On Benchmarking Benchmarking

Dr.-Ing. Christian Goerick
Honda Research Institute Europe GmbH

HONDA
The Power of Dreams

On Benchmarking Benchmarking

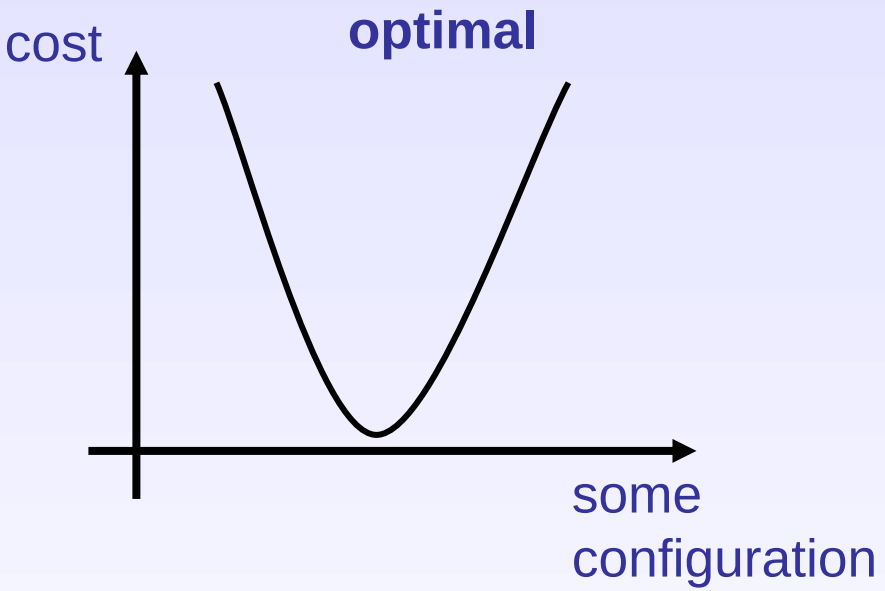
Honda Research Institute Europe

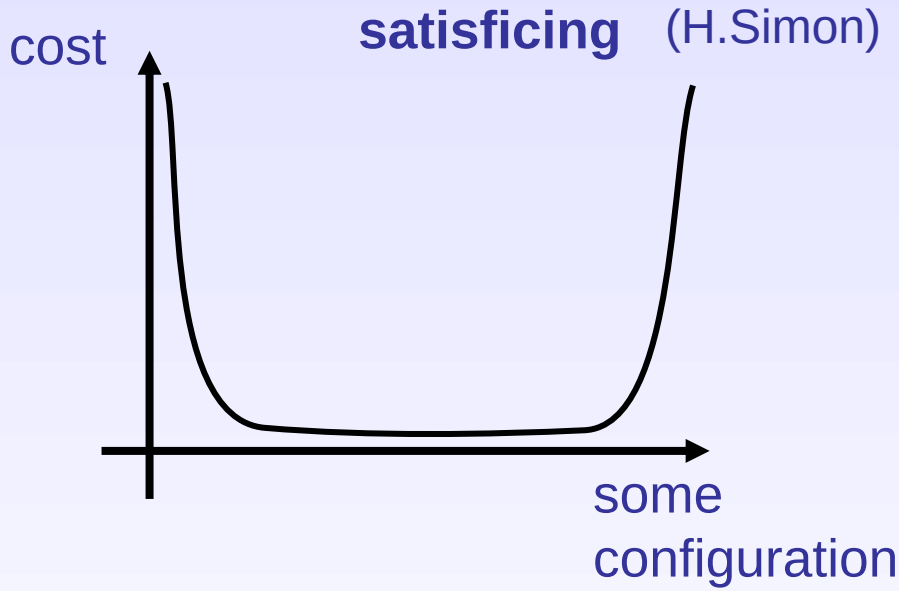
What are abstract requirements in cognitive robotics and how can they be transformed into benchmarks?

- robustness
- learning & autonomous development
- flexibility
- cognitive diversity
- openness of representations
- ... your favorite here ...

i.e. the issues we admire our own brains for ... maybe not so easy to formalize ...

Optimality vs. Satisficing





traditional engineering design:

- target for one optimal solution
- broader optimum for robust solution

natural cognitive systems:

- have viability range
  e.g. homeostasis
- allow for several solutions to one problem

Satisficing: “As long as it does it within acceptable bounds it is ok.”

Specificity vs. Generality

Honda Research Institute Europe

- Specific solution always outperform general solutions on fixed tasks (e.g. on benchmarks)
- nevertheless, general solution that can be adapted to any task are preferred

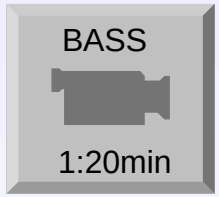
→ benchmarks should include the requirement for adapting one system to qualitative different tasks

robo soccer: changing requirements over a longer time scale with similar goal

Value of error recovery / treatment

- What kind of value do we attribute to errors the system can overcome?

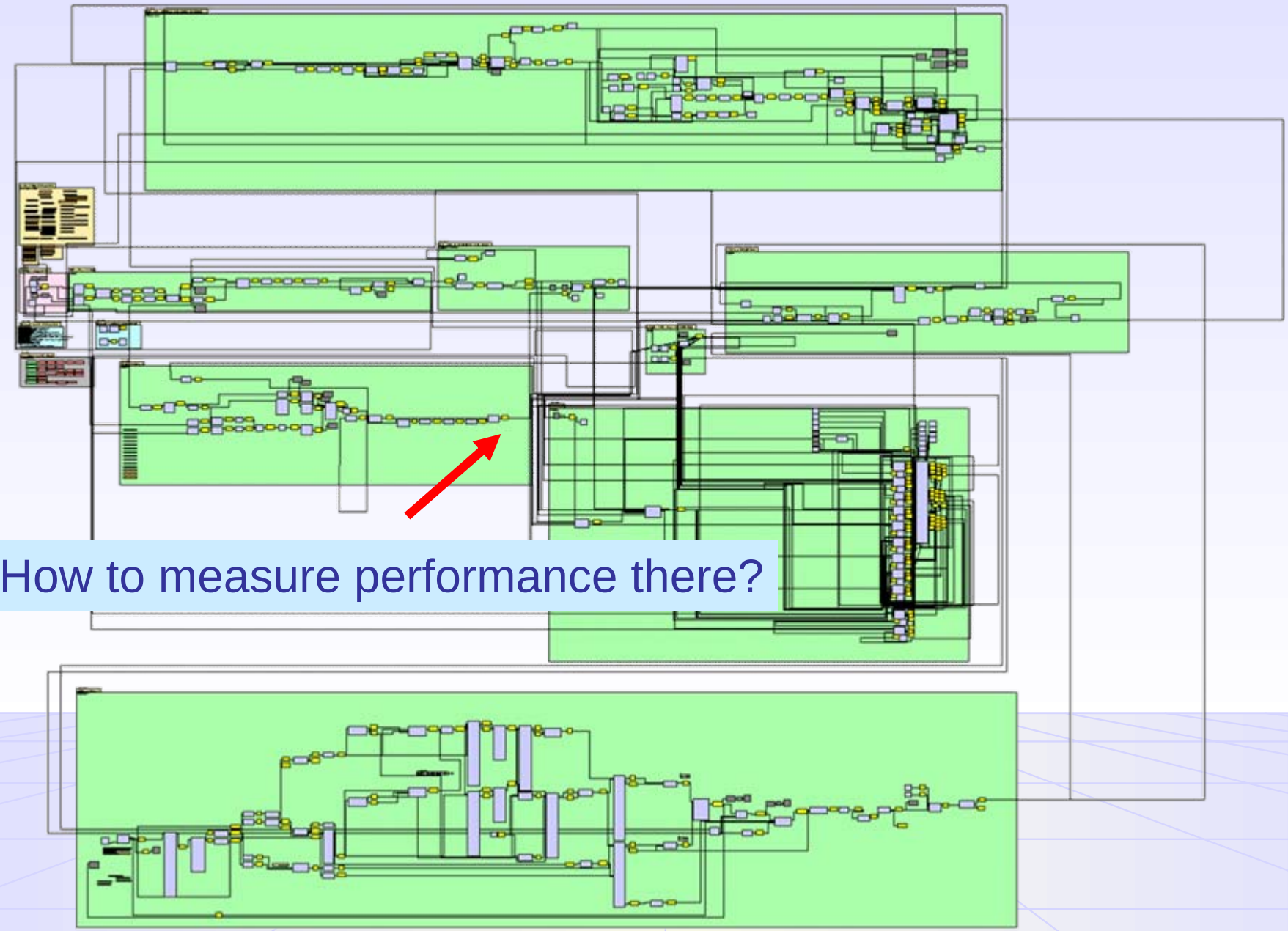
[video shows interactive error correction during online learning of visual object]



- Distinguish between
  - playful learning (where errors are necessary) and
  - hard evaluation (where errors are maybe not acceptable).
- benchmarks should account for this

Intermediate Process Benchmarking I

Honda Research Institute Europe



How to measure performance there?

Intermediate Process Benchmarking II

Honda Research Institute Europe

- Independent optimization of intermediate results may not contribute to a good overall performance

  → classical credit assignment problem

  → build and evaluate complete systems

- intermediate benchmarking must be congruent with the overall processing

  example: statistical covering of input and optimizing reconstruction error maybe unsuited for solving any tasks

→ intermediate benchmarking is maybe necessary but never sufficient

Static vs. Dynamic Benchmarks

Honda Research Institute Europe

- **static benchmarks** := system does not influence sensory data, e.g. image databases, good for papers or baseline comparison
- **dynamic benchmarks** := system influences the generation of the sensory data, e.g. active camera on humanoid robot, interactive scenarios, closed loop systems
- Dynamic benchmarks may include situations where the predictions of the system fail
  → interesting case, learning starts here
- good static performance does not guarantee a good dynamic performance

→ target dynamic interaction as early as possible

Temporal coherence of systems I

- -- or – do we always want to start from scratch?
- For one specific scenario / task infinitely many systems can be build (e.g. look at Alife wrt. locomotion, result from theory of nonlinear dynamical system)

→ sequence of scenarios itself does not lead to temporally congruent series of incremental systems

Temporal coherence of systems II

Honda Research Institute Europe

→ three different views, related, each with a temporal sequence

- **Method oriented:** quality of the internal methods, e.g. vision recognition, motion control
- **Skill oriented:** capabilities of the overall system, observable, independent of context
  e.g. inspection and manipulation of proto-objects
- **Scenario oriented:** context to show a capability
  e.g. learn what my cup looks like and fetch it later from the cupboard

→ Benchmark is often just one view to a system (e.g. scenario),

neglecting the other views may lead to an inferior system

Natural benchmarks?

Honda Research Institute Europe

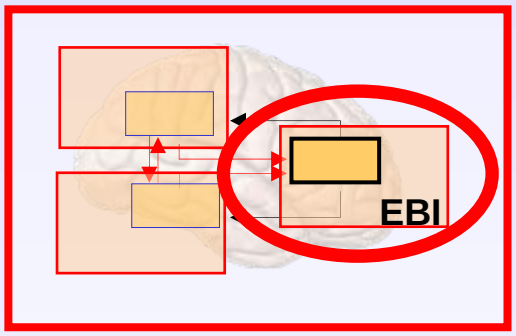
- Since we are targeting for cognitive systems, aren't there "benchmarks" for natural cognitive systems?
- Screening tests for development of children?
  - focus primary on sensing and development of the body, cognitive skills are focused on later
  - time line is not clear, variance of occurrence of a specific skill is in the order time of the occurrence itself
  - e.g. speaking starts between 12 month and 24 month, walking starts between 10 month and 24 month
- Psychology
  - I.Q. tests etc.
  - currently far beyond abilities of technical systems

What do we do at HRI-EU?

- pursue analysis by synthesis approach for understanding real cognitive systems
- follow the biological paradigm in various levels
- check system performance in skill level
  - Method oriented: quality of the internal methods
  - Skill oriented: capabilities of the system, observable, independent of context
  - Scenario oriented: context to show a capability

EBI - Embodied Brain-Like Intelligence

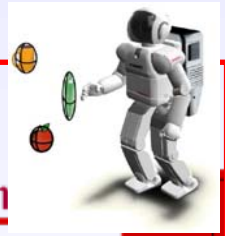
Honda Research Institute Europe



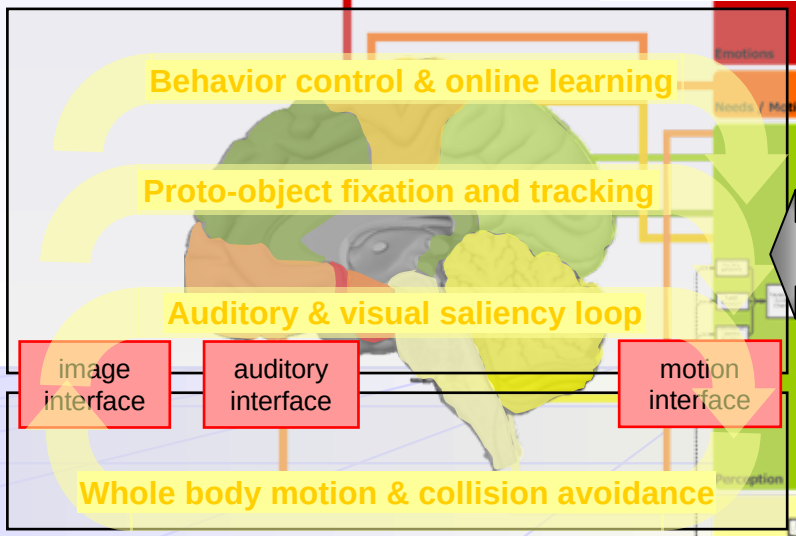
Objectives

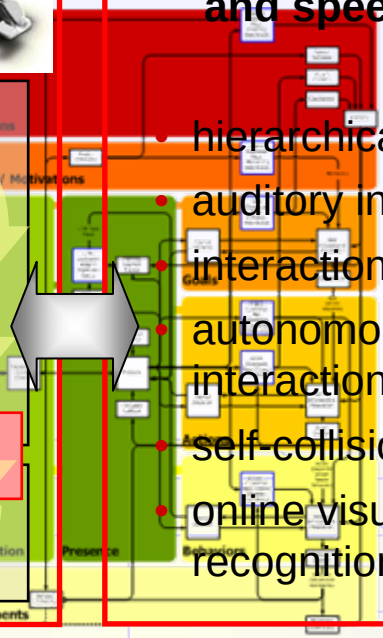
- develop individual cognitive abilities based on practical experience
- evolve system from initially few innate abilities towards an autonomous compliant partner

- **Autonomous Learning & Interaction System (ALIS) on ASIMO**



Building a coherent system





- **Integrated system for vision, behavior and speech I/O comprising**
  - hierarchical processing architecture
  - auditory interaction initiation
  - interaction related object representation
  - autonomous behavior generation and interaction
  - self-collision free online whole body motion
  - online visual object learning and recognition

Episodes 2:00 | NoCut 2:42 | Internal 5:58 | External 4:45

Thank you for sharing our thoughts