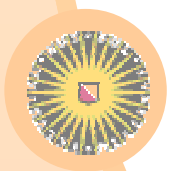
Training Robots

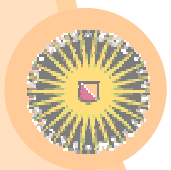
based on

1. YAMADA Seiji, YAMAGUCHI Tomohiro **Training AIBO like a Dog**
2. Frederic Kaplan, Pierre-Yves Oudeyer, Eniko Kubinyi,Adam Miklosi, **Robotic clicker training**



Overview

1. Learning
2. Teaching methods
3. Cues and conditioning
4. Clicker training
5. Robotic clicker training
6. Social robots and adaptation
7. Learning actions
8. Discussion

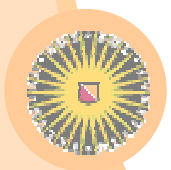


Learning

- Learn ontology/names
- Learn actions
- Learn to perform actions
- Learn social cues
- Supervised learning, reinforcement learning, …

We need to be specific about what exactly we want the robot to learn.

The domain determines the type of learning techniques available and appropriate



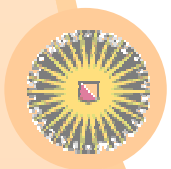
Learning new behaviour

What is needed?

- Combinable basic behaviours
- Explorative behaviour
- “Reinforcement” mechanism

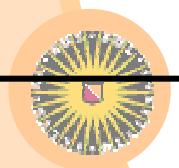
Why teaching and not programming?

- Users don’t know how to program
- Teaching behaviour is fun by itself



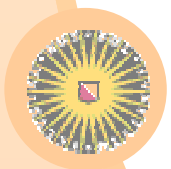
Teaching Methods

Training Techniques	Sequences of Actions	Unusual Actions	Usability with Animals	Usability for Autonomous robots
Modelling	No	Difficult	Seldom used	Difficult
Luring	Difficult	Difficult	Good for simple actions	Seldom used
Capturing	No	No	Good	Good
Imitating	Yes	Yes	Seldom used	Difficult
Shaping	Yes	Yes	Very good	Seldom used



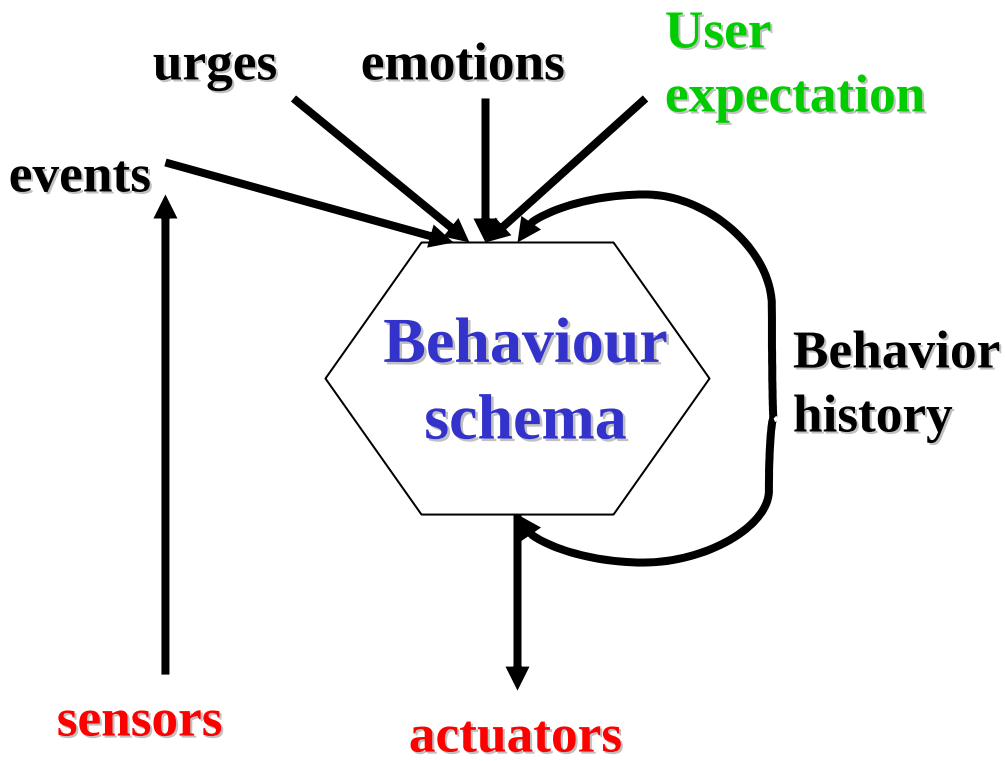
Clicker training

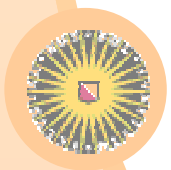
1. Charging up the clicker: associate the click with a reward
2. Getting the behavior: click on each movement in right direction, give reward after full behavior
3. Adding the command word: trainer says the word just before or after the behavior
4. Testing the behavior: Refine and test in different situations. Rewards only given after full performance.



Robotic clicker training

Control architecture of behaviour schemas





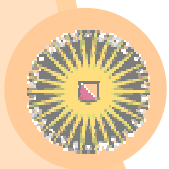
Learning secondary reinforcers

Needed

- Primary reinforcer
- Association between reinforcer and secondary reinforcer

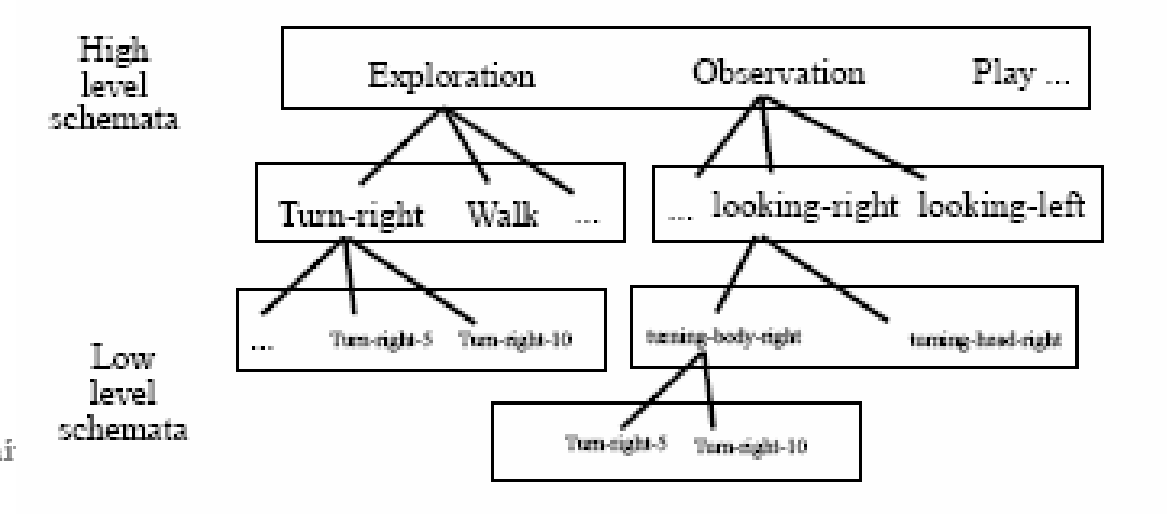
Implemented:

1. Primary reinforcer is fixed. E.g. pat head sensor
2. Associate other input by proximity in time >30 times. E.g. say “good”
3. After association is fixed robot indicates association by feedback (e.g. wagging tail) after secondary reinforcer input



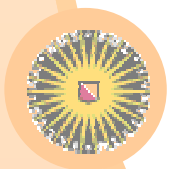
Guiding the robot

- Indicate start of “training”.
  - The command “try” activates the “user expectation module” to start making a model of behavior that the user wants
- Reinforcement of exhibited behavior inhibits behavior schemas. Consequent secondary reinforcement leads to a final complete expected behavior that is rewarded.



Adding the command word

- After hearing a word and having a new behavior in the user expectation module a new temporary schema is created (the exact sequence of actions in the schema might be too short or long)
- The temporary schema is part of a “please user” schema
- The activation trigger is the word heard
- The schema becomes permanent if the schema is used often (reinforced)



Testing the new schema

- Upon hearing the command perform the new schema.
- If no reward is given, mutate the schema and apply new schema next time command is given
- Repeat until reward is given and schema is correct.

Example:

```
Later, TRAINER says "spin"
ROBOT performs the sequence
(WALK-4, TURN-LEFT-3, TURN-RIGHT-5, TURN-RIGHT-10, TURN-RIGHT-10,
TURN-RIGHT-10, TURN-RIGHT-10)
It walks forward, makes 3 steps in the left direction
and then moves to the right in a chaotic manner.
TRAINER says nothing and after a while repeats "spin"
ROBOT performs the modified sequence (WALK-4, TURN-RIGHT-15
TURN-RIGHT-10, TURN-RIGHT-10, TURN-RIGHT-10), etc.
```

Test the clicker training with naïve users

Difficulties:

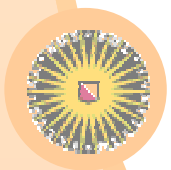
1. **Speech recognition!**
2. **Unexpected behavior due to design of the behavior tree**

Mutual adaptation

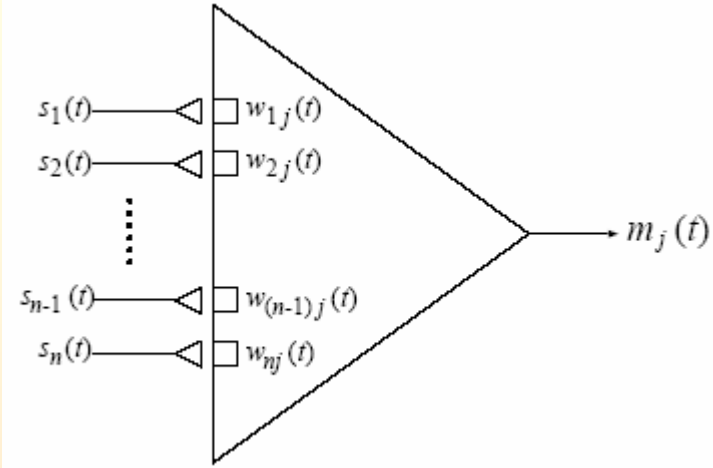
Test:

The human has to teach an AIBO to move forward, then sit and then move backward, by using only the three buttons on the back.

No indication is given which button should be used for what purpose. This can be developed different by each user.



Computational model of classical conditioning



Klopf neuron model

$$m_j(t) = \sum_{i=1}^{n} w_{ij}(t)s_i(t) \quad (1)$$

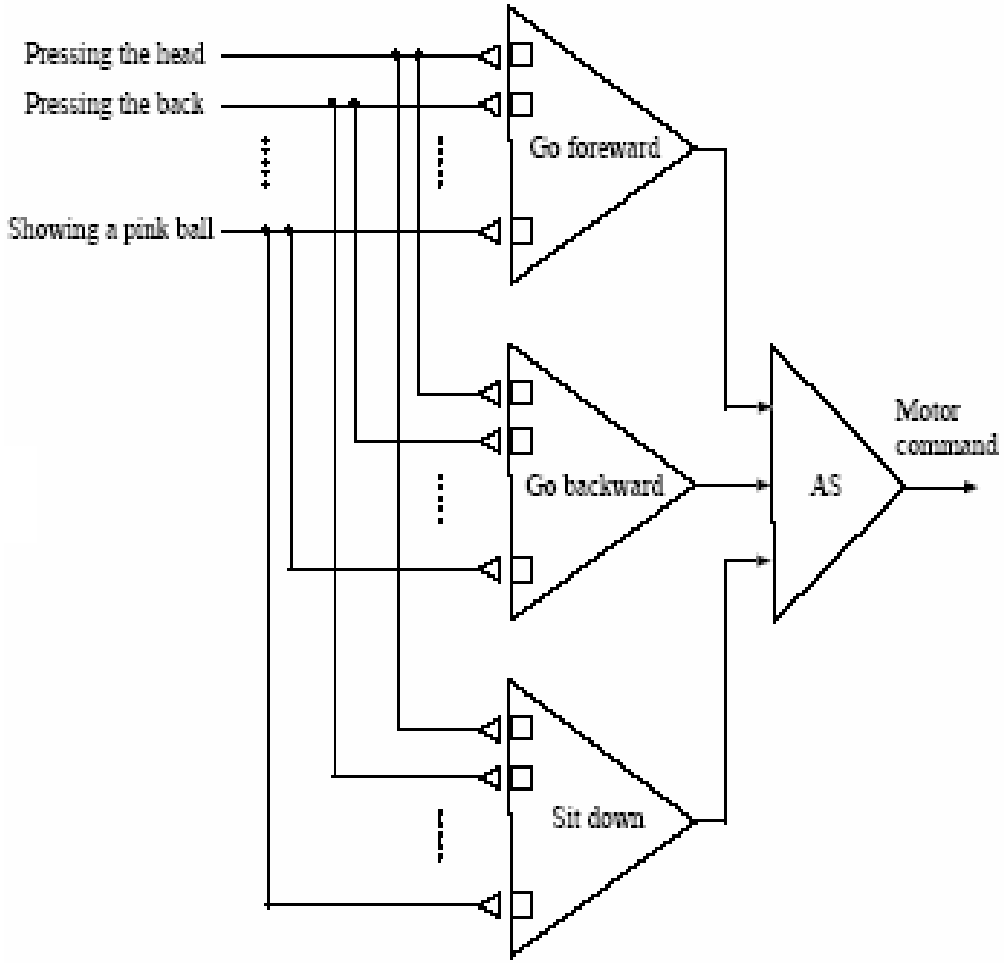
$$w_{ij}(t) = w_{ij}(t-1) + \Delta w_{ij}(t) \quad (2)$$

$$\Delta w_{ij}(t) = \Delta m_j(t) \sum_{k=1}^{\tau} c_k |w_{ij}(t-k)| \Delta s_i(t-k) \quad (3)$$

$$\Delta w_{ij}(t) = w_{ij}(t+1) - w_{ij}(t) \quad (4)$$

$$\Delta m_j(t) = m_j(t) - m_j(t-1) \quad (5)$$

$$\Delta s_i(t-k) = s_i(t-k) - s_i(t-k-1) \quad (6)$$



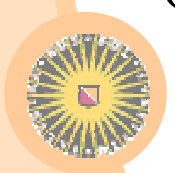
Sensors used

1. Pressing the head
2. Pressing the back
3. Pressing the chest
4. Pressing the chin
5. Showing pink ball

Human needs to find out which are the primary reinforcers (unconditional stimuli)

Human needs to search for the right secondary reinforcers (conditioned stimuli)

(the three buttons on the back).



Program the AIBO like a dog

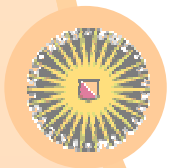
Explore the similarities between AIBO's and real dogs. Users will tend to search the US in the same areas as for real dogs.

1. If AIBO sees pink ball it goes forward
2. If human presses back it will sit down
3. If human presses chest it will go backward


Pink ball

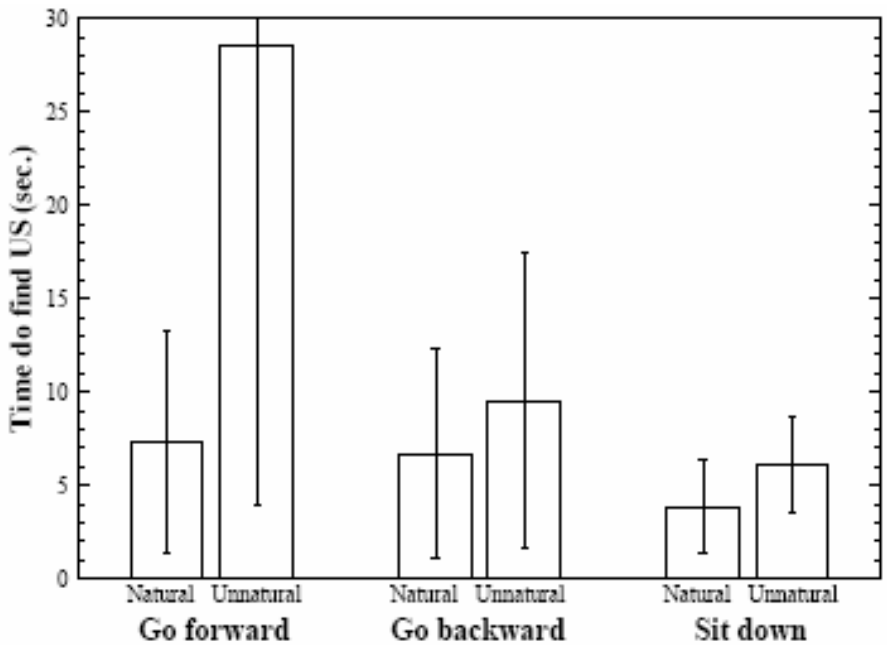

Press back

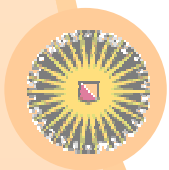

Press chest



Experiment

- Natural vs. unnatural behavior on stimuli
  - Unnatural:
    - AIBO moves back on seeing pink ball
    - AIBO sits down on pressing head
    - AIBO moves forward if pressing chest
- Results:



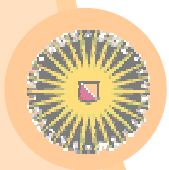


Now condition the AIBO

Explain the clicker training

Let subjects train AIBO

Results: “promising”



Discussion

- Who adapts to who?
- Is clicker training feasible for training more behaviours?
- Is it useful for training goal directed behavior?
- When to train and when to program?
- Combining behaviours?