual skills sets & capabilities
- situation awareness
- intent & plan
- adequacy of reaction
- …

The Technological and Scientific Challenges — A Virtually Endless List



Distributed Sensing

Distributed Planning & Control

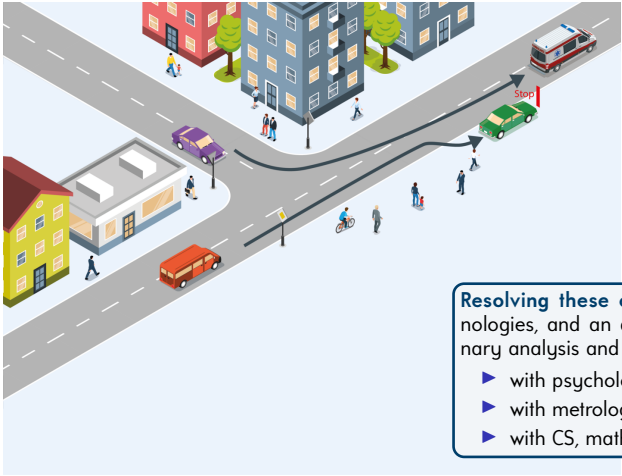
Communication

Interoperability & Syst. Evolution

Humans in the Loop

- human state
- emotions
- workload
- variation in task performance
- individual skills sets & capabilities
- situation awareness
- intent & plan
- adequacy of reaction
- …

The Technological and Scientific Challenges — A Virtually Endless List



Distributed Sensing
Distributed Planning & Control
Communication
Interoperability & Syst. Evolution
Humans in the Loop

Resolving these challenges requires methods, technologies, and an architecture permitting transdisciplinary analysis and design

- with psychology, sociology, ethics, law,
- with metrology, control, communications,
- with CS, math, physics, …

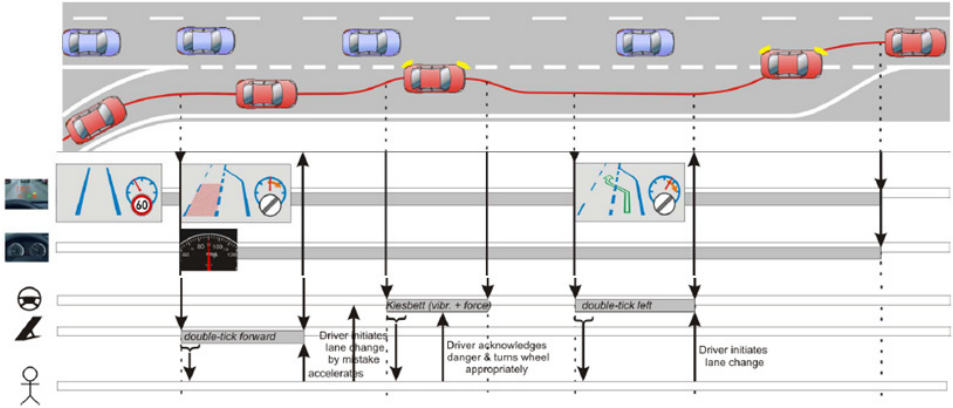
Agenda

1. Human-Cyber-Physical System (HCPS)
2. The safety targets
3. Co-modelling of H+C+P system components
4. Simulation-based quantitative safety analysis
5. Strategy synthesis based on computational cognitive models
6. System design and safety verification based on learned models for human state detection and behaviour prediction
7. Model translation

High Integrity, Safety-Critical Human-Cyber-Physical Systems

Which systems and which safety targets are we after?

Deriving Quantitative Safety Targets: A Safety-Critical Manoeuvre



Hardness: What Safety Targets are we After?

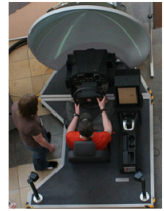
- Statistically[1], we have one injury per $1.46 \cdot 10^6$km road traffic in Germany.
  If every tenth collision is assumed to lead to some injury, this means one collision every $1.5 \cdot 10^5$km on average.
- I.e., likelihood of collision during filtering in should be of order $5 \cdot 10^{-6}$ to 10^{-5} per manoeuvre — the latter if we accept filtering in being of above average risk.

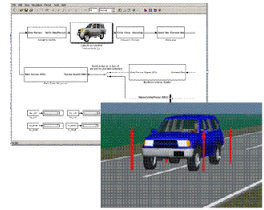
[1] Ingeborg Vorndran: Unfallstatistik — Verkehrsmittel im Risikovergleich. Statistisches Bundesamt, 12/2010

Model-Based Analysis in the Design of Advanced Driver Assistance









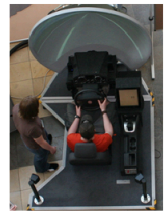
realism
cost & time
risk

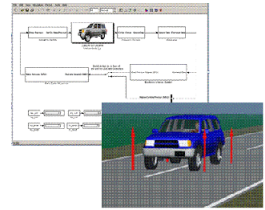
availability
speed
coverage

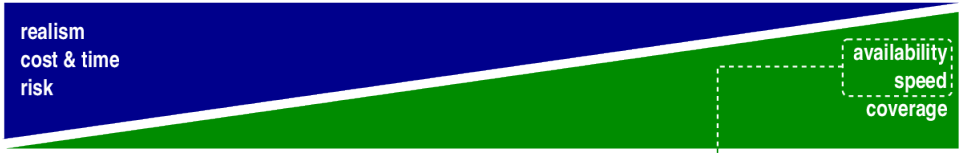
Model-Based Analysis in the Design of Advanced Driver Assistance









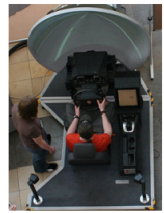


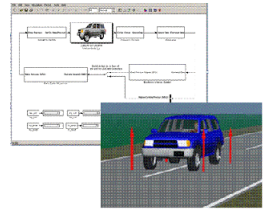
All the usual benefits from model–based design:
A thousand design flaws eliminated before the first expensive prototype is built.

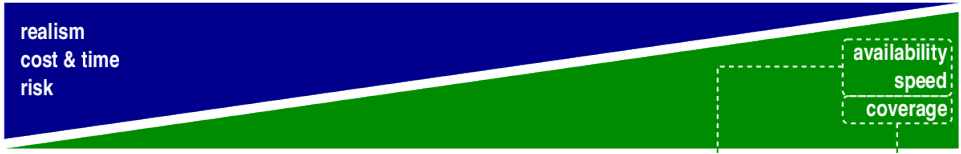
Model-Based Analysis in the Design of Advanced Driver Assistance









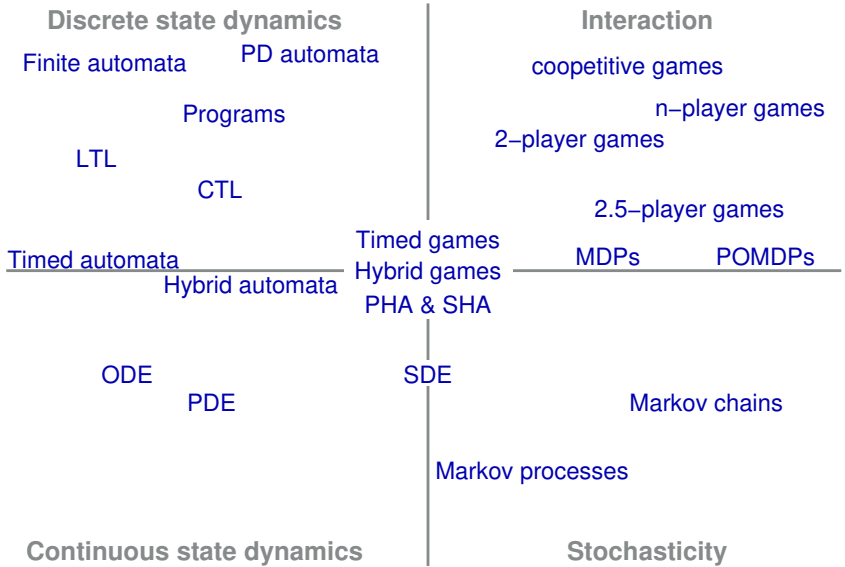


All the usual benefits from model–based design:
A thousand design flaws eliminated before the first expensive prototype is built.

Enhances safety analysis:
Improved coverage of risky situations, rare situations, hard to provoke situtations...

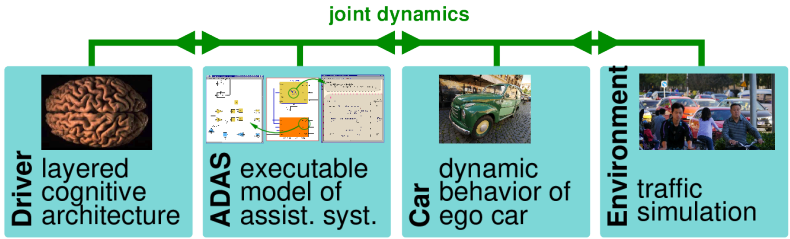
Facilitating Model-based Design and Analysis of HCPSes

Human-Cyber-Physical (Co-)Modelling

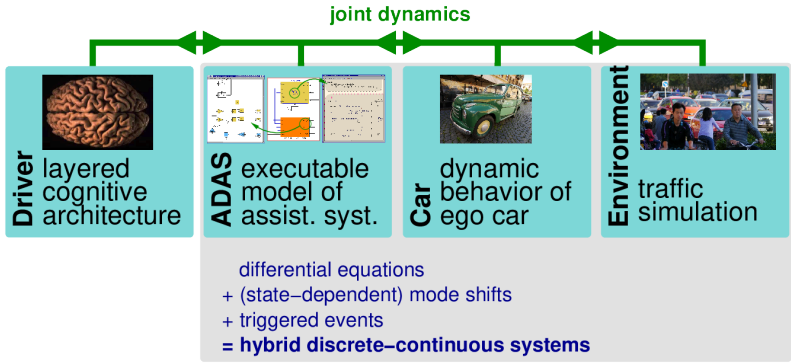


The Model Zoo

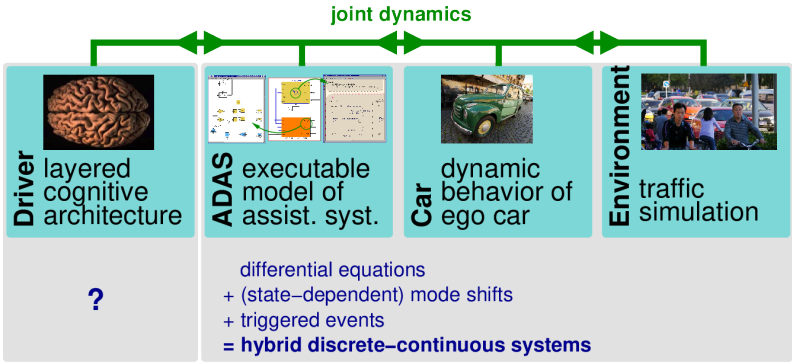
Would Human Cognition & Control Add Another MoC?



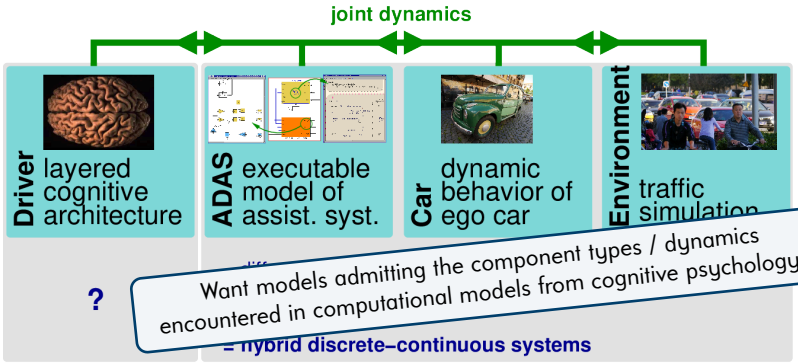
Would Human Cognition & Control Add Another MoC?



Would Human Cognition & Control Add Another MoC?



Would Human Cognition & Control Add Another MoC?



Modeling Options for Human Behavior

1. ML model mining
   - comprehensive coverage of behaviourally observable effects
   - relatively simple build process
   - rare-event coverage hard to achieve

Modeling Options for Human Behavior

1. ML model mining
   - comprehensive coverage of behaviourally observable effects
   - relatively simple build process
   - rare-event coverage hard to achieve

2. Import of computational models from cognitive psychology, social sciences, …
   - thoroughly validated models, albeit often in very specific experimental contexts
   - edge cases covered by general cognitive theories
   - purpose- and sometimes even scenario-specific

From Cognitive Psychology to Technology:
A Plea for Computational Modeling

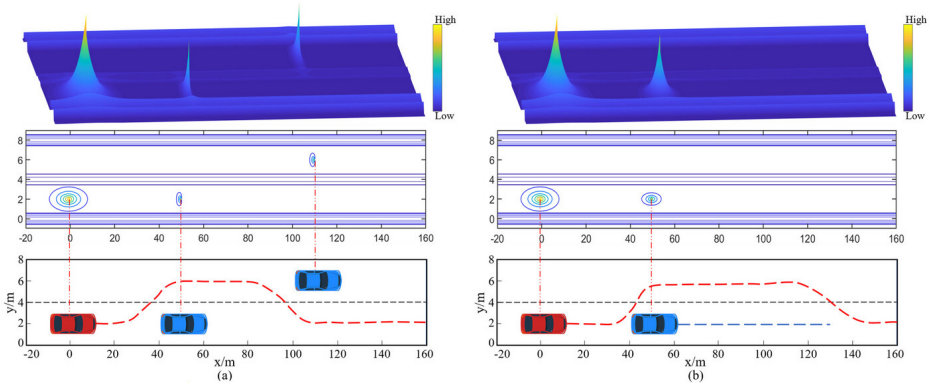
> *"Although modern cognitive psychology contains a wealth of knowledge of human behavior, it is not a simple matter to bring this knowledge to bear on the practical problems of design [of interactive computer-based devices]"*

S.K. Card, T.P. Moran, A. Newell 1983: The psychology of human-computer interaction, p. vii

Models from Cognitive Psychology and Social Sciences I

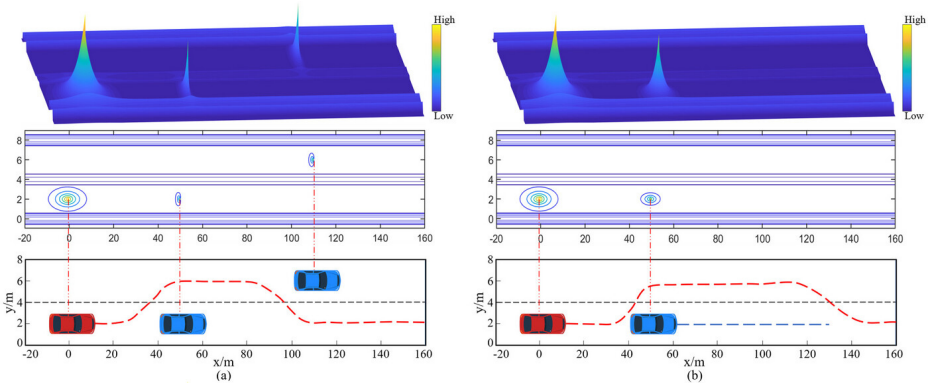
Potential-Field Models

Potential-Field Models: Guidance by Induced Forces



[Sijing Guo, Shengwen Zheng, Ji Li, Quan Zhou, Hongming Xu 2024]

Potential-Field Models: Guidance by Induced Forces



[Sijing Guo, Shengwen Zheng, Ji Li, Quan Zhou, Hongming Xu 2024]

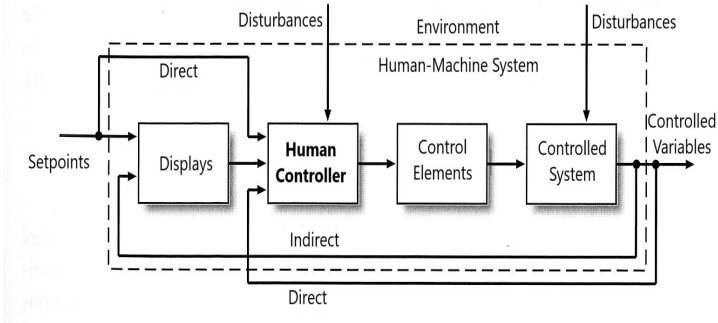
- Established model with long tradition in traffic psychology (beginning in the 1920s)
- Explains most local trajectory selection well (preference f. open space, evasive manoeuvres)
- Hard to fit to varying geometric contexts; unwanted equilibria or stable oscillations; no strategic lookahead

Models from Cognitive Psychology and Social Sciences II

Human-as-a-Controller Models

Origin

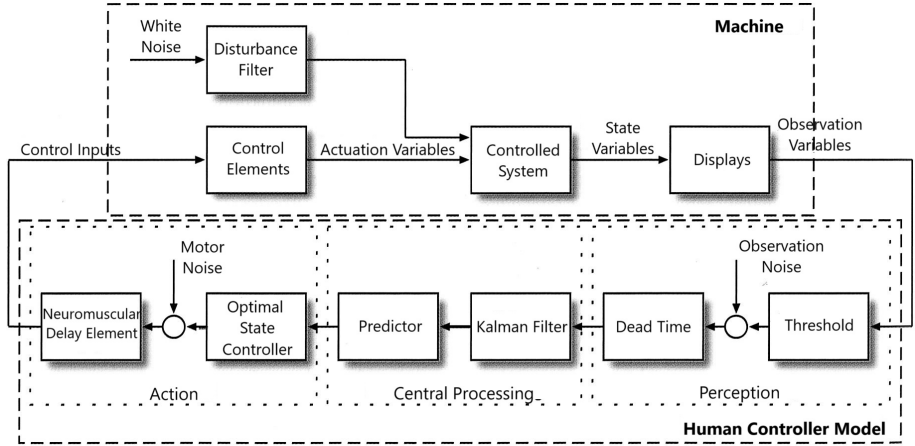
- The (now somewhat historic) success of cybernetics popularized use of feed-back loops as general paradigm for describing reactive behavior
- Leading to human-as-a-controller models for interaction of humans with their physical environment:



Structure of manual control due to Johannsen, 1993

Optimization-theoretic Model

due to Kleinmann, Baron und Levison, 1970:



Structure after Johannsen, 1990

[Nonlinearities in observability thresholds, disturbances, control path]

General Problems in Employing Control-Theoretic Models

- Structure and parameters of the controllers are hard or even impossible to determine experimentally in the nonlinear case, given the inherently limited number and diversity (risk!) of experiments with humans
- Impact of (assumed to be) symbolic cognitive functions not directly representable.

General Problems in Employing Control-Theoretic Models

- Structure and parameters of the controllers are hard or even impossible to determine experimentally in the nonlinear case, given the inherently limited number and diversity (risk!) of experiments with humans
- Impact of (assumed to be) symbolic cognitive functions not directly representable.

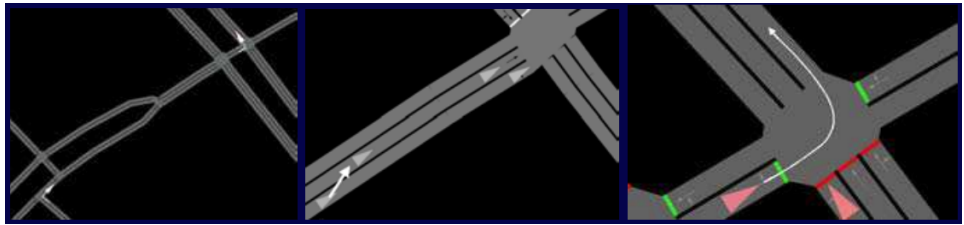
Solution: Hybrid multi-layer models

- split into diverse layers of heterogeneous nature
- split into state-based layer models with “thin” interfaces between layers.

Models from Cognitive Psychology and Social Sciences III

Three Layer Models

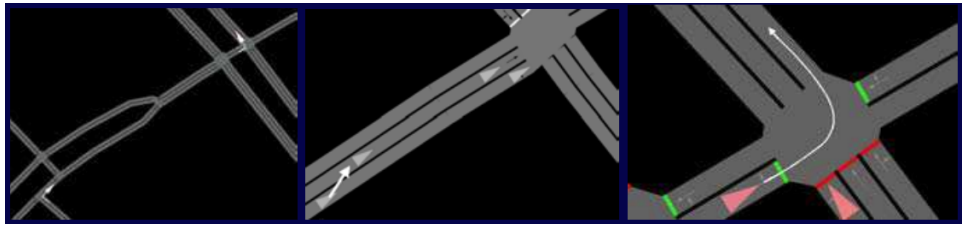
Three Layer Model: Rationale after Ranney, 1994



A.o., Ranney (1994) identifies three experimentally distinguishable layers in human car control:

1. navigation level (abstract route within the road network),
2. tactical level (e.g., decision to change the lane,)
3. control level (following the lane geometry)

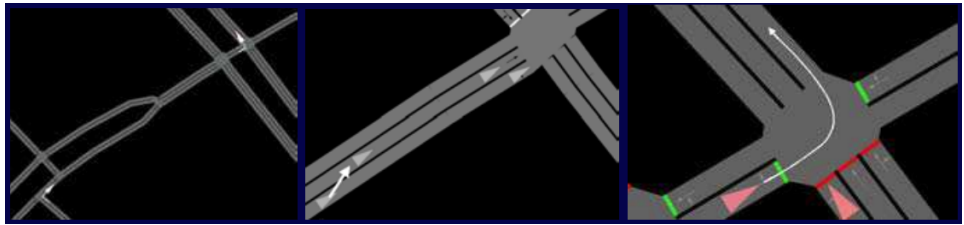
Three Layer Model: Rationale after Ranney, 1994



A.o., Ranney (1994) identifies three experimentally distinguishable layers in human car control:

1. navigation level (abstract route within the road network), "Navigation", $>$ 1 minute
2. tactical level (e.g., decision to change the lane,)
3. control level (following the lane geometry)

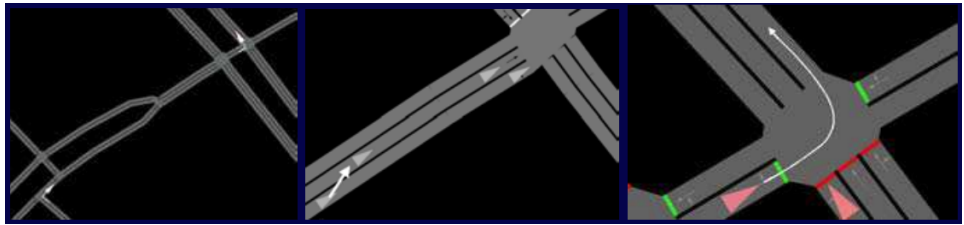
Three Layer Model: Rationale after Ranney, 1994



A.o., Ranney (1994) identifies three experimentally distinguishable layers in human car control:

1. navigation level (abstract route within the road network), "Navigation", > 1 minute
2. tactical level (e.g., decision to change the lane,) "Guidance", 1 sec. – 1 min.
3. control level (following the lane geometry)

Three Layer Model: Rationale after Ranney, 1994

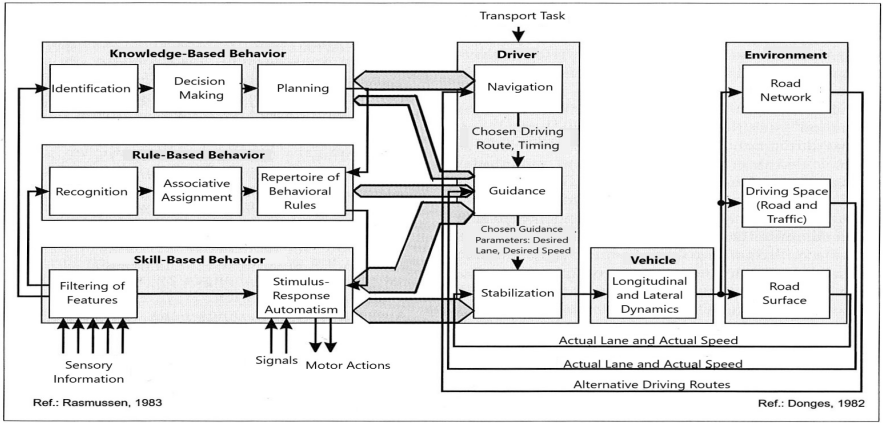


A.o., Ranney (1994) identifies three experimentally distinguishable layers in human car control:

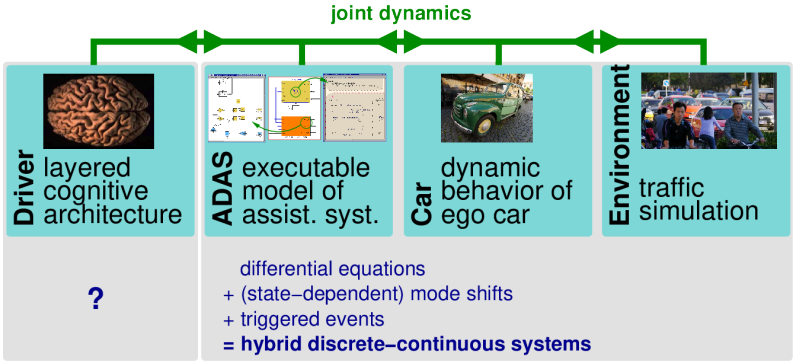
1. navigation level (abstract route within the road network), "Navigation", $>$ 1 minute
2. tactical level (e.g., decision to change the lane,) "Guidance", 1 sec. – 1 min.
3. control level (following the lane geometry) "Stabilization", $<$ 1 second

Modeling paradigms may vary between layers!

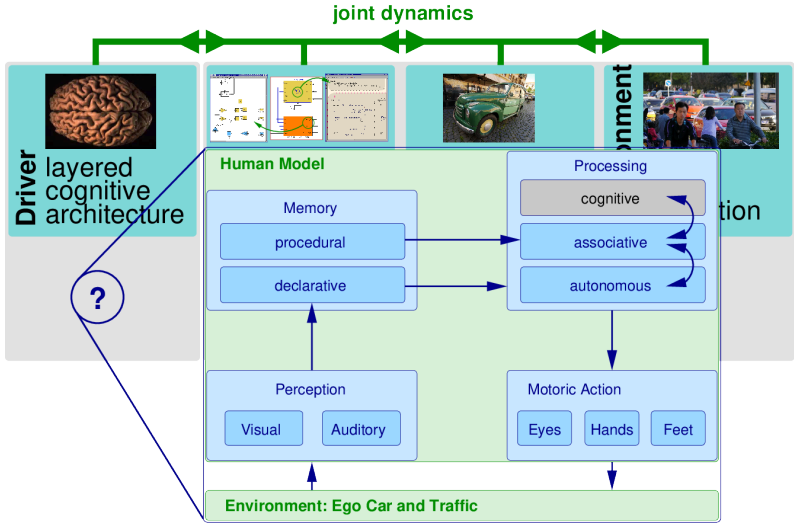
Three Layer Model: Heterogeneous MoCs after Donges and Rasmussen, 1982/83



A MoC for Human Cognition: Cognitive Architectures



A MoC for Human Cognition: Cognitive Architectures



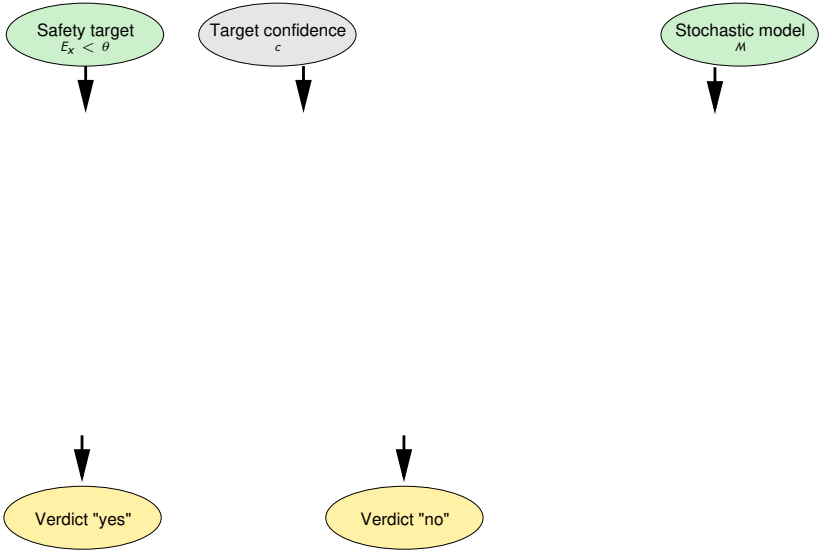
Cognitive Architectures

- Provide modular and heterogeneously structured (“codelets”), executable models of human cognition
  - instantiable to various contexts
  - modularity permits substantial reuse when transferring to a new domain
  - module descriptions / module programming employs heterogeneous MoCs:
    stochastic IO automata for perception and action, continuous signal transducers or predefined controllers for describing autonomous control skills, logic programming for goal decomposition at tactical and reactive level, associative memory for percept processing, linear filters plus tresholds for motoric dynamics, Bayesian networks, …
- Established frameworks available:
  - ACT-R [J.R. Anderson, C. Lebiere 1990–]
  - IDA/LIDA [S. Franklin, A. Kelemen, L. McCauly 1998–2003, J. Baars, S. Franklin 2009–]
  - Soar [J. Laird, A. Newell, P. Rosenbloom 1983–2012]
  - CASCaS [A. Lüdtke et al. 2004–]
- Co-simulation with CPS models feasible
  - using simulator coupling frameworks like the High-Level Architecture HLA [DoD 1998, IEEE 2000, 2010, 2025] or the Functional Mock-Up Interface FMI [MAP 2010–] or
  - embedding as (pre-compiled) functional blocks into Simulink etc. or
  - embedding as (pre-compiled) functional blocks into domain-specific simulators

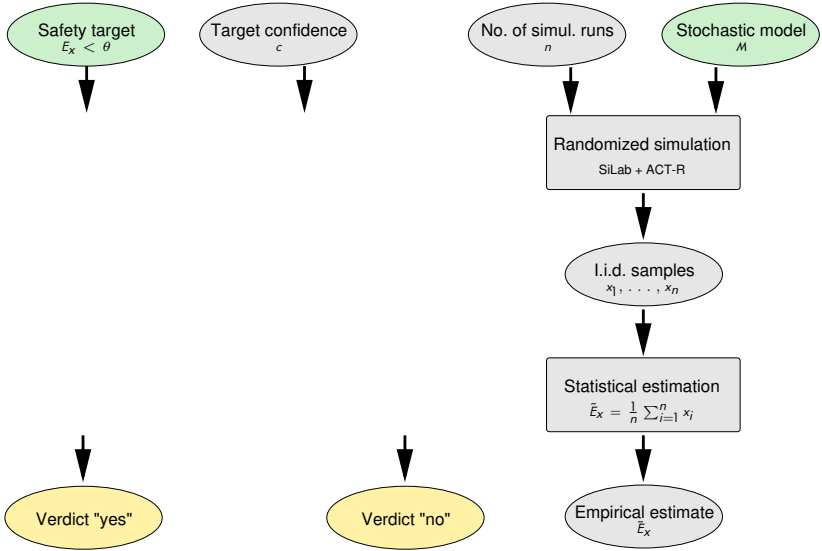
Model-based Quantitative Safety Evaluation

Exploiting Co-Simulation

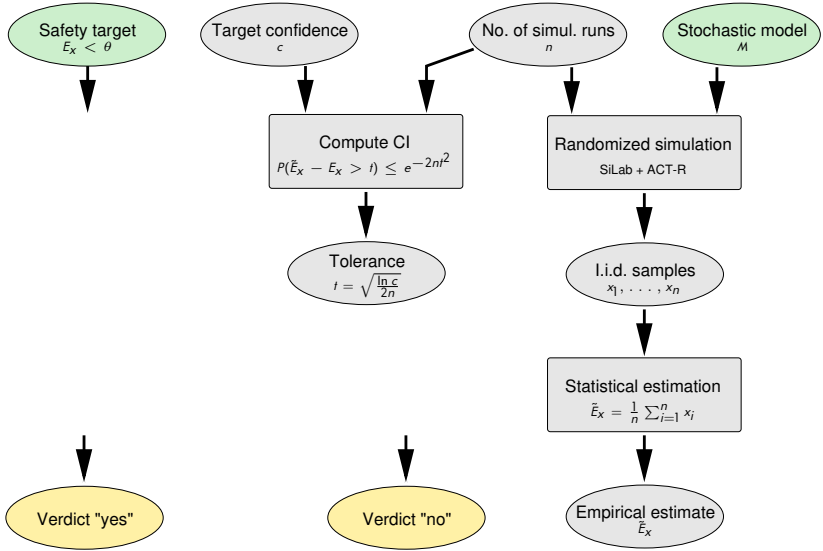
Simulation-based Verification: Statistical Model Checking [Younes, Simmons 2002–]



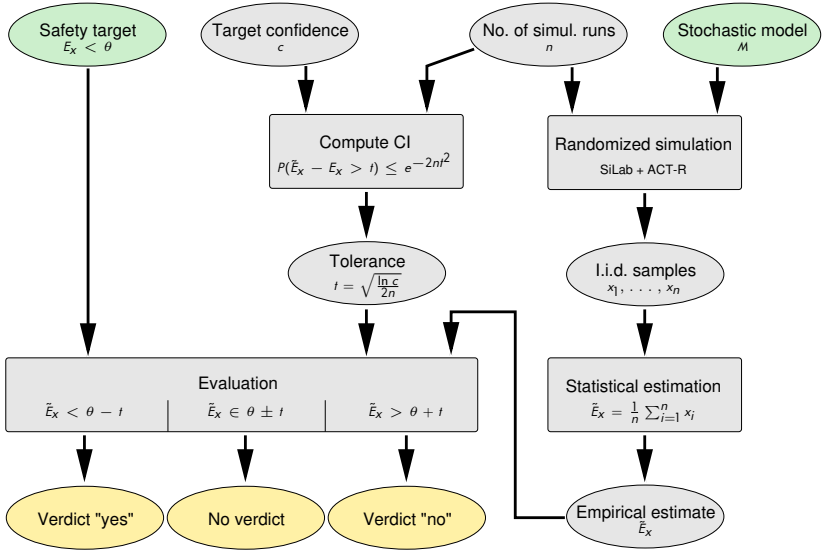
Simulation-based Verification: Statistical Model Checking [Younes, Simmons 2002–]



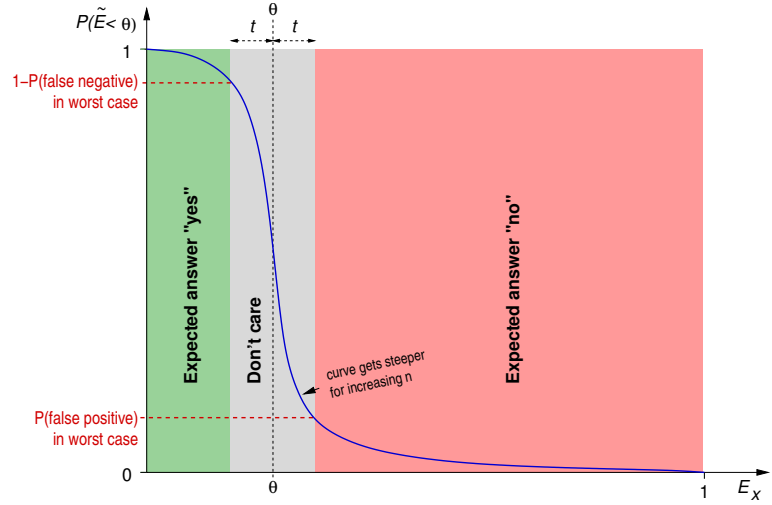
Simulation-based Verification: Statistical Model Checking [Younes, Simmons 2002–]



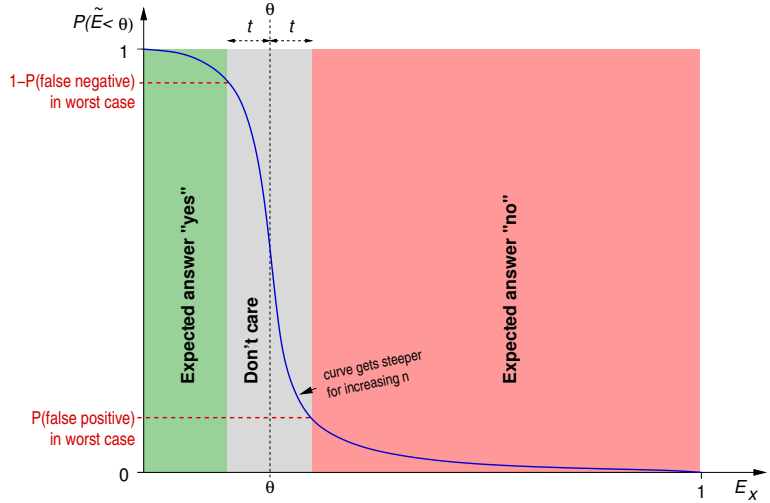
Simulation-based Verification: Statistical Model Checking [Younes, Simmons 2002–]



Interpretation of Result

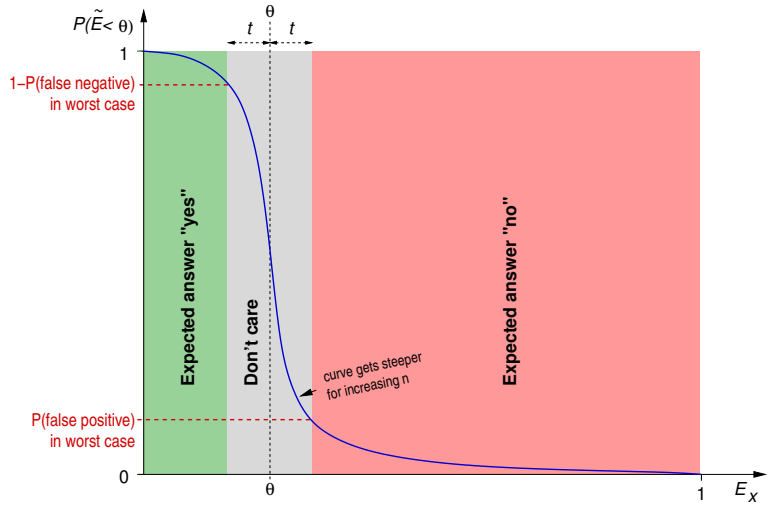


Interpretation of Result



Obs.: $P(\text{false certificate}) \to 0$ if $n \to \infty$ or $t \to \max(\theta, 1-\theta)$.

Interpretation of Result



But what if safety targets are high and interesting events are rare?

Statistical Evaluation: Findin the Needle in the Haystack



Scenario: Driving at a winding road with varying curve radii

- Target speed $100\,\frac{\mathrm{km}}{\mathrm{h}}$
- Secondary in-vehicle task: number read-back

Question: How likely is the distraction to cause a near hit of a bridge pillar far down the track?

Statistical Evaluation: Findin the Needle in the Haystack



Scenario: Driving at a winding road with varying curve radii

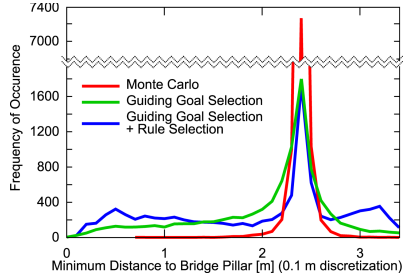
- Target speed $100\,\frac{\mathrm{km}}{\mathrm{h}}$
- Secondary in-vehicle task: number read-back

Question: How likely is the distraction to cause a near hit of a bridge pillar far down the track?

Analysis Method: Randomized co-simulation of car dynamics, assistant system, driver model, …

Result: Within 10k simulations ($\approx$ 1 week)

- Monte Carlo simulation won't spot any critical case,
- criticality-guided simulation [Frä., Gerwinn, Puch, Wortelen 2013–18] provides a useful yield & statistics of near-hits.



Statistical Evaluation: Findin the Needle in the Haystack



Scenario: Driving at a winding road with varying curve radii

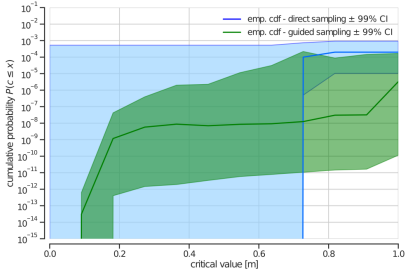
- Target speed $100\,\frac{\mathrm{km}}{\mathrm{h}}$
- Secondary in-vehicle task: number read-back

Question: How likely is the distraction to cause a near hit of a bridge pillar far down the track?

Analysis Method: Randomized co-simulation of car dynamics, assistant system, driver model, …

Result: Within 10k simulations ($\approx$ 1 week)

- Monte Carlo simulation won't spot any critical case,
- criticality-guided simulation [Frä., Gerwinn, Puch, Wortelen 2013–18] provides a useful yield & statistics of near-hits.



Criticality-Guided Simulation

General idea: Suggest a continuous criticality function that is correlated with risk

- e.g., nominal time to collision vs. actual collisions
- or quantitative 'robust' semantics of temporal logic (e.g., Duration Calculus [Frä., Hansen 2005], Signal Temporal Logic [Donzé, Maler 2010]) induce such criticality functions [Eddeland et. al 2017], [Miremadi, Fabian, Åkeson 2017]
- continuity permits use of gradient descent to find and boost 'bad' instances

Criticality-Guided Simulation

General idea: Suggest a continuous criticality function that is correlated with risk

- e.g., nominal time to collision vs. actual collisions
- or quantitative 'robust' semantics of temporal logic (e.g., Duration Calculus [Frä., Hansen 2005], Signal Temporal Logic [Donzé, Maler 2010]) induce such criticality functions [Eddeland et. al 2017], [Miremadi, Fabian, Åkeson 2017]
- continuity permits use of gradient descent to find and boost 'bad' instances

Use cases: Falsification or improved frequency statistics

- Falsification 'only' tries to improve likelihood of finding some fatality during system simulation if there is such

  (suggested/applied in various projects w. industry, a.o. V&V-Methoden, Pegasus, Set-Level 4–5)
- Improved frequency statistics changes probabilities during sampling to improve rare-event statistics
  - usually variants of importance sampling or importance splitting
  - statistics for orig. probability space is reconstructed by importance weighting [Puch, Frä., Gerwinn 2018]

That was cumbersome and computationally expensive!

Can we do with simpler models and get beyond a-posteriori verification?

Yes, particular snippets of a CA often do the job:

Exploiting cog-sci attention models in strategy synthesis

When to Provide an Explanation?

[Bairy, MF '22–'25]

- Complex technical items demand self-explanation capabilities
- Surprisingly few research has gone into beneficial explanation timing
  - Course-grain timing, i.e. before/during/after an event has been explored thoroughly, e.g., [Koo, Shin, Steinert, Leifer '16], [Haspiel, Du, Meyerson et al. '18+'19], [Chen, Liao, Sundar '24]
  - but favorable presentation time within these groups, esp. "during", hasn't gained much attention.
- Explanation timing is instrumental to workload
  - Receiving & interpreting explanations induces human workload
    $\Longrightarrow$ too early / too frequent explanations are bad
  - Self-scanning for information induces high workload
    $\Longrightarrow$ late / missing explanation is even worse

When to Provide an Explanation?

[Bairy, MF '22–'25]

- Complex technical items demand self-explanation capabilities
- Surprisingly few research has gone into beneficial explanation timing
  - Course-grain timing, i.e. before/during/after an event has been explored thoroughly, e.g., [Koo, Shin, Steinert, Leifer '16], [Haspiel, Du, Meyerson et al. '18+'19], [Chen, Liao, Sundar '24]
  - but favorable presentation time within these groups, esp. "during", hasn't gained much attention.
- Explanation timing is instrumental to workload
  - Receiving & interpreting explanations induces human workload
    $\Longrightarrow$ too early / too frequent explanations are bad
  - Self-scanning for information induces high workload
    $\Longrightarrow$ late / missing explanation is even worse

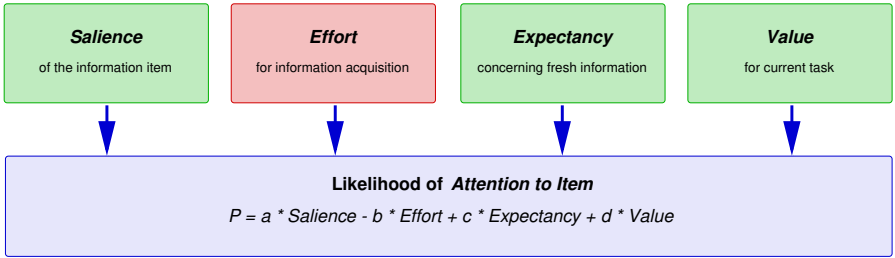
Have to provide explanations as rarely and as late as possible, but before human scan for information likely commences!

Likelihood of Paying Attention — the SEEV Model

[Wickens '03][Wickens, Carley '08]

SEEV is a computational model of attention distribution and likelihood:

- Records *Salience, Effort, Expectancy, Value* f.e. information item accessible to the human.
- Translates these into a time-variant probability of attention to the information item:



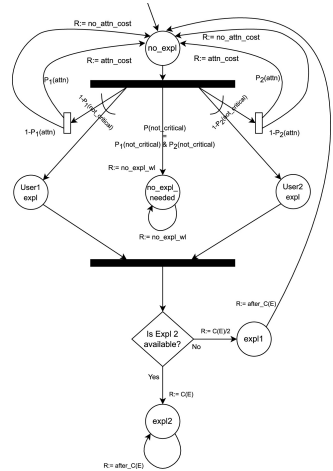
- where *Salience, Effort, Expectancy, Value* all vary situationally and depending on history, thus constitute dynamic processes to be described by hybrid automata.

Embedding SEEV into a Stochastic Game

[Bairy, Frä. '22–'25]

1. SEEV modulates instantaneous probabilities of the uncontrollable events (paying attention);
2. Decisions to display an explanation reset some SEEV variables;
3. Cost variables trace the accumulated workload.

Cost-optimal strategy for this stochastic game minimizes expected workload.



Embedding SEEV into a Stochastic Game

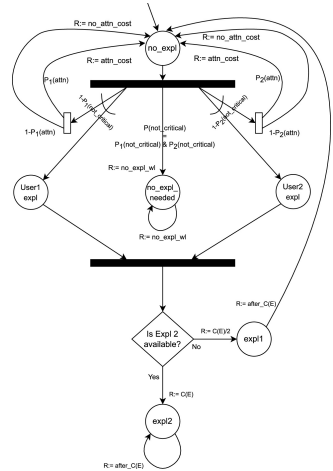
[Bairy, Frä. '22–'25]

1. SEEV modulates instantaneous probabilities of the uncontrollable events (paying attention);
2. Decisions to display an explanation reset some SEEV variables;
3. Cost variables trace the accumulated workload.

Cost-optimal strategy for this stochastic game minimizes expected workload.

Synthesize that strategy!

- with individualised parameters for SEEV dynamics, workload of actions, …
- resulting in individually varied presentation strategies!

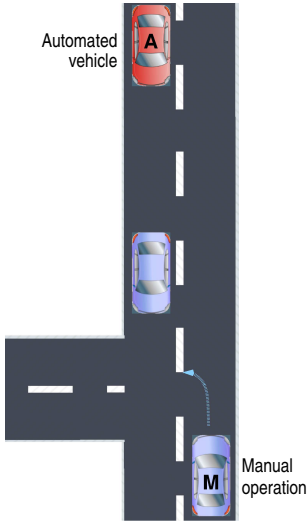


Yes, (combinations of) machine-learned models often do the job:

ACPS design & verification involving statistically qualified ML components

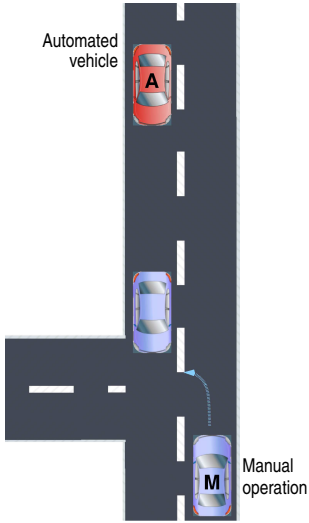
AV Safety by BCI-Enabled Introspection & Prediction

[Damm, Frä., Lüdtke, Rieger, Trende, Unni et al. 2019–]



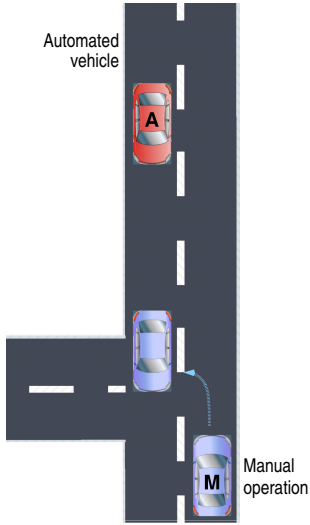
AV Safety by BCI-Enabled Introspection & Prediction

[Damm, Frä., Lüdtke, Rieger, Trende, Unni et al. 2019–]



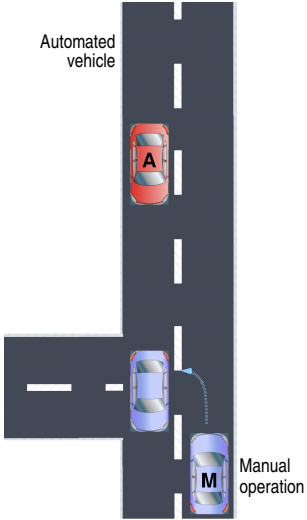
AV Safety by BCI-Enabled Introspection & Prediction

[Damm, Frä., Lüdtke, Rieger, Trende, Unni et al. 2019–]



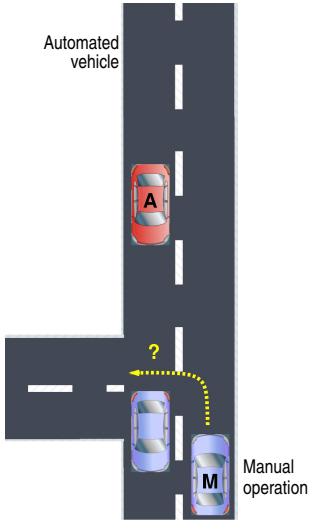
AV Safety by BCI-Enabled Introspection & Prediction

[Damm, Frä., Lüdtke, Rieger, Trende, Unni et al. 2019–]



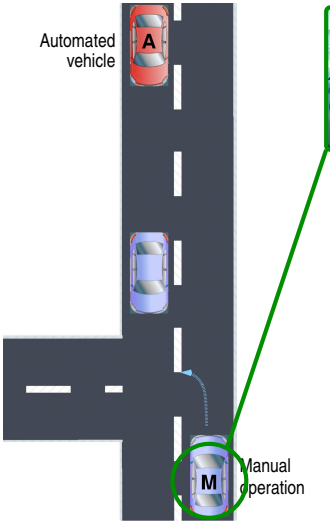
AV Safety by BCI-Enabled Introspection & Prediction

[Damm, Frä., Lüdtke, Rieger, Trende, Unni et al. 2019–]



AV Safety by BCI-Enabled Introspection & Prediction

[Damm, Frä., Lüdtke, Rieger, Trende, Unni et al. 2019–]

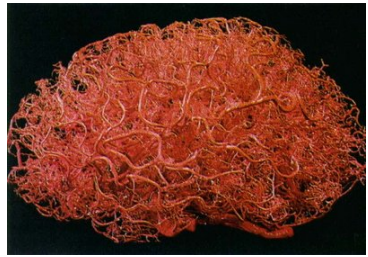




Infrared optical sensing detects delayed (4s – 6s) BOLD response in several cortical areas.

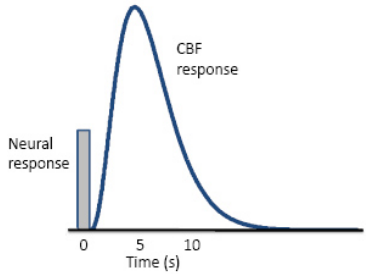
Ingredient 1: Neurophysiological Sensing by FNIRS





©Scholarpedia

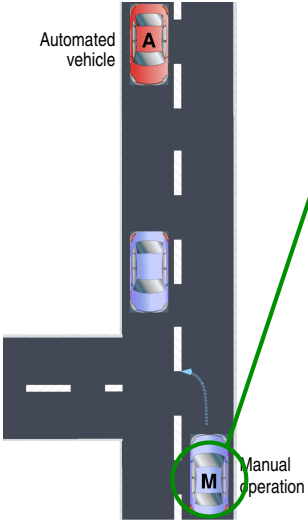
- Infrared optical sensing to detect BOLD (blood oxygen level dependent) response in cortical areas.
- BOLD follows changes in neural activity due to neurovascular coupling.
- More discriminating than peripheral physiological measurements of arousal (heart rate, pupil dilatation, skin conductance, …)
- Signal latency 4–6 s.



©Scholarpedia

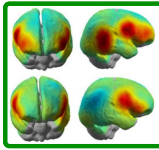
AV Safety by BCI-Enabled Introspection & Prediction

[Damm, Frä., Lüdtke, Rieger, Trende, Unni et al. 2019–]





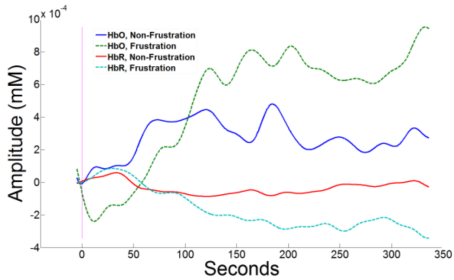
Infrared optical sensing detects delayed (4s – 6s) BOLD response in several cortical areas.

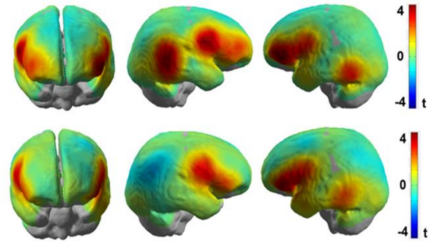


Evaluate image to discriminate frustration from non–frustration.

~ 78% accuracy

Ingredient 2: Detecting Frustration by Automatic Image Classification



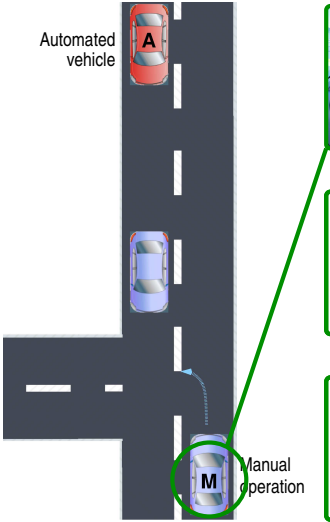


Trained neural network discriminates frustrated from non-frustrated driving intervals with 78.1% accuracy.

(Per subject basis, not yet transferable!)

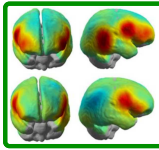
AV Safety by BCI-Enabled Introspection & Prediction

[Damm, Frä., Lüdtke, Rieger, Trende, Unni et al. '19–'22]



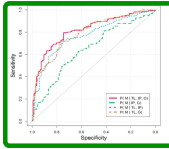


Infrared optical sensing detects delayed (4s – 6s) BOLD response in several cortical areas.



Evaluate image to discriminate frustration from non–frustration.

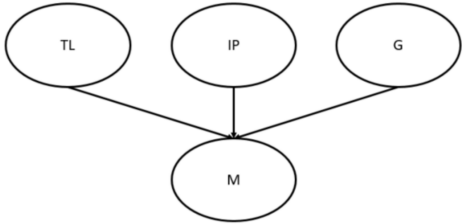
~ 78% accuracy



Predict gap selection from frustration level + environm. conditions.

~ 75% accuracy

Ingredient 3: Behaviour Prediction
— Setting up a Baysean Network Predicting Likelihood of Risky Merge

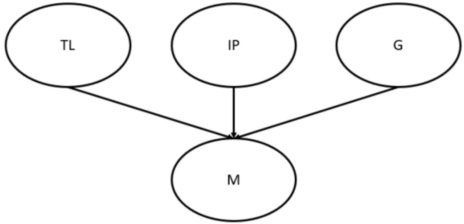


Probability of merging decision as a function of

- frustration level & gap length,
- interaction partner type (human/AV),
- driver's gender.

Determined from driving experiments with human subjects.

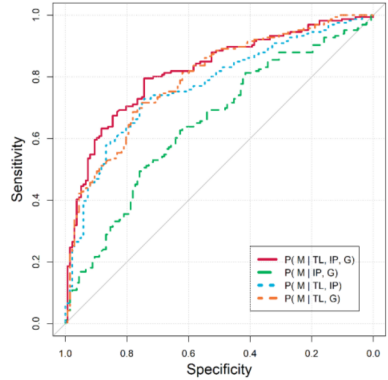
Ingredient 3: Behaviour Prediction
— Setting up a Baysean Network Predicting Likelihood of Risky Merge



Probability of merging decision as a function of

- frustration level & gap length,
- interaction partner type (human/AV),
- driver's gender.

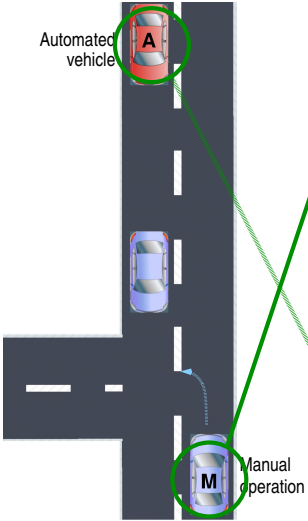
Determined from driving experiments with human subjects.



10-fold cross validation proves ≈75% accuracy.

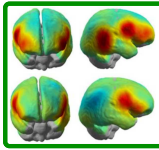
AV Safety by BCI-Enabled Introspection & Prediction

[Damm, Frä. , Lüdtke, Rieger, Trende, Unni et al. 2019–]



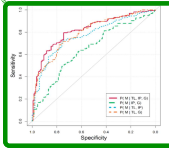


Infrared optical sensing detects delayed (4s – 6s) BOLD response in several cortical areas.



Evaluate image to discriminate frustration from non–frustration.

~ 78% accuracy

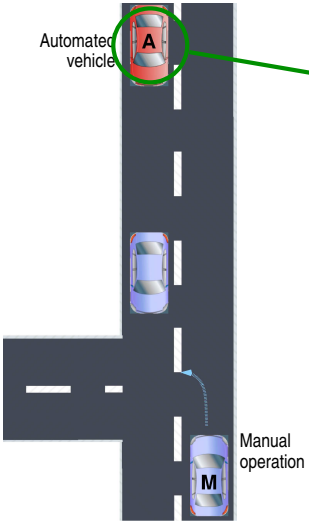


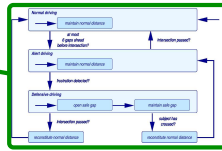
Predict gap selection from frustration level + environm. conditions.

~ 75% accuracy

AV Safety by BCI-Enabled Introspection & Prediction

[Damm, Frä. , Lüdtke, Rieger, Trende, Unni et al. 2019–]

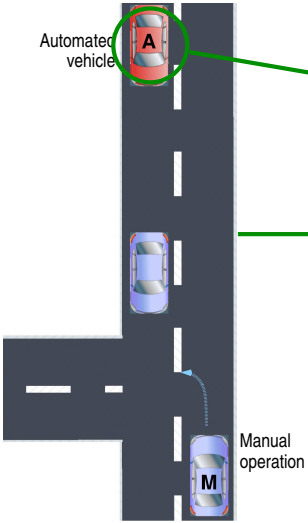


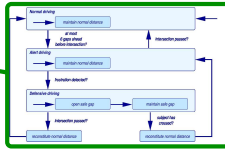


Generation of hybrid–state control strategy robust under uncertainty.

AV Safety by BCI-Enabled Introspection & Prediction

[Damm, Frä. , Lüdtke, Rieger, Trende, Unni et al. 2019–]





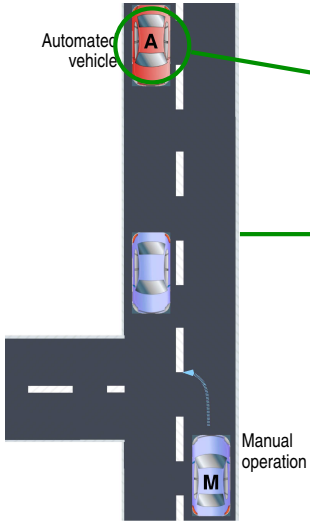
Generation of hybrid–state control strategy robust under uncertainty.

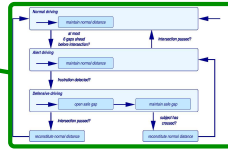
Pursue quantitative safety verification of full system by model checking.

Worst–scenario likelihood of intruding into critical gap reduced ~ 3.3–fold!

AV Safety by BCI-Enabled Introspection & Prediction

[Damm, Frä. , Lüdtke, Rieger, Trende, Unni et al. 2019–]





Generation of hybrid–state control strategy robust under uncertainty.

Pursue quantitative safety verification of full system by model checking.

Worst–scenario likelihood of intruding into critical gap reduced ~ 3.3–fold!

	Physiological Measurement
+	Cognitive Modeling
+	Models of CPS
+	Robust Planning & Control
+	Stimulation of Human Action
+	Formal System Verification
=	Significant Gains in Safety & Performance

Yes, mimicking the cognitive architecture opens new pathways:

Employing automata learning to feed cognitive architectures into algorithmic control synthesis

Cognitive Architectures and Control Synthesis Don't Like Each Other

Problem: Cognitive architectures are extremely heterogeneous MoCs, containing and seamlessly integrating a.o.

- Continuous controllers
- Continuous signal transducers
- Finite automata
- Rule-based goal decomposition akin to logic programming
- Stochastic disturbances & randomized conflict resolution in goal selection
- Associative memory (with stochastically timed decay rules)
- Learning
- …

No current control synthesis tool comes close to addressing this combination.

Cognitive Architectures and Control Synthesis Don't Like Each Other

Problem: Cognitive architectures are extremely heterogeneous MoCs, containing and seamlessly integrating a.o.

- Continuous controllers
- Continuous signal transducers
- Finite automata
- Rule-based goal decomposition akin to logic programming
- Stochastic disturbances & randomized conflict resolution in goal selection
- Associative memory (with stochastically timed decay rules)
- Learning
- …

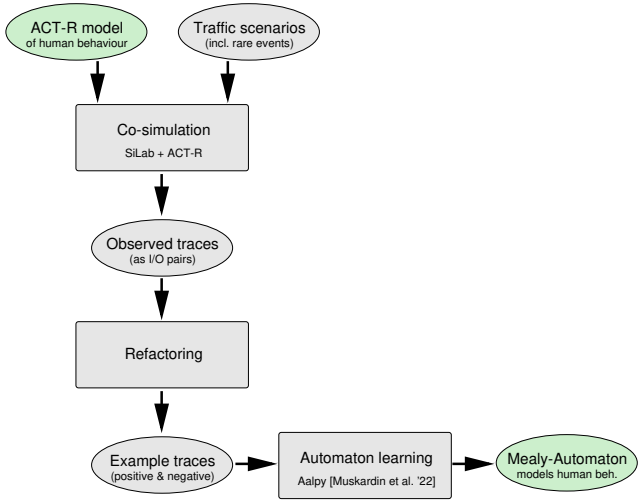
No current control synthesis tool comes close to addressing this combination.

Remedy: Generate (hybrid) automata models mimicking the CA

- by syntactic translation [Langenfeld, Westphal, Podelski '19–]
- by automata learning from traces [Hajnorouzi, Frä. '26]

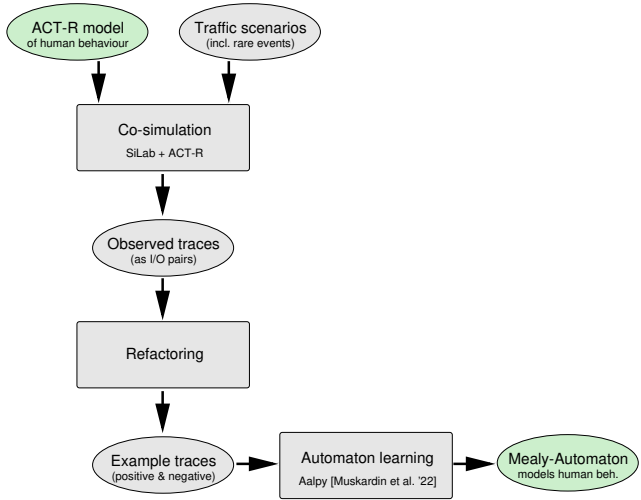
Active Automata Learning "Translating" the ACT-R Model

[Hajnorouzi, Frä. '26]



Active Automata Learning "Translating" the ACT-R Model

[Hajnorouzi, Frä. '26]

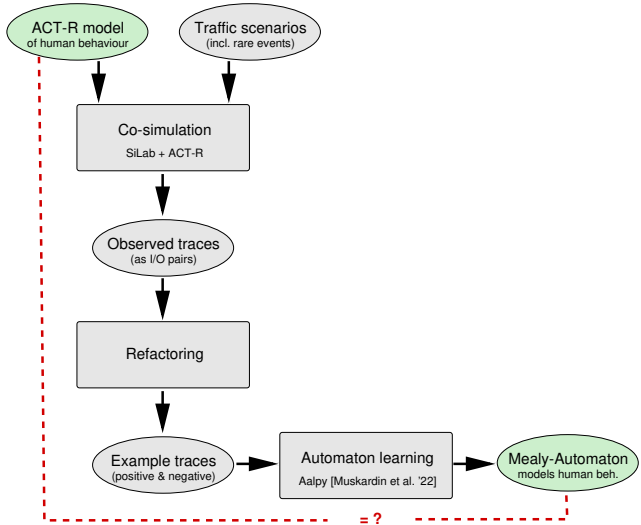


Extracted automaton amenable to

- exhaustive analysis of dynamic properties by model checking
- discrete-event control synthesis, when interpreted as reactive game graph

Active Automata Learning "Translating" the ACT-R Model

[Hajnorouzi, Frä. '26]

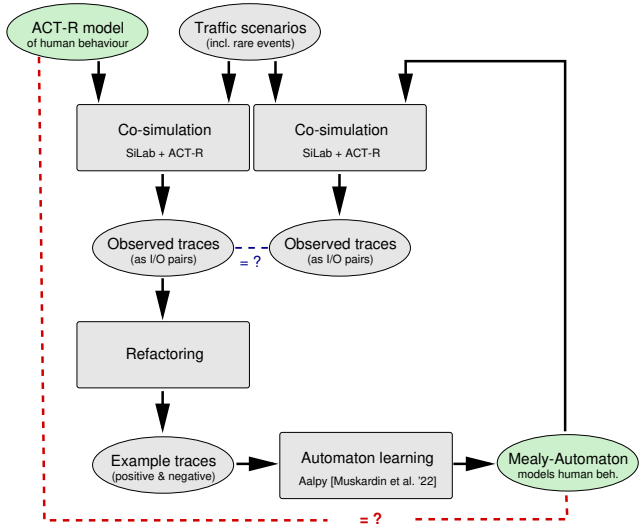


Extracted automaton amenable to

- exhaustive analysis of dynamic properties by model checking
- discrete-event control synthesis, when interpreted as reactive game graph

Active Automata Learning "Translating" the ACT-R Model

[Hajnorouzi, Frä. '26]

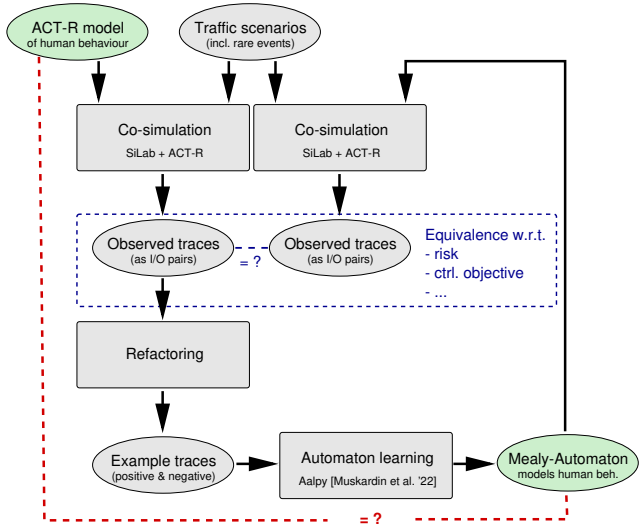


Extracted automaton amenable to

- exhaustive analysis of dynamic properties by model checking
- discrete-event control synthesis, when interpreted as reactive game graph

Active Automata Learning "Translating" the ACT-R Model

[Hajnorouzi, Frä. '26]



Extracted automaton amenable to

- exhaustive analysis of dynamic properties by model checking
- discrete-event control synthesis, when interpreted as reactive game graph

Feasibility and Precision: Car-following Case Study

- 336 test scenarios
- Includes comprehensive rare-event coverage:
  84 cases uncontrollable by “normally skilled” ACT-R model
- Automata learning achieved perfect match w.r.t. relevant properties:

	ACT-R	Automaton
Collision	84	84
Non-progress	173	173
Safe progress	79	79

(on an independent set of test scenarios)

- Matches not only frequentistically, but also case-to-case!

Summing Up

Did We Make Some Mileage?

Summary

1. Behavioural models of HCPSes can be built
   - by seamlessly coupling models from cognitive psychology etc. with established modelling frameworks for cyber-physical systems
   - the resulting models can in principle be developed to any level of detail and accuracy
   - but validation, inter-subject variation, etc. remain a problem, especially when transferring them between subjects, scenarios, operational design domains, or even betw. application domains

Summary

1. Behavioural models of HCPSes can be built
   - by seamlessly coupling models from cognitive psychology etc. with established modelling frameworks for cyber-physical systems
   - the resulting models can in principle be developed to any level of detail and accuracy
   - but validation, inter-subject variation, etc. remain a problem, especially when transferring them between subjects, scenarios, operational design domains, or even betw. application domains

2. The resulting heterogeneous models can be analysed for quantitative safety properties etc.
   - yet only by simulation due to their heterogeneity and complexity, especially the vast and heterogeneous state spaces of the embedded cognitive architectures
   - uncovering critical situations and generating accurate risk statistics requires advanced simulation mechanisms

Summary

1. Behavioural models of HCPSes can be built
   - by seamlessly coupling models from cognitive psychology etc. with established modelling frameworks for cyber-physical systems
   - the resulting models can in principle be developed to any level of detail and accuracy
   - but validation, inter-subject variation, etc. remain a problem, especially when transferring them between subjects, scenarios, operational design domains, or even betw. application domains

2. The resulting heterogeneous models can be analysed for quantitative safety properties etc.
   - yet only by simulation due to their heterogeneity and complexity, especially the vast and heterogeneous state spaces of the embedded cognitive architectures
   - uncovering critical situations and generating accurate risk statistics requires advanced simulation mechanisms

3. Use cases beyond statistical verification require MoC-constrained cognitive models, obtained by
   - "slicing" the cognitive architecture,
   - machine-learned models,
   - translation to / approximation by simple and homogeneous MoCs (e.g. automata)

Summary

1. Behavioural models of HCPSes can be built
   - by seamlessly coupling models from cognitive psychology etc. with established modelling frameworks for cyber-physical systems
   - the resulting models can in principle be developed to any level of detail and accuracy
   - but validation, inter-subject variation, etc. remain a problem, especially when transferring them between subjects, scenarios, operational design domains, or even betw. application domains

2. The resulting heterogeneous models can be analysed for quantitative safety properties etc.
   - yet only by simulation due to their heterogeneity and complexity, especially the vast and heterogeneous state spaces of the embedded cognitive architectures
   - uncovering critical situations and generating accurate risk statistics requires advanced simulation mechanisms

3. Use cases beyond statistical verification require MoC-constrained cognitive models, obtained by
   - "slicing" the cognitive architecture,
   - machine-learned models,
   - translation to / approximation by simple and homogeneous MoCs (e.g. automata)

4. ML as a model-mining tool is gaining impetus, but induces exciting new problems.