Combining Reflexes and Reinforcement Learning in Evolving Ecosystems for Artificial Animals

Victor Skoglund, Hans Glimmerfors

Supervisor: Claes Strannegård
Examiner: Torbjörn Lundh

2021-06-15

Introduction

- Simulating ecosystems with sexual reproduction, inheritance and evolution
- Part of the Ecotwin project
    - Lotka-Volterra Population Dynamics
    - Homeostatic Regulation
    - **Reproduction and evolution**

CHALMERS
UNIVERSITY OF TECHNOLOGY

Goals

- Investigate the need for evolution <u>and</u> learning in ecosystems
  - Reflexes are created by evolution
  - Learning is done by reinforcement learning
- Is asexual- and sexual reproduction more advantageous for survival in some environments?
- Is there an evolutionary purpose to death?

Simulating Ecosystems

- Historically ecosystems were analyzed mathematically
  - Using differential equations
    - Lotka-Volterra
  - Handle the species on a population-wide level
- Problems
  - Many assumptions made
  - No notion of individuals

$$\begin{cases} \dfrac{dx}{dt} = \alpha x - \beta xy \\[2ex] \dfrac{dy}{dt} = \delta xy - \gamma y \end{cases}$$

Simulating Ecosystems

- With reinforcement learning we can simulate intelligent individuals
- The hope is to create similar population-wide behaviours
    - But since it is based on individuals, it is more realistic

+

=

?

CHALMERS
UNIVERSITY OF TECHNOLOGY

Classical Predator-Prey Systems

- Predator eats prey
- Predator only dies due to starvation
- Prey only dies due to being eaten
- Predator reproduces % of time when consuming a prey
- Prey reproduces % continuously when not eaten
- Only mathematical, thus we only know the ratios of species

- Difficult and time-consuming to find parameters

$$\begin{cases} \dfrac{dx}{dt} = \alpha x - \beta xy \\ \\ \dfrac{dy}{dt} = \delta xy - \gamma y \end{cases}$$

CHALMERS
UNIVERSITY OF TECHNOLOGY

Multiagent-based Predator-Prey

- Predator feeds on prey
- Prey feeds on plants (or nothing)
  - Various implementations tried previously
  - In our model prey must eat
- Animats die from starvation, being eaten, or reaching a maximum age

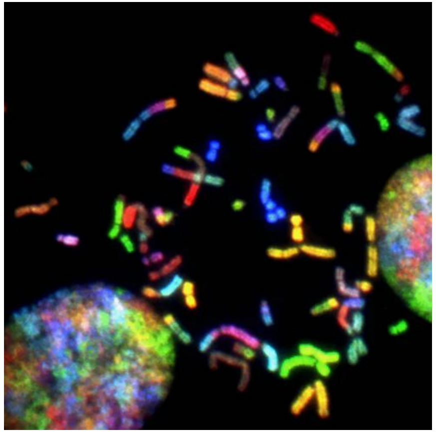
Multiagent-based Predator-Prey

- Animats reproduce
  - Automatically (done previously)
  - Consciously (our model)
- Species representation
  - Prey
    - Static
    - Agent (e.g. our model)
  - Predator
    - Agent

CHALMERS
UNIVERSITY OF TECHNOLOGY

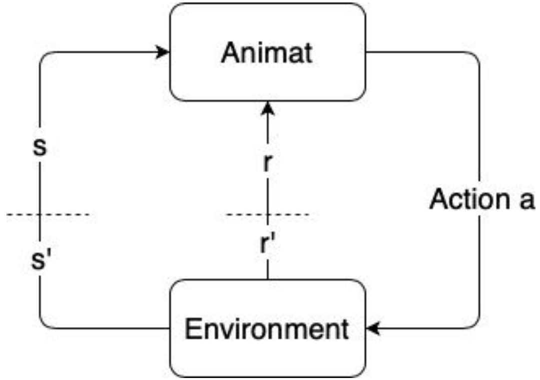
Animats

- Artificial animals (“ANImal” + “MATerials”)
- Based on biology
  - Decisions made by nervous systems
  - Observations and actions dependent on inheritable genes
- Adapted for computer science
  - Decisions made by neural networks
  - Genes represented as floating-point numbers
    - Combine genes of parents for inheritance
    - Operations possible: sum/product/max/min/etc.
      (we use one of the parents’ genes)



Reinforcement Learning

- Learn through rewards when achieving goals
- Try to find the actions maximizing the rewards
- Two choices:
  - *Exploit* policy for instant reward
  - *Explore* policy for possibly greater future rewards

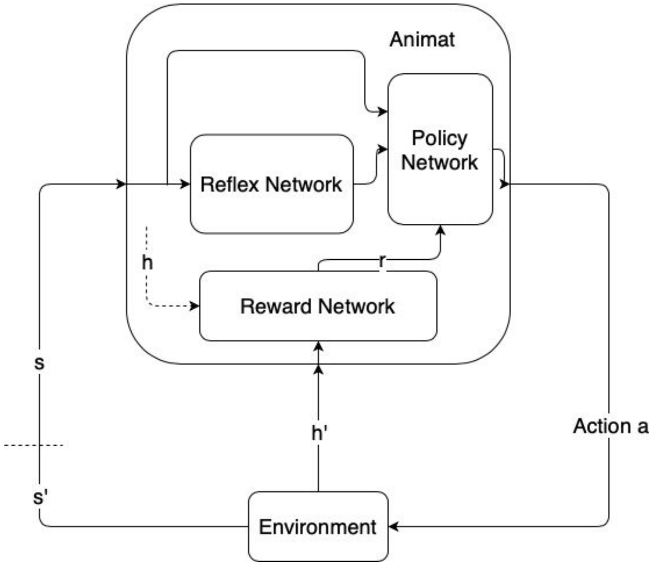


CHALMERS
UNIVERSITY OF TECHNOLOGY

Cognition

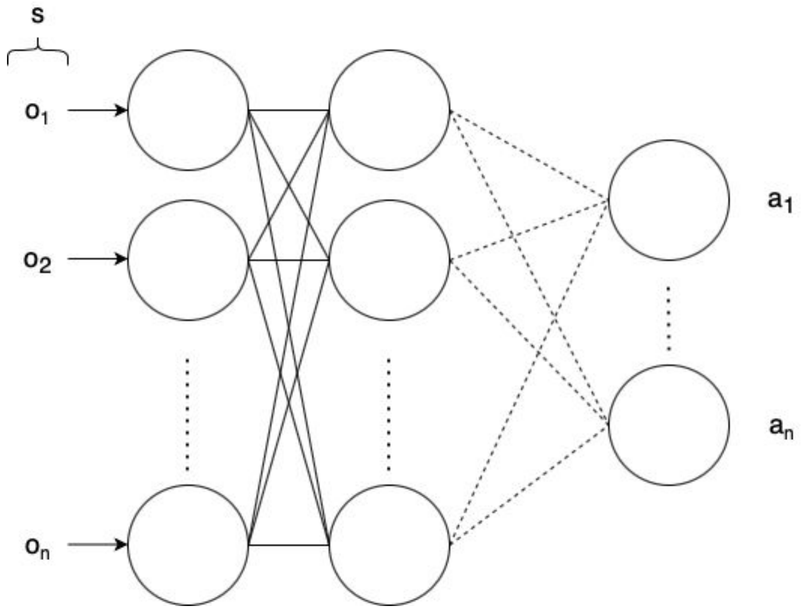
- Corresponds to nervous systems in animals

- Policy network
- Reward network
- Reflex network
- Prediction network (future research)



Policy Network

- Set of observations are sent to a neural network
- The neural network predicts the action in the observed state with the highest reward
- By adjusting the weights and biases for each node the animat is able to find better predictions

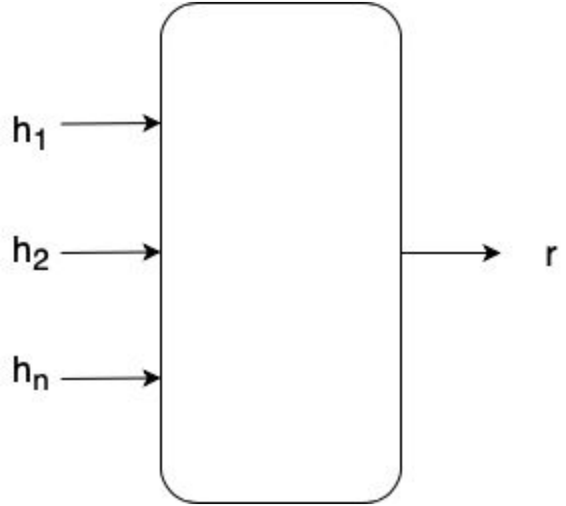


Reward Network

- Based on an implementation presented by Kleve and Ferrari

$$happiness_t(H_t) = \prod_{h \in H} (a_h + w_h u_h(h_t))$$

- Reward given in every step based on happiness



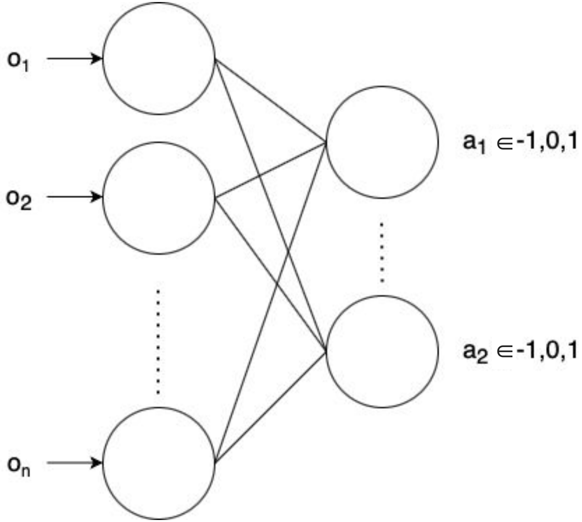
Reflexes

Animats

- Innate behaviors
  - Pecking behavior in gull hatchlings
  - Turtle hatchlings journey to the sea
- Reflex actions
  - Knee-jerk reflex
  - Diving reflex

Reflex Network

- Hardwires a stimuli to a certain action
- A reflex may:
  - Forbid a certain action
  - Force a certain action
- If the policy network chose a forbidden action:
  - The second best action is chosen
- If an action is forced:
  - The policy network's choice is ignored

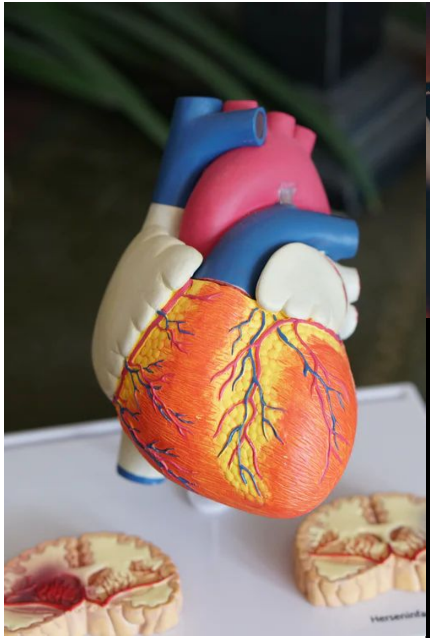


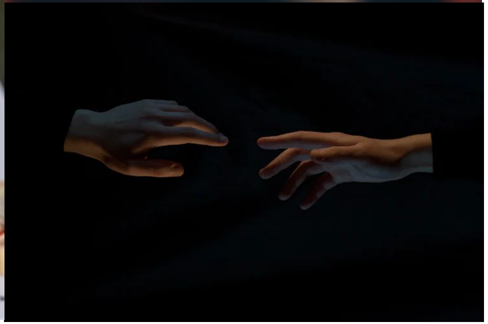
CHALMERS
UNIVERSITY OF TECHNOLOGY

Observations - Senses

- Internal/Homeostatic Senses
- Vision
- Smell
- Touch

- Inform animat of the current state





Homeostatic Senses

- Energy
- Libido

$$happiness_t(H_t) = \prod_{h \in H} (a_h + w_h u_h(h_t))$$

- Energy is a critical sense: if it is not managed correctly, the animat dies
  - Happiness from energy is maximized by increasing energy
- Libido affects only happiness: controlled by the animats reproductive cycle
  - Happiness from libido is maximized by decreasing libido

Energy

$$C = M + R + G + E$$

C = Energy Consumption

M = Maintenance Cost

R = Reproduction Cost

G = Growth Cost

E = Energy Stored

Energy

$$C = M + R + G + E$$

Animats

Maintenance cost:

- BMR + Movement Cost
- Simulates BMR: energy depletes over time
- Kleiber's Law: BMR = $k \cdot m^{3/4}$
- Cost for moving corresponds to the work performed

CHALMERS
UNIVERSITY OF TECHNOLOGY

Energy

$$C = M + R + G + E$$

Reproduction Cost:

- Small constant cost for mating
- Energy cost for giving birth corresponds to the difference in weight between the mother and the offspring

Energy

$$C = M + R + G + E$$

Growth Cost:

- Growth cost corresponds to the change in mass between each time step
- Animats grow linearly until their maturity age

Energy

$$C = M + R + G + E$$

Animats

Energy Stored:

- All animats can use all of their energy at any time
- Can store a fixed amount of energy

Energy

$$C = M + R + G + E$$

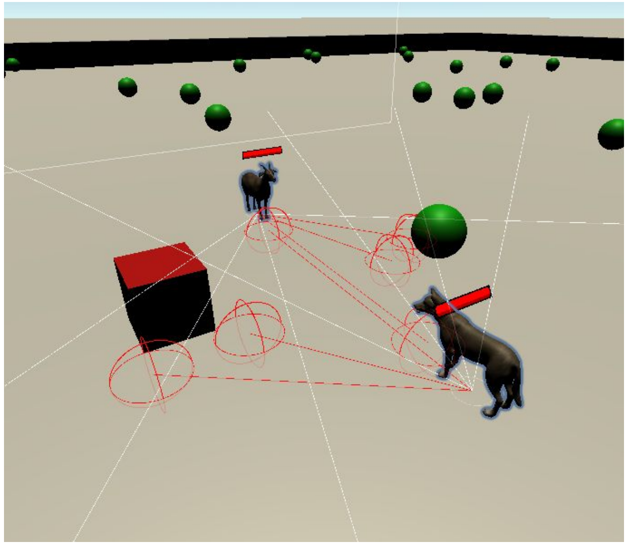
Animats

Energy Acquisition:

- Energy gained by eating food
- Each food have a specific nutritional value
- Food can either fill or drain an animats energy depending on if it is toxic

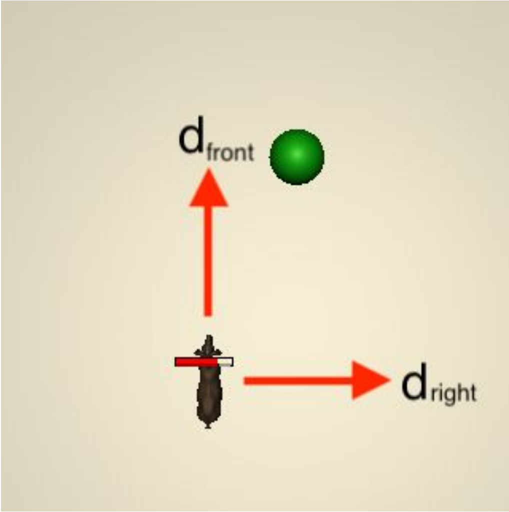
Vision

- Simulated using raycasts
- Each animat has a set of rays spanning in a 108º angle
  - Goats: 26 rays, Wolves: 13 rays
- Each ray registers:
  - Type of object
  - Distance to object
  - Color of the object
- Number of rays heavily affects performance



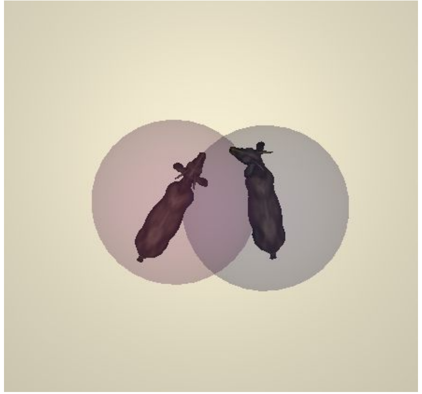
Smell

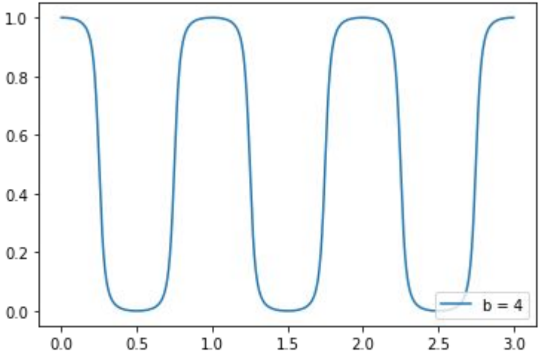
- Animats can smell certain objects within a distance r_{smell}
- For each object type within reach, the distance and direction to each object is calculated and normalized
  - (The normalized vector divided by the square of its magnitude)
- The animat perceives the sum of each vector normalized vertically and horizontally to the animat's direction



Reproduction

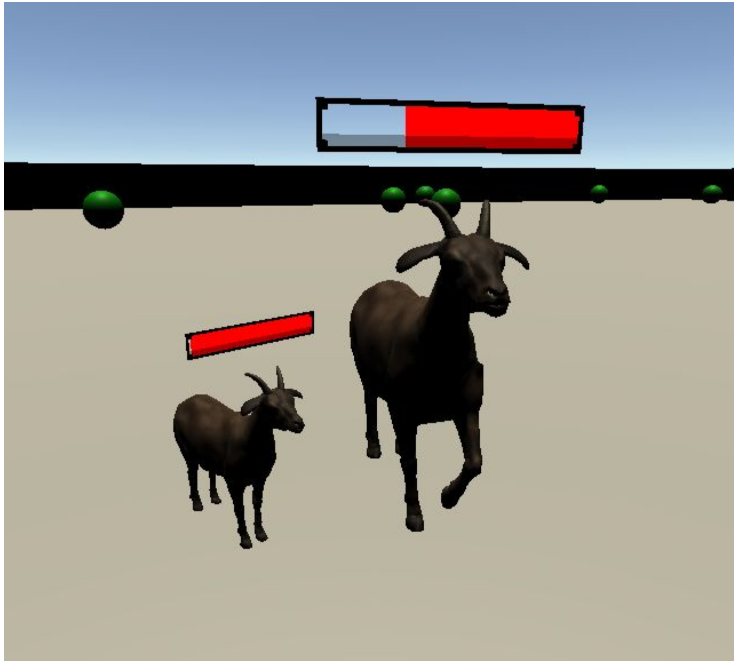
- Sexual Reproduction:
  - Male and female must choose “Reproduce” action
  - Female’s fertility depends on the mating season
- Libido varies depending on season
  - Animats only want to mate when their libido is high
- After creating offspring
  - Libido is lowered
    - Increase in happiness
  - Female is infertile while recovering
    - Represents pregnancy
    - But birth is instant





Reproduction

- Offspring are born small
  - Move slower, eat less, die easier
  - Consume less energy
  - Need to grow before becoming fertile
- Only animats of same species can reproduce



Action Space

- Idle
- Eat
- Accelerate
- Decelerate
- Turn Left
- Turn Right
- Attack (available only to predator)
- Reproduce

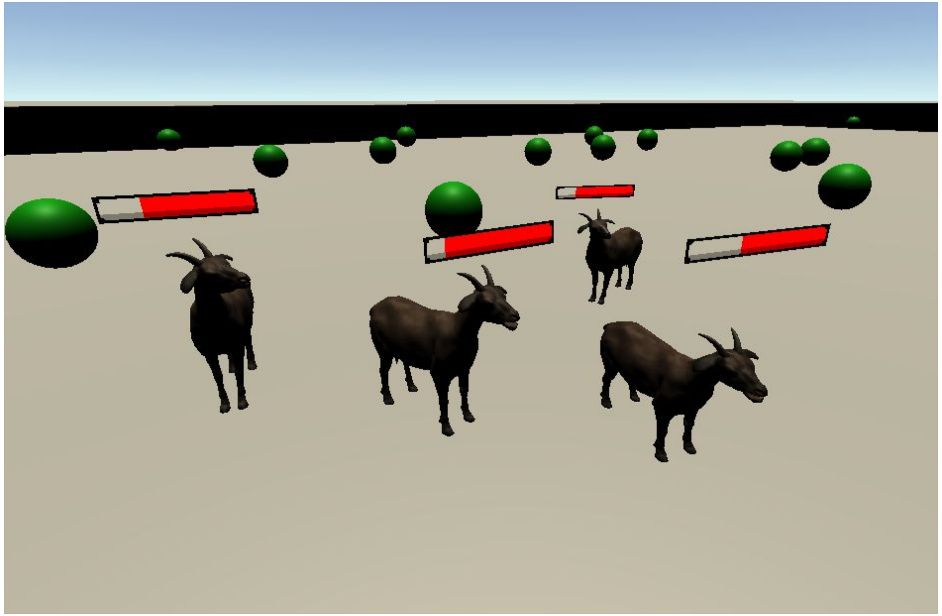
Action Space

Animats

- Idle - No action is taken
- Eat - Eats nearby food
- Accelerate
- Decelerate
- Turn Left
- Turn Right
- Attack (available only to predator) - Damages nearby prey
- Reproduce - Attempts to mate with nearby animat of same species and opposite sex

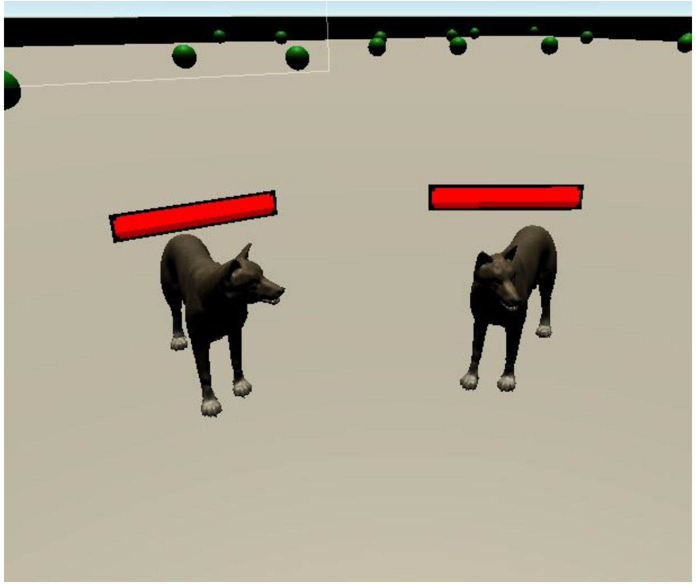
Goats

- Feed on plants
- Reproduce in autumn/winter
- Male has a very high libido
  - Expected to mate with up to ~8 females per mating season



Wolves

- Feed on meat
  - Need to first kill goats
- Reproduce in a winter/autumn
  - Only for simplicity during tests
  - Real wolves mate in spring
- Male has a medium libido
  - Expected to mate with ~4 females per mating season

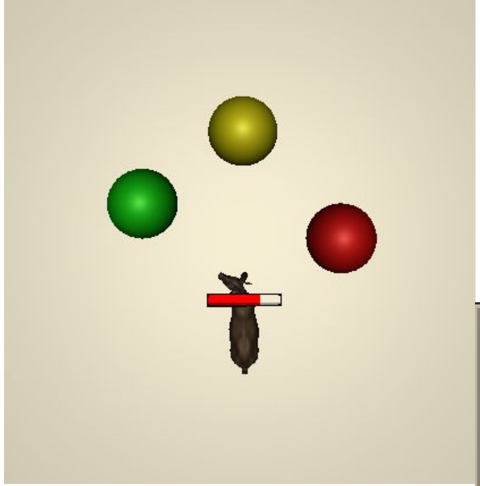


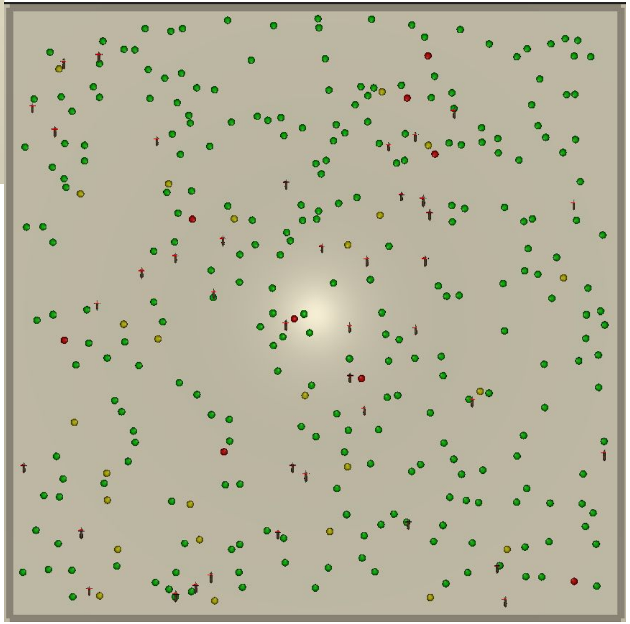
Environments

- Lethal Food
  - Unrealistic food spread
  - Similar to previous implementations
- Advanced Plants
  - More realistic food spread

Lethal Food

- 3 types of food:
  - Good Food
    - Replenishes energy
  - Bad Food
    - Depletes energy
  - Lethal Food
    - Sets energy to 0
- The eating of the types of food is restricted by reflex genes





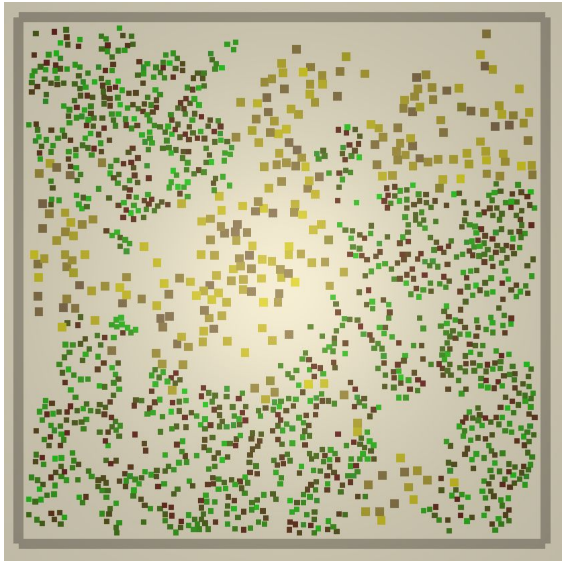
CHALMERS
UNIVERSITY OF TECHNOLOGY

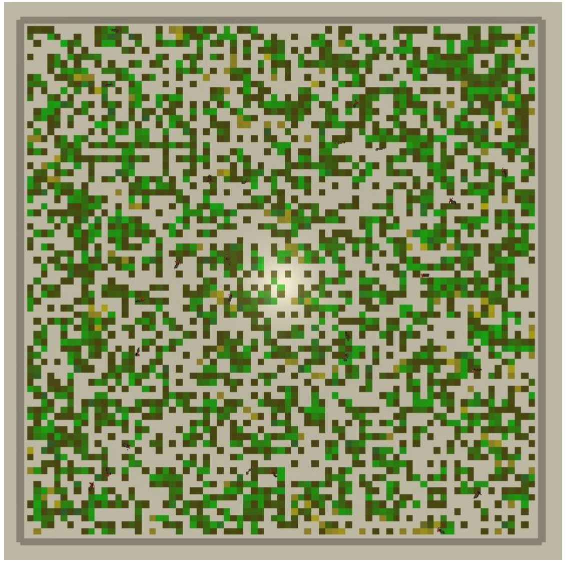
Advanced Plants

- Two types of food for prey species:
    - Grass
        - Spreads in nearby area (root spread)
    - Dandelions
        - Spreads to near or far away positions (pollen spread)
- The spawn of new food is handled by food objects
    - Every food object is a spawner
    - New positions are dependent on the spawner
- Plants compete for spawn positions
- Impossible to spawn within a distance $r_{hostile}$ (or r_{grace} if of same species) of another plant

Advanced Plants

- Food decays over time





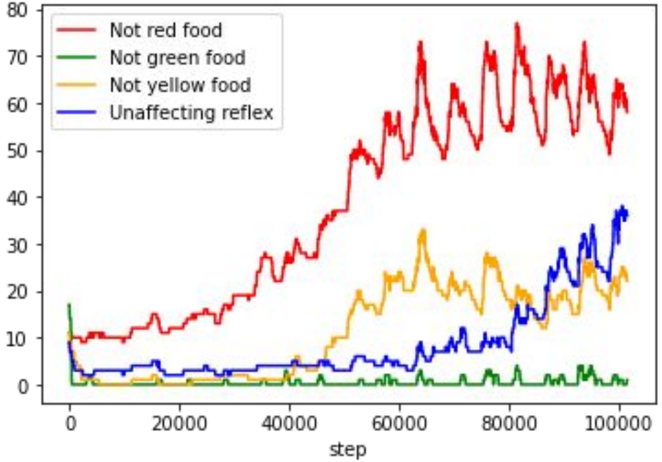
Green: grass
Yellow: dandelion

Red: decayed grass
Brown: decayed dandelion

CHALMERS
UNIVERSITY OF TECHNOLOGY

Reflexes

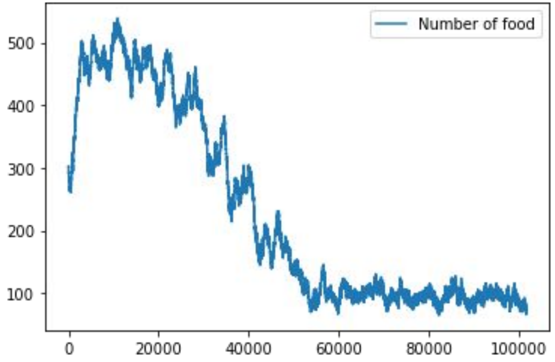
Results

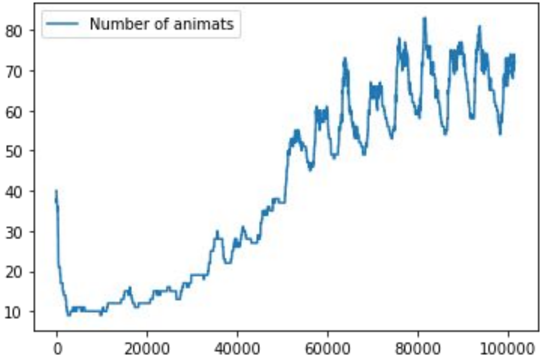


: Lethal Food

: Bad Food

: Good Food





CHALMERS
UNIVERSITY OF TECHNOLOGY

Reflexes: Asexual- vs. Sexual Reproduction

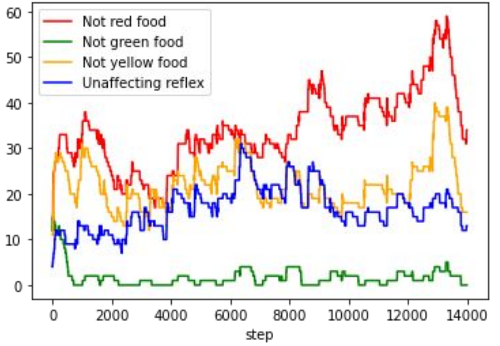
Results


: Lethal Food

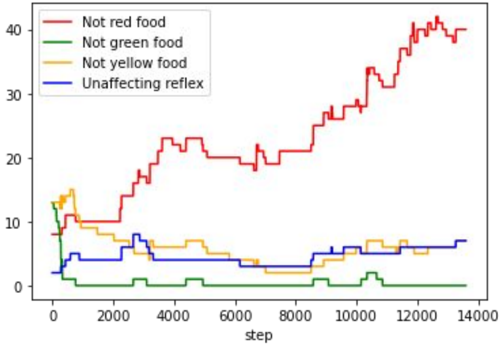

: Bad Food


: Good Food

Asexual Reproduction:



Sexual Reproduction:



CHALMERS
UNIVERSITY OF TECHNOLOGY

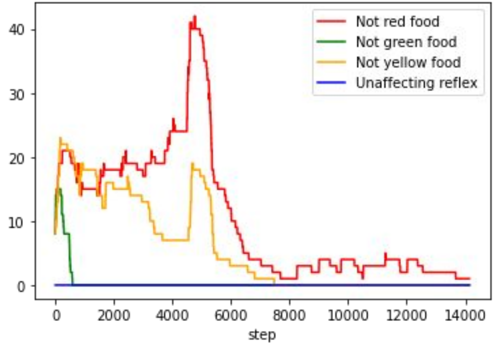
Reflexes: Asexual- vs. Sexual Reproduction


: Lethal Food

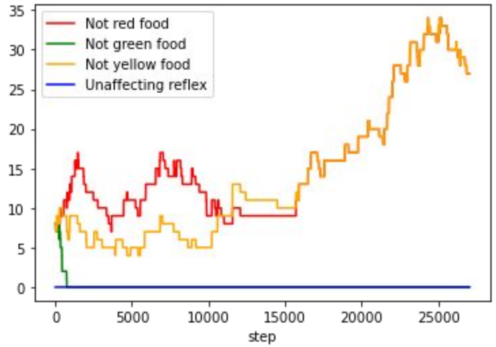

: Lethal Food


: Good Food

Asexual Reproduction:

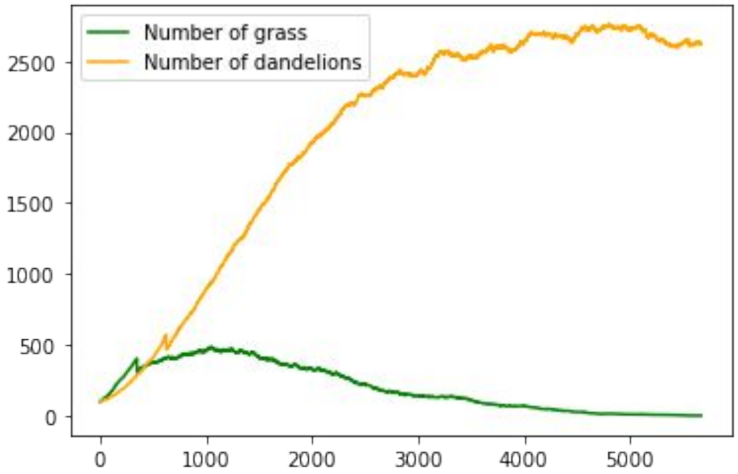


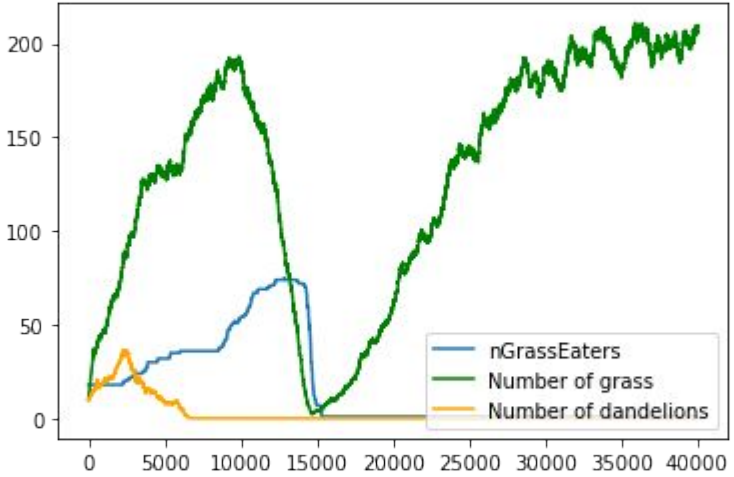
Sexual Reproduction:



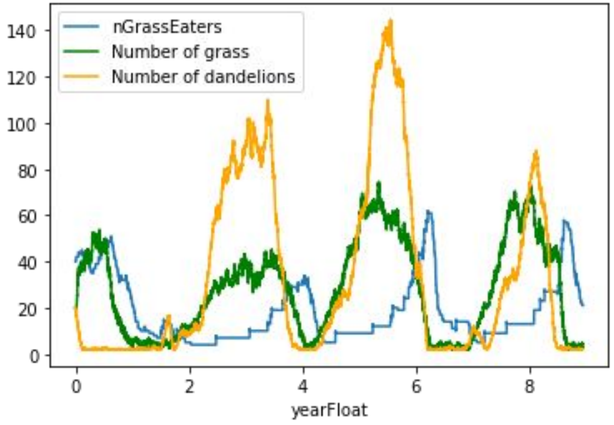
CHALMERS
UNIVERSITY OF TECHNOLOGY

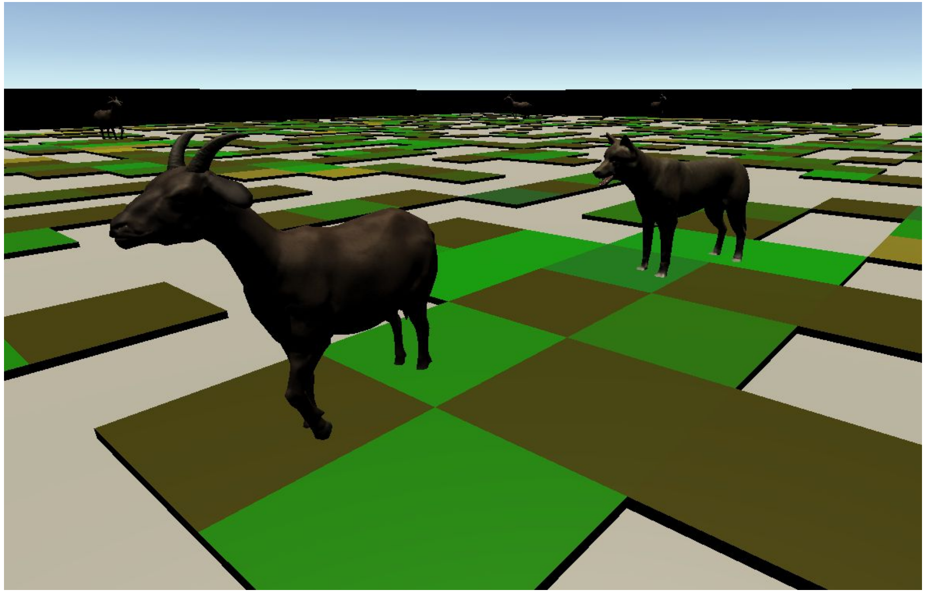
Advanced Plants with Herbivores





Advanced Plants with Herbivores

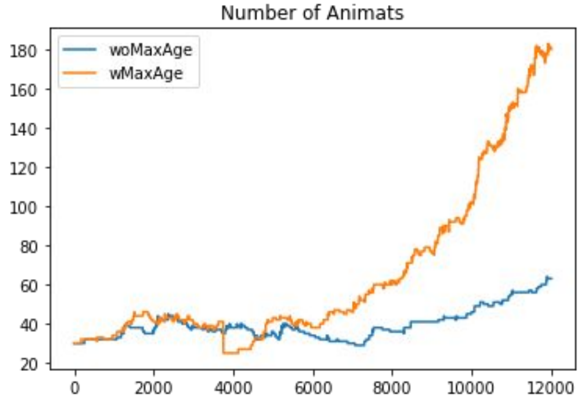


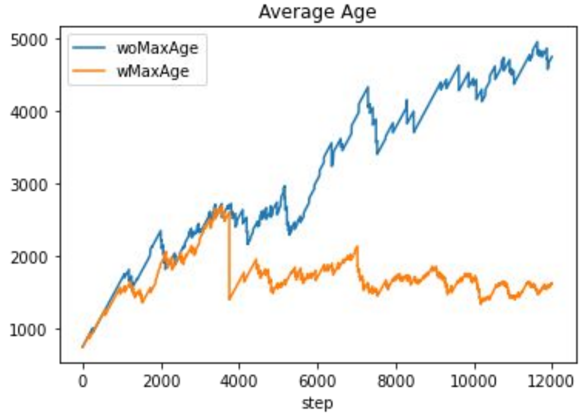


Advanced Plants with Herbivores

Comparing animats with a maximum age limit to animats without an age limit

Oooops...

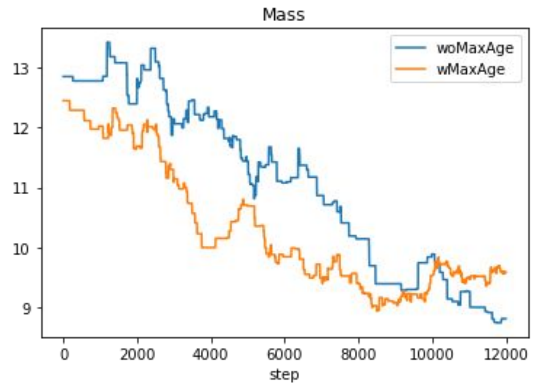


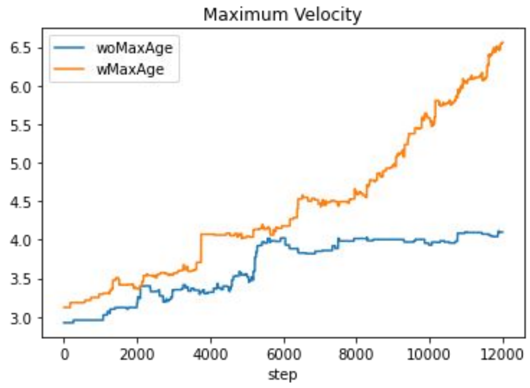


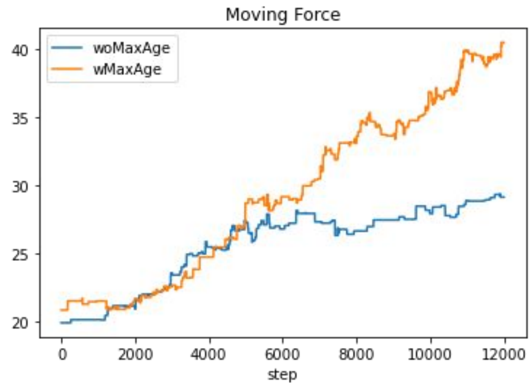
CHALMERS
UNIVERSITY OF TECHNOLOGY

Advanced Plants with Herbivores

Results

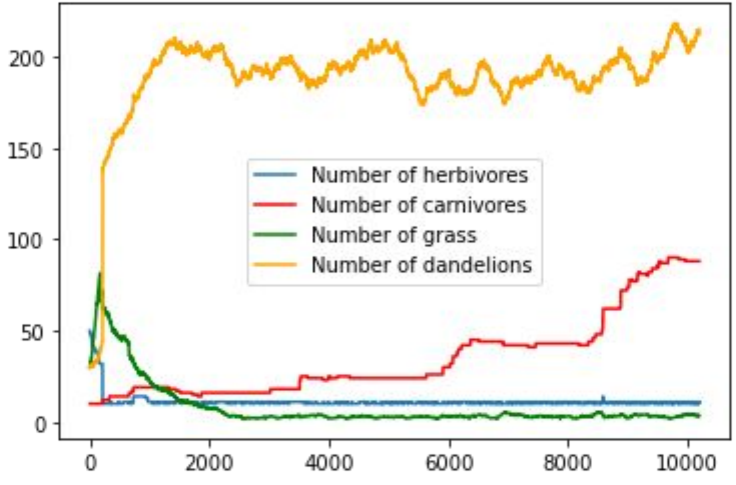


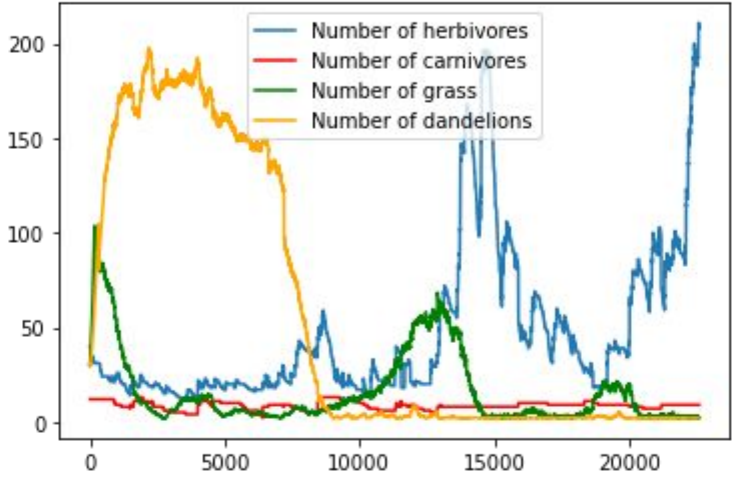




CHALMERS
UNIVERSITY OF TECHNOLOGY

Three-Species Population Dynamics

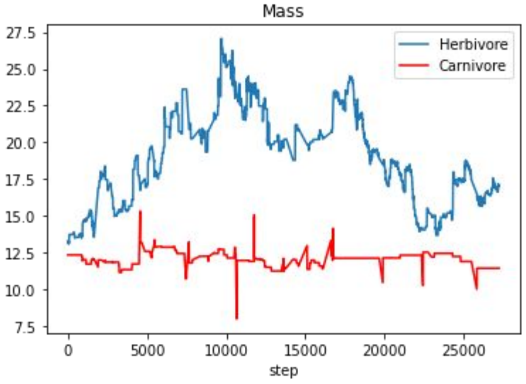


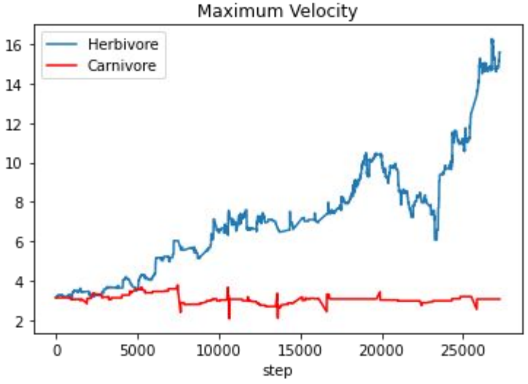


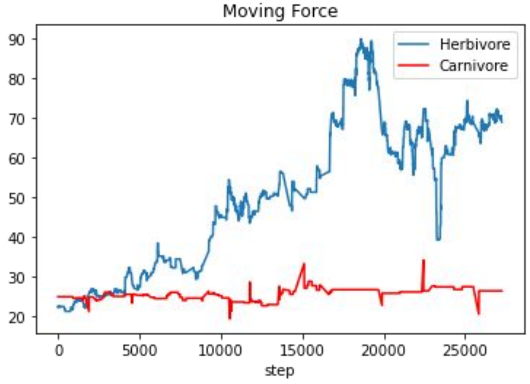
CHALMERS
UNIVERSITY OF TECHNOLOGY

Three-Species Population Dynamics

Results







CHALMERS
UNIVERSITY OF TECHNOLOGY

Conclusions

Discussion

- Reflexes and evolution is necessary in specific environments
- Sexual reproduction is favored and required for survival in dangerous environments
- Asexual reproduction performs better in non-dangerous environments

- Able to achieve Lotka-Volterra like population dynamics in ecosystem with plants and herbivores
- Difficult to balance population dynamics with more levels of the food chain with this complexity, it requires careful parameter tuning

CHALMERS
UNIVERSITY OF TECHNOLOGY

Future Work

- Ecotwin
- Stable three-species Lotka-Volterra with sexual reproduction and evolution
- More realistic gene models
  - Study genes affecting multiple features
  - Study features being affected by multiple genes
- More realistic births
  - Lay eggs
  - Pregnancy
- Introduce actions such as “drink” and homeostatic variables such as “thirst”
- Prediction Network allowing animats to plan their future actions more carefully