Präsentiert von

Henning Butz

Autonome Mensch Maschine Systeme (MMS) im Zeichen des Evolutionsgesetzes: The Survival of the Fittest

Self-Governing Human Machine Systems (HMS) Directed by the Law of Evolution: The Survival of the Fittest

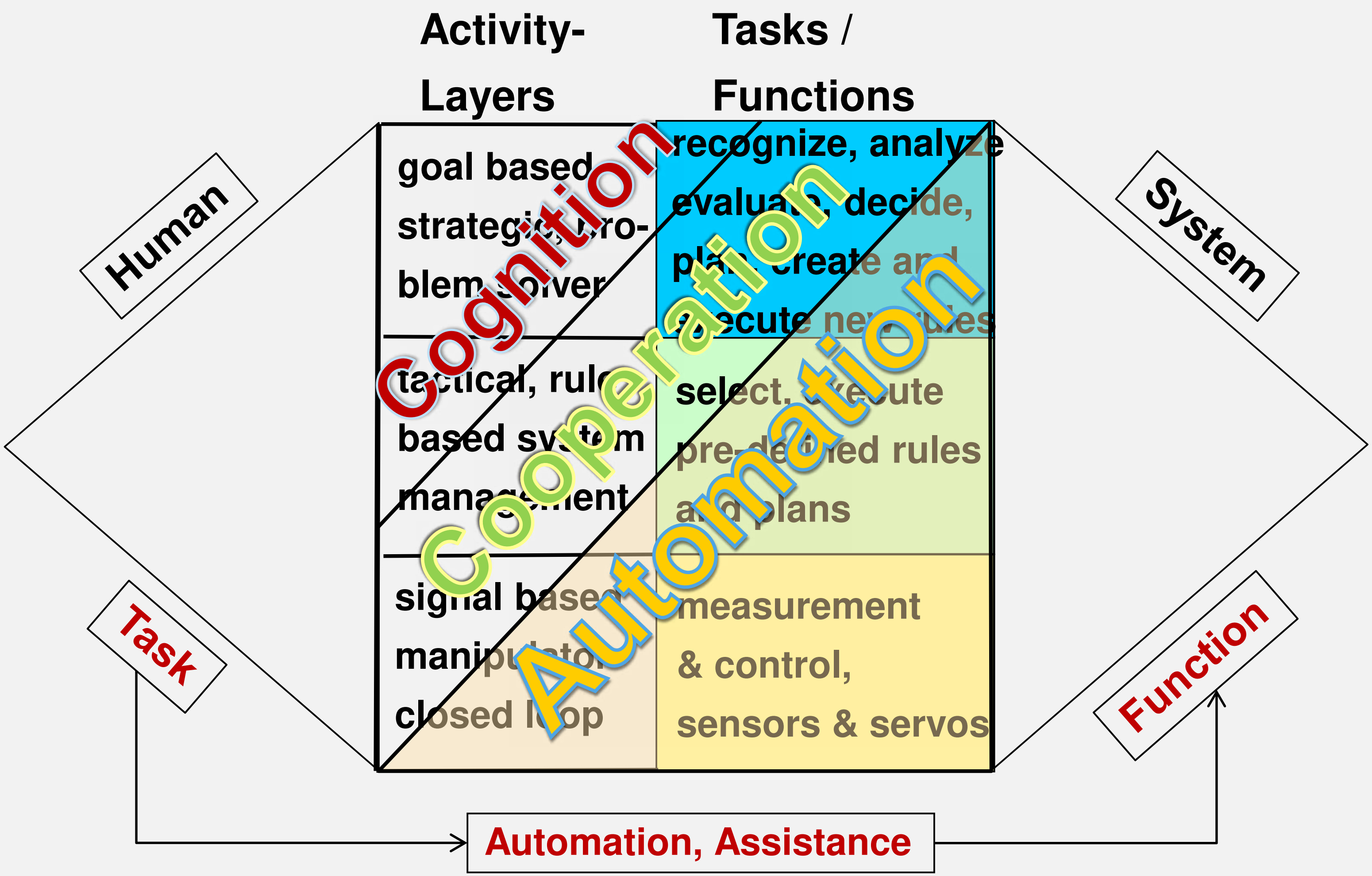
MESCONF, München, September 11th 2019

We are going to talk about …

- Automation Autonomy Levels
- Issues with Autonomy when Human is in the Loop
- Cognition (AI) as a Feature of Autonomy
- Evolution of AI World Models by Semiotic Triangles
- Going beyond Human Perception and Understanding
- Goal Driven Cognitive Automation
- The Survival of the Fittest

ASES Advanced Systems Engineering Solutions

Automation Activity Layers vs. Tasks & Functions



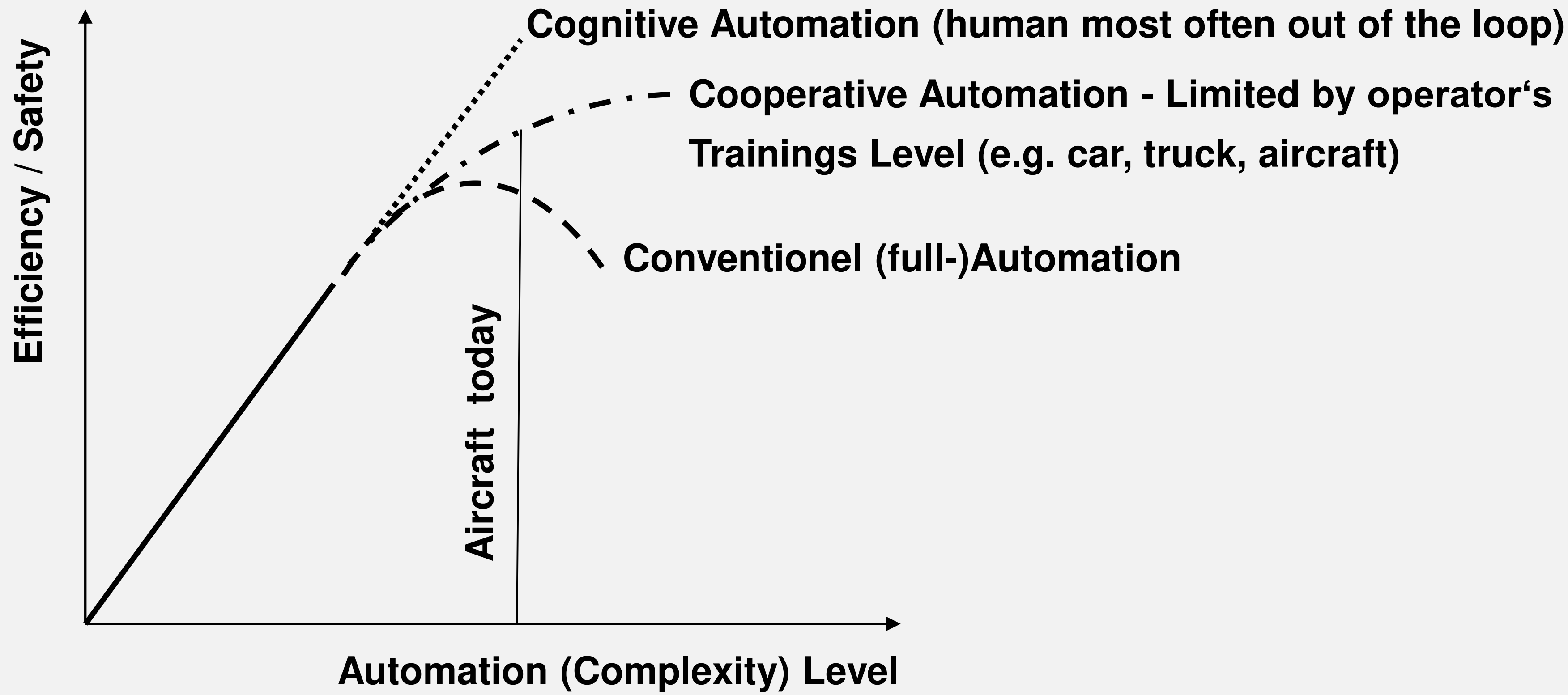
- Automation lowers the *Complexity of Tasks*, while boosting the *Function Complexity*
- When the Automation fails, the *Task Complexity* goes up instantly
- *Cooperation = f (Human-Machine-Communication) ➔ SEMANTIC!*

Examples for Automation Activity Layers

Level:	manipulator servo, sensor	rule based tactical	goal based strategic
Human:	muscle, sens	discern	cognition
Car:	power steering	ESP / ABS	collision avoidance
Aircraft:	cabin pressure control	flight control protections	flight management (performance indexed)
Defense:	weapon stabilizer	mission mana- gement system	low level flight assistant (goal driven)

How Complex Automation relate to Human Factors

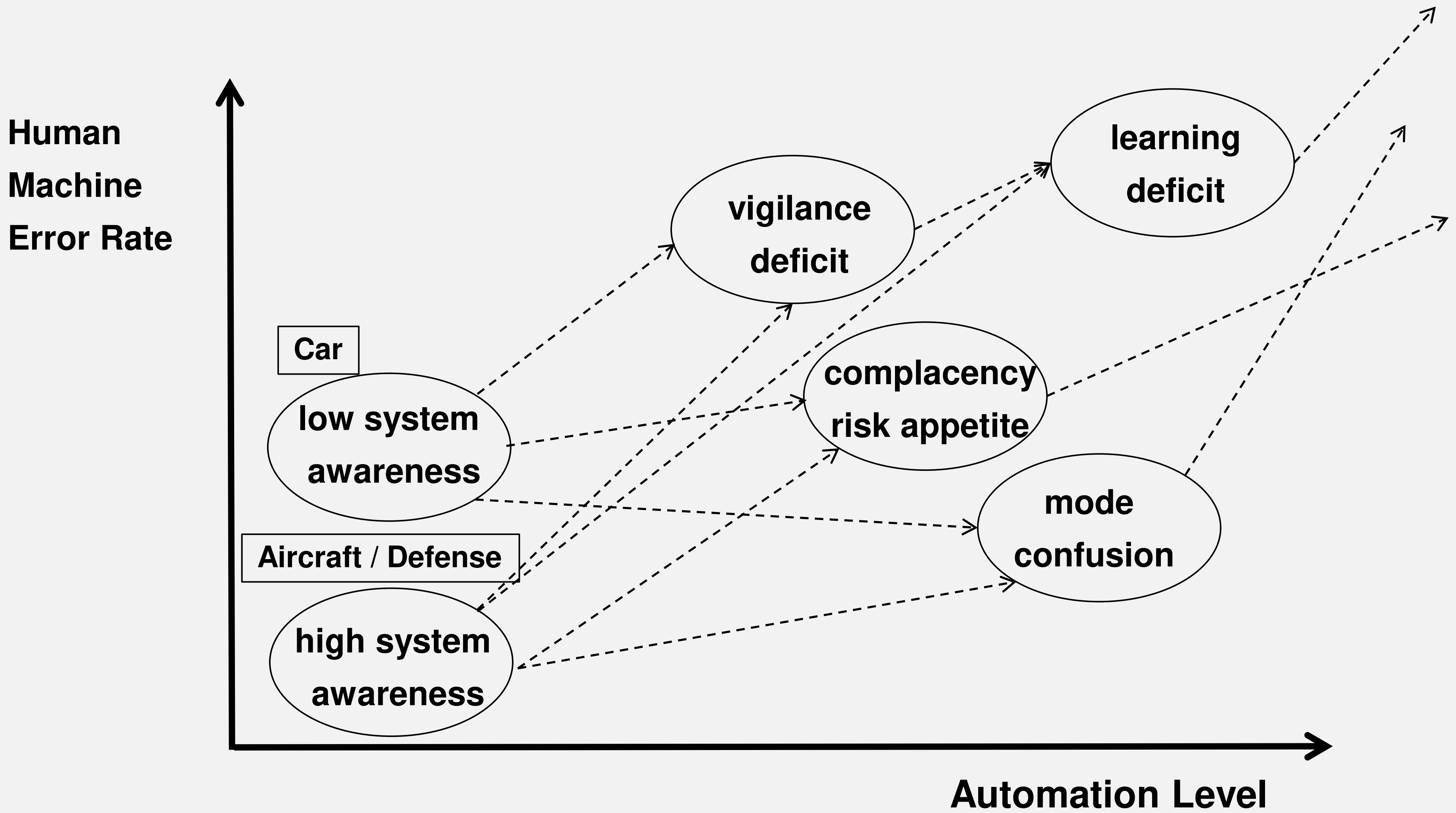
Relation between Level of Automation (Complexity) and Automation Objective



The point where the effect starts to invert or to become stagnant depends on the operator's capabilities (training), the type of automation, and the HMI compatibility.

ASES Advanced Systems Engineering Solutions

Reasons for Safety & Performance Degration



ASES Advanced Systems Engineering Solutions

Cooperative Automation as a Countermeasure

- **Mode and mode change indication**
- **Vigilance monitoring**
- **Pre-indications**
- **Protection and warnings**
- **Emergency process guidance and sequencing**

Cooperative automation demands for well trained operators, who know about the semantics of the HMI and the logic rules and the sequencing of the state machine.

- **all rules must be implemented a-priori**
- **precise a-priori situation anticipation is crucial for designers**
  - **➔Thorough validation is crucial!**
  - **➔ (for automotive guys: SOTIF)**

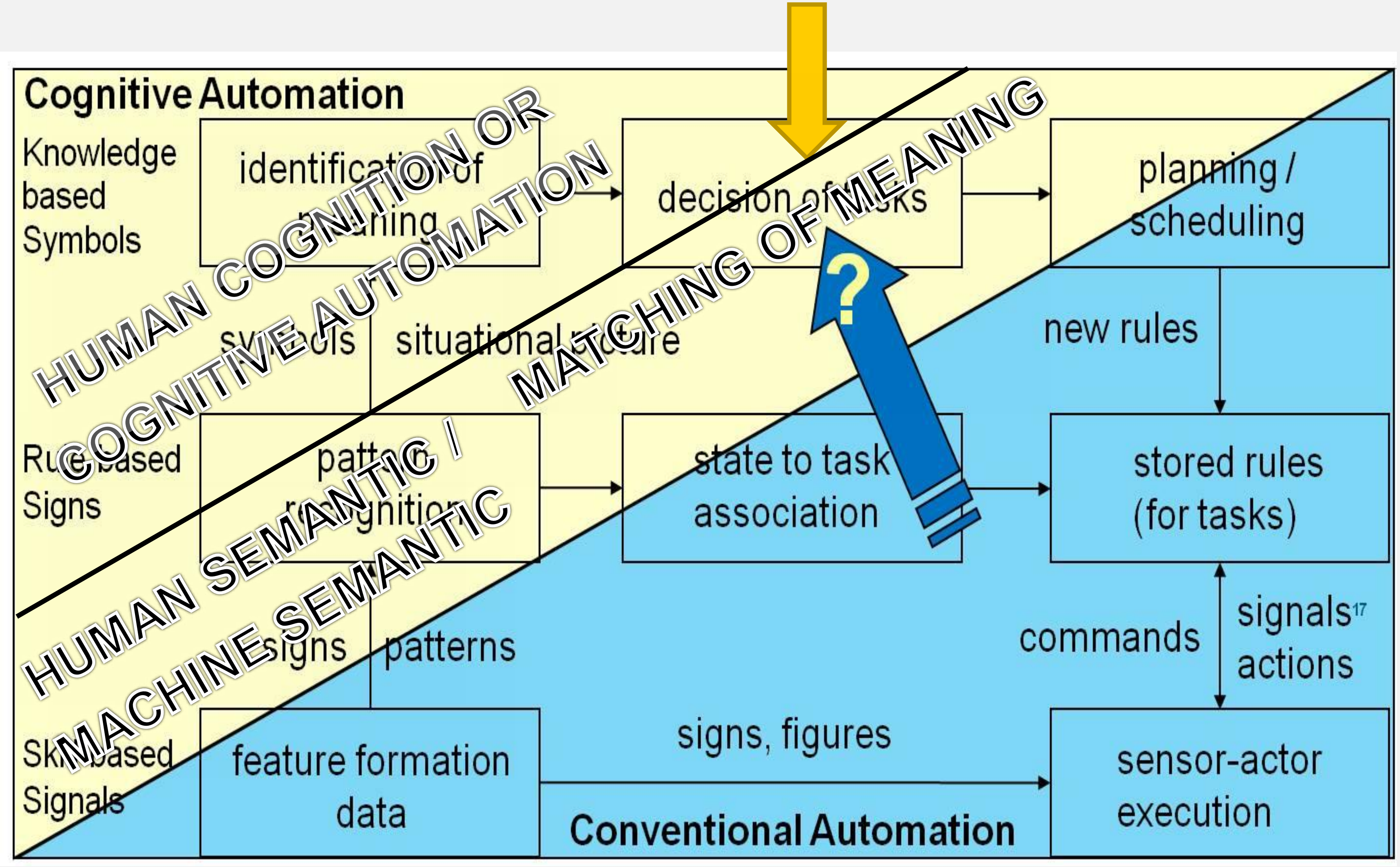
ASES Advanced Systems Engineering Solutions

AI Automation vs. Conventional Automation

- AI automation is designed to track "goals"
- The goals generate "rules" or "plans", which best (in terms of a performance index) meet the goals in reference to the actual operational situation ➔ a-posteriori generation of rules
- The rules / plans are derived from a dynamic, machine generated knowledge base (semantic "world model"), which acquires its data from plenty of non-human sensors (RADAR, LIDAR, IR, Ultra-Sound, LTE-Link, Camera etc.) or other state machines
- Thus, the "world model" can be expected to feature a non-human semantic

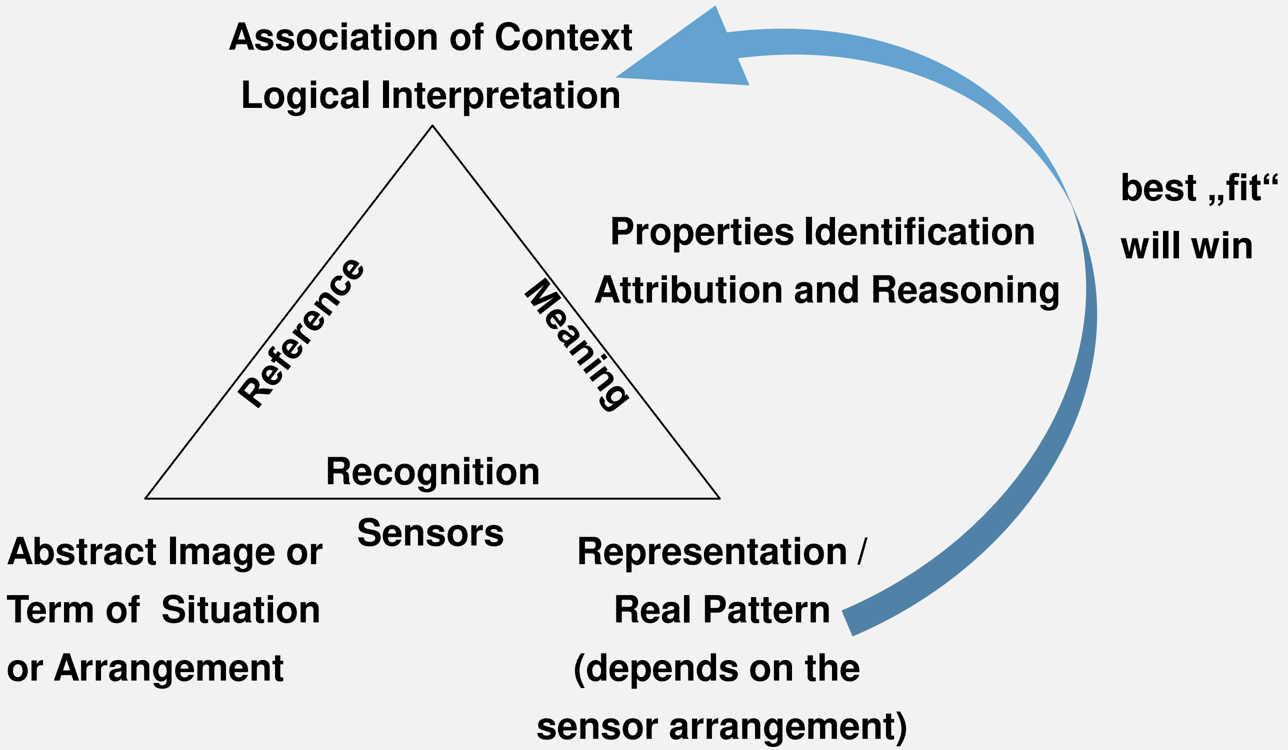
Cognitive Automation (Rasmussen's Model of human knowledge processing, 1983):

Symbols ➔ Meaning ➔ Goals ➔ Rules & Plans



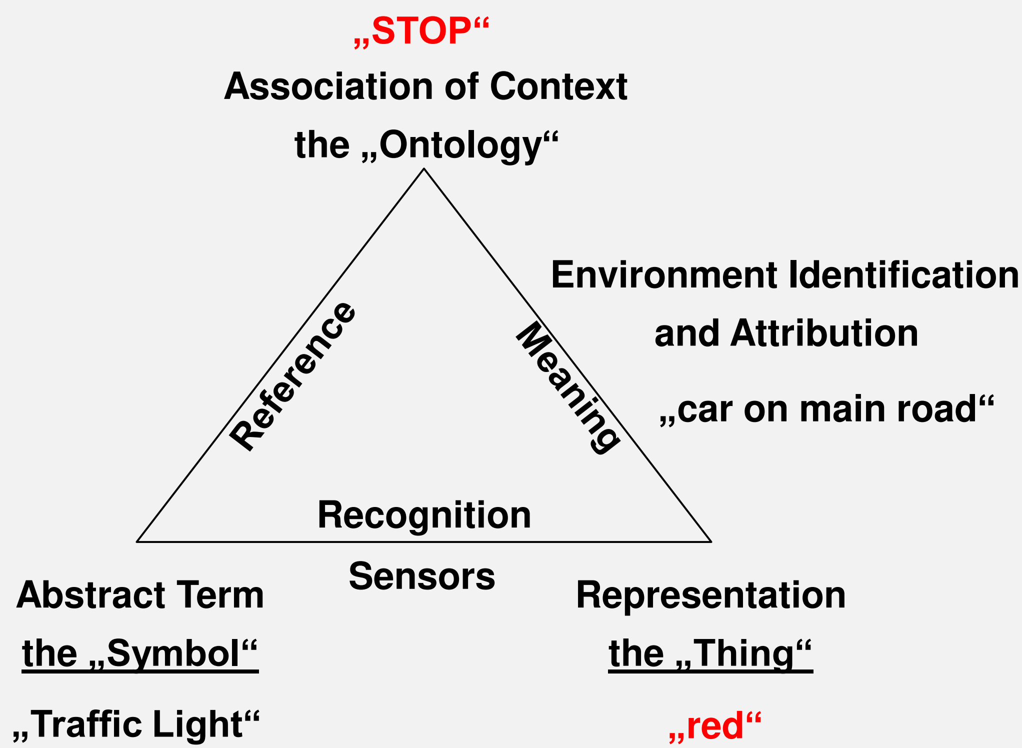
The Semiotic Triangle

Pattern recognition and interpretation



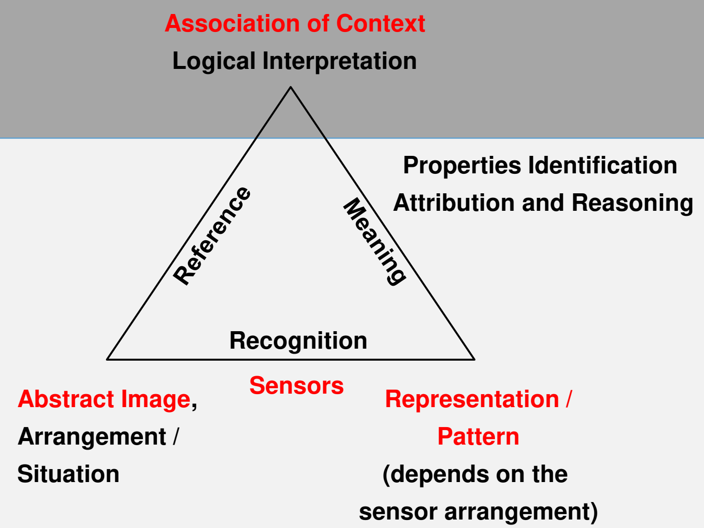
Generation of the World Model

The semiotic triangle: a simple example



ASES Advanced Systems Engineering Solutions

Evolution of the World Model



- **Over time the "Association of Context" at the top of the semiotic triangle develops to a "Model of Situations and Meaning" (Ontologies)**
- **The paradigm of this "Model" depends on the set of sensors that recognize the actual "Representation / Pattern" (right lower edge) of the "Abstract Image / Situation" (left lower edge)**
- **This "Representation" and its paradigm may be far beyond human perception**
- **As a consequence, the AI machine may recognize conflicts / risks much earlier than a human operator and may act in a way a human operator will not understand at all**
- **Human interference may impair the "mutant"**

ASES Advanced Systems Engineering Solutions

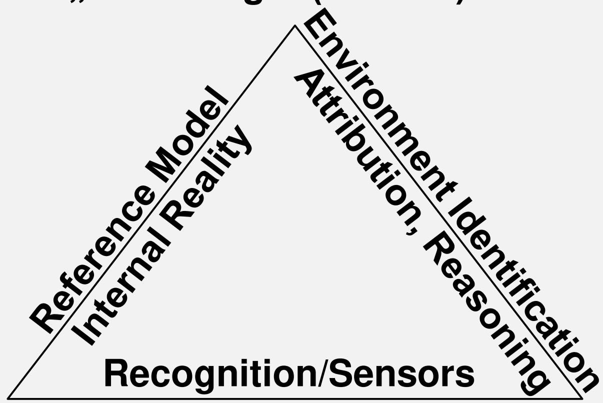
Example: Model of Situations and Meaning

Association, Meaning:

Normal: „STOP“ – „Prepare for GO“ – „GO“ – „Slow Down“

Abnormal: „Traffic Light (Control) Out of Order“

survival of the „fittest“



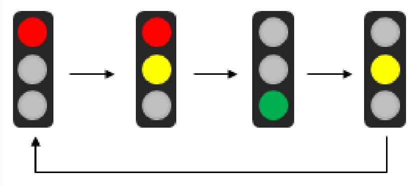
External Reality

Abstract Image, Arrangement / Situation

Traffic Light on Main Street

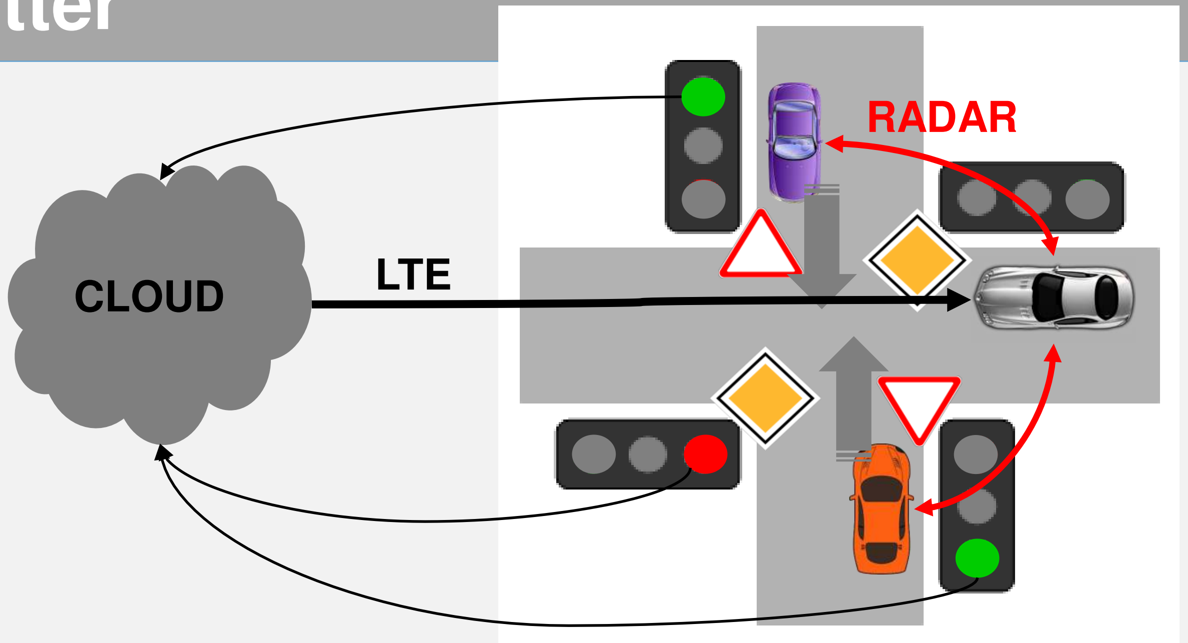
Representation / Pattern

Nominal States:



Error States:

Extension of the World Model one, which fits better



CASE: Driving on the main street, traffic light "off"

- **Trivial meaning: go ahead and cross the subordinate road**
- **Critical meaning: CAUTION! your traffic light failed – it is "RED"**
- **Extended World Model needed: <Traffic Light "Out of Order">**
- **Additional information from CLOUD or by RADAR surveillance**
- **The AI Engine got a complete picture – much better and earlier than the driver ➔ AUTO-BREAK? ➔ timely communication necessary**

Managing Conflicting Goals

GOALS

Safety	Threat Avoidance	Mission Accomplishment
• Drive slowly • Watch out • Keep maximum separation • SG-4 • SG-5 ….	• Keep clear from No-Go-Areas • Avoid pedestrian areas • Keep traffic rules • SG-x • SG-y • ….	• Drive max. (admissible) speed • Follow fastest Nav. Path • SG-III • SG-IV • ….
Collision Avoidance	Detect RADAR Traps	Catch the Train, Be On Time

Abstract Goals

Specific Goals, Plans & Rules

Conflicts !!

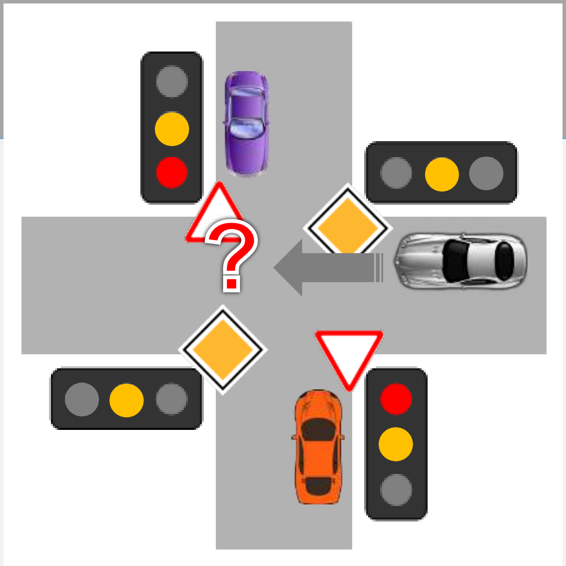
Subsequent Tasks

Traffic Jam !!! ➔ Task – Deviation ➔ Sub-Goal – Violation

Generate NEW Rules to meet Abstract Goals

ASES Advanced Systems Engineering Solutions

Linking the World Model to Goals



Arrangement: traffic light on main street switches from “green” to “yellow”

Question: how to drive reasonably?

- **Two goals: Safety & Mission Accomplishment**
- **Mission system knows about: car speed – distance car-X-ing – duration of “yellow light” phase – traffic situation behind X-ing**
- **Safety goal prevails: car prefers to slow down and stop at TL**
- **Mission Accomplishment: car accelerates in order to pass traffic light before it turns to “red”. Then decelerates to admissible speed**
- **Add a third goal – Threat Avoidance: car speeds up to a “safe” speed in order to avoid speed camera traps, but only if the TL can be passed before it turns “red” ➔ otherwise: “slow down”**
- **Here again: the “fittest” will survive**

Lessons learnt

1. A cognitive (AI) Mission Control System (AI-MCS) generates a-posteriori (own) rules, plans and control strategies
2. Those are a function of …

- An a-priori given set of goals
- The weight of each goal
- The available information of the actual situation as gained from data links and sensors, other state machines, …
- The correlation of the recognized pattern with the stored and trained semantic reference pattern of the AI-MCS “World Model”

➔ The prediction of the system is difficult / impossible. It may perform differently even in seemingly similar situations

➔ Depending on the MCS anticipation capability and semantic ontology it may feature “unexpected”, “surprising” or even “illogical” execution in terms of human perceptions and categories of the situational reality

ASES Advanced Systems Engineering Solutions

Constraint Monitoring for AI-Cognitive Systems

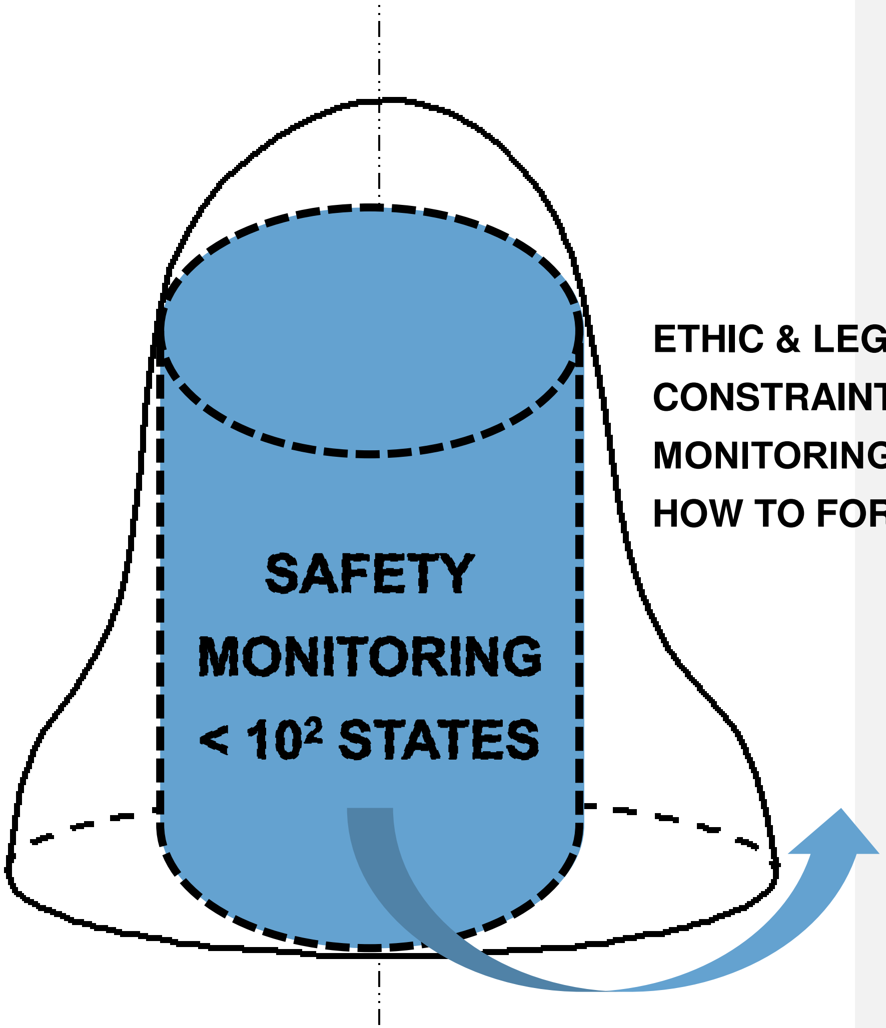
1. Many of the constraints
- Safety margins
- Material stress protection
- Security issue
- Economy and comfort

   will be monitored in the same way as for classical systems:

   "State Constraint Monitoring"

2. The challenge is
- To monitor the goal related rules and plans to stay within given limits (legal, ethical, material, performance, …), e.g. to prevent a system to turn into a "RAMBO" by putting the emphasis to the "Mission Accomplishment" goal
- To achieve "Goal Constraint Checking" with Conflicting Goals
- To find an appropriate semantic for the monitoring

The Issue with AI: Legal / Ethic Monitoring



ETHIC & LEGAL
CONSTRAINTS
MONITORING ….
HOW TO FORMALIZE ?

Self-learning ontologies may bypass the monitoring containment by application of a non-human logic

ASES Advanced Systems Engineering Solutions

How to prevent from Ethical / Moral Monitoring ?

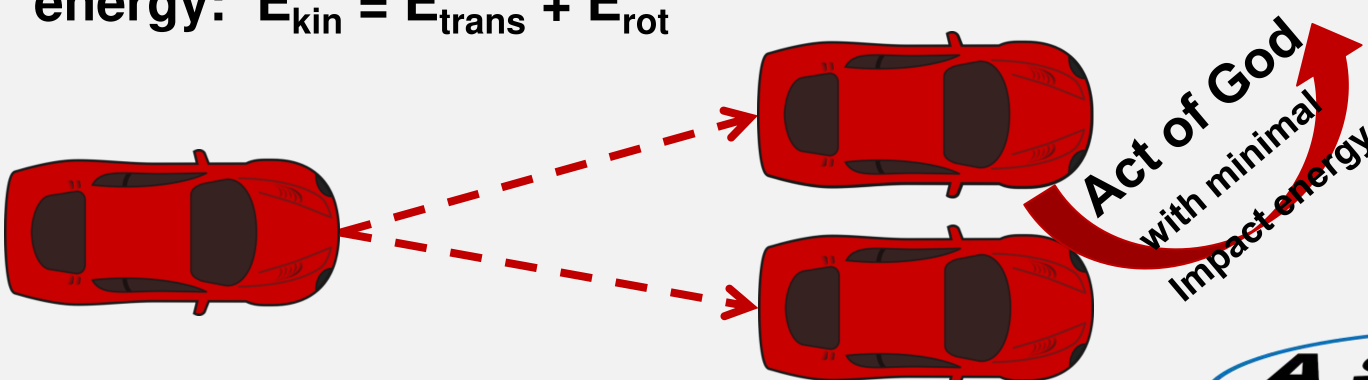
Example: fatal accident with three impact options

Child - Old Guy - Concrete Wall

Moral / Ethical Decisions generally are Not Necessary if they are Substituted by "Clever Goals":

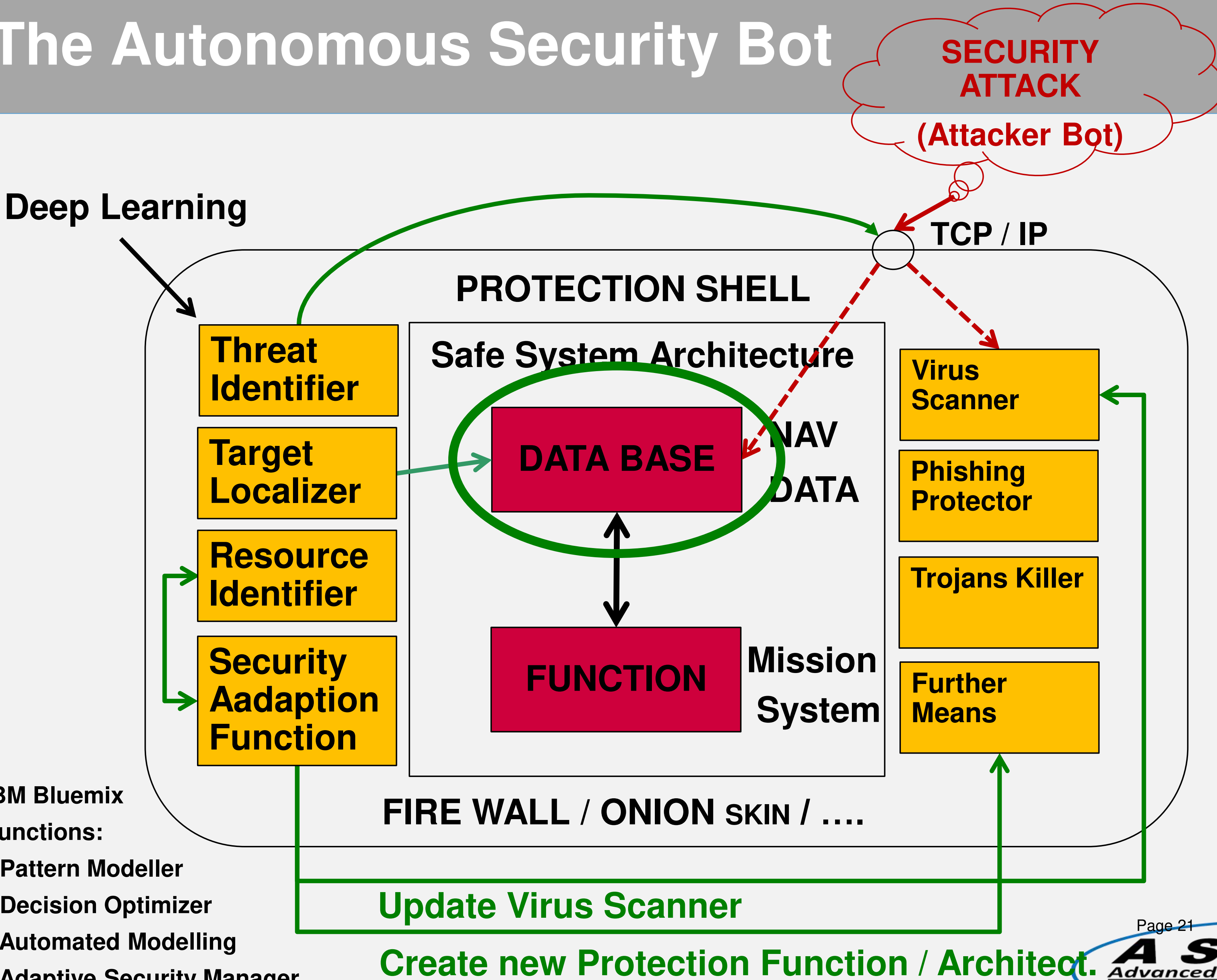
"In Case of Impact, collide with Minimal Energy"

Strategy: differential breaking ➔ convert translational into gyratory energy: $E_{kin} = E_{trans} + E_{rot}$



ASES
Advanced Systems
Engineering Solutions

The Autonomous Security Bot



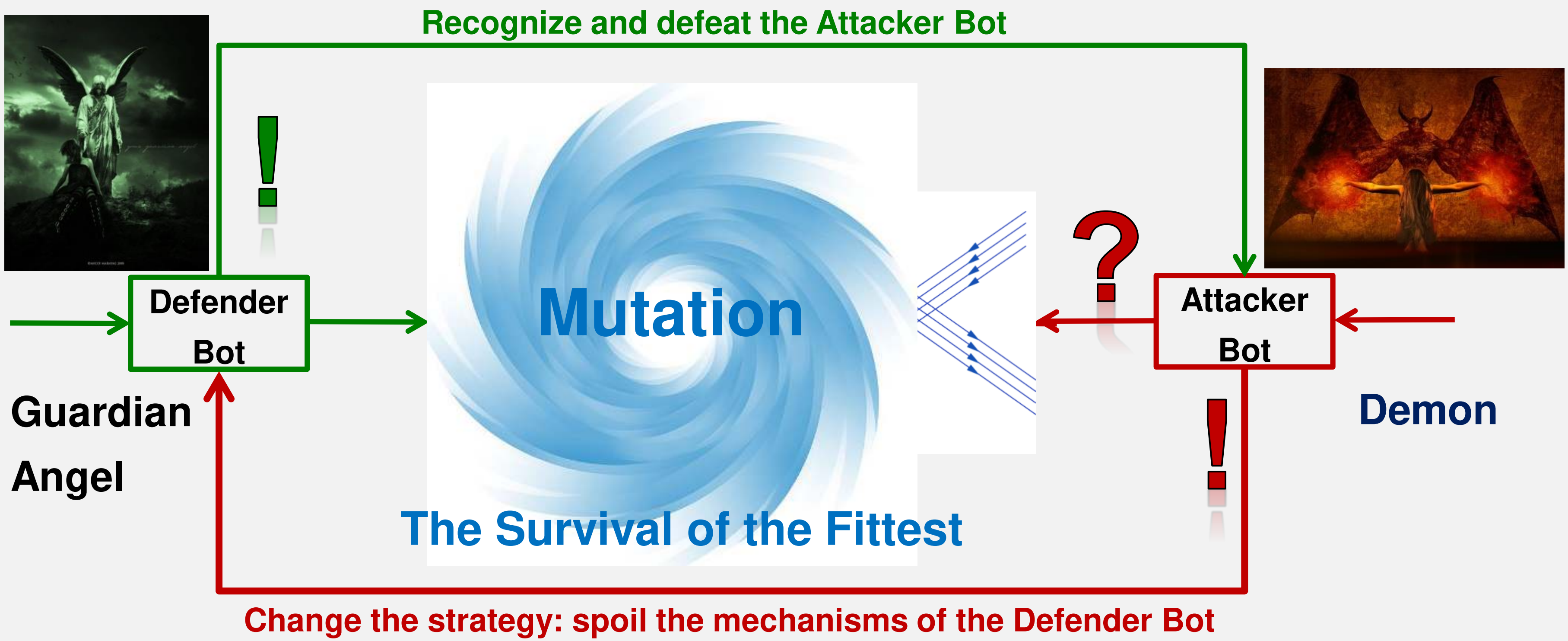
IBM Bluemix Functions:

- Pattern Modeller
- Decision Optimizer
- Automated Modelling
- Adaptive Security Manager

A S E S Advanced Systems Engineering Solutions

Mutual Evolution of Competing Attacker vs Defender Bot

Autonomous recognition of new malware patterns and adaptive generation of intelligent counter measures by an autonomous defender bot vs. counter-counter measures of the attacker bot



We neither see nor understand them, but we perceive and depend on them

Finally: A Meta-Physical Interpretation

Conclusion

- **Highly autonomous systems exclude humans from the loop**
- **Cognitive systems are “goal driven”**
- **Autonomous, self learning systems develop by mutation and selection**
- **The “world model” of a cognitive system cannot be followed up by humans (due to “non-human” sensors, ontology, evolution speed, …)**
- **This is a real challenge to mankind**

Thanks for Attention !

Questions?

mailto:henningbutz@web.de