Design and Implementation of an Agent Architecture combining Emotions and Reasoning

Janos Tapolczai
Matrikelnummer 0825077

Wien, 09.05.2016

Content

Our Goal

- What is AI supposed to be able to do?

Our Goal

- What is AI supposed to be able to do?
- Many approaches based on search/logic exist:
  - A* search,
  - Iterated deepening DFS,
  - Answer-set programming, etc.

Our Goal

- What is AI supposed to be able to do?
- Many approaches based on search/logic exist:
  - A* search,
  - Iterated deepening DFS,
  - Answer-set programming, etc.
- These do well, but world-ontology has to be encoded in explicit rules!

Our Goal

- Neural networks are another approach:
  - supervised learning,
  - unsupervised learning,
  - hierarchical learning, etc.

Our Goal

- Neural networks are another approach:
  - supervised learning,
  - unsupervised learning,
  - hierarchical learning, etc.
- The network must approximate some function with the help of training data.

Our Goal

- Neural networks are another approach:
  - supervised learning,
  - unsupervised learning,
  - hierarchical learning, etc.
- The network must approximate some function with the help of training data.
- We must have labelled examples/cost functions available.

Our Goal

- Neural networks are another approach:
  - supervised learning,
  - unsupervised learning,
  - hierarchical learning, etc.
- The network must approximate some function with the help of training data.
- We must have labelled examples/cost functions available.
- It's not obvious what a neural network "understands".

Our Goal

- We want to closely imitate the biological brain.

Our Goal

- We want to closely imitate the biological brain.
- We are interested in
  - *realistic behaviour*
  - and *realistic cognition*.

Our Goal

- We want to closely imitate the biological brain.
- We are interested in
  - *realistic behaviour*
  - and *realistic cognition*.
- Our inspirations:
  - *evolutionary neurobiology* and
  - nouvelle AI.

Our Goal — Nouvelle AI

Our Goal — Nouvelle AI

One aim of nouvelle AI is the relatively modest one of producing systems that display the same level of intelligence as insects.

Our Goal — Nouvelle AI

One aim of nouvelle AI is the relatively modest one of producing systems that display the same level of intelligence as insects.

A central idea of nouvelle AI is that the basic building blocks of intelligence are very simple behaviours [. . .] *More complex behaviours "emerge" from the interaction of these simple behaviours.*

Our Goal — Nouvelle AI

> *One aim of nouvelle AI is the relatively modest one of producing systems that display the same level of intelligence as insects.*
>
> *A central idea of nouvelle AI is that the basic building blocks of intelligence are very simple behaviours* [. . .] *More complex behaviours "emerge" from the interaction of these simple behaviours.*

Jack Copeland. What is Artificial Intelligence?

`http://www.alanturing.net/turing_archive/pages/Reference%20Articles/what_is_AI/What%20is%20AI11.html`

Our Goal — Evolutionary Neurobiology

- In what order did the components of the brain evolve?
- What is their relation to each other?
- How well is the brain “designed”?

Our Goal — Evolutionary Neurobiology

- In what order did the components of the brain evolve?
- What is their relation to each other?
- How well is the brain “designed”?

> *It is not worth asking how to define consciousness, how to explain it, how it evolved, what its function is, etc., because theres no one thing for which all the answers would be the same. Instead, we have many sub-capabilities, for which the answers are different: e.g., different kinds of perception, learning, knowledge, attention control, self-monitoring, self-control, etc.*

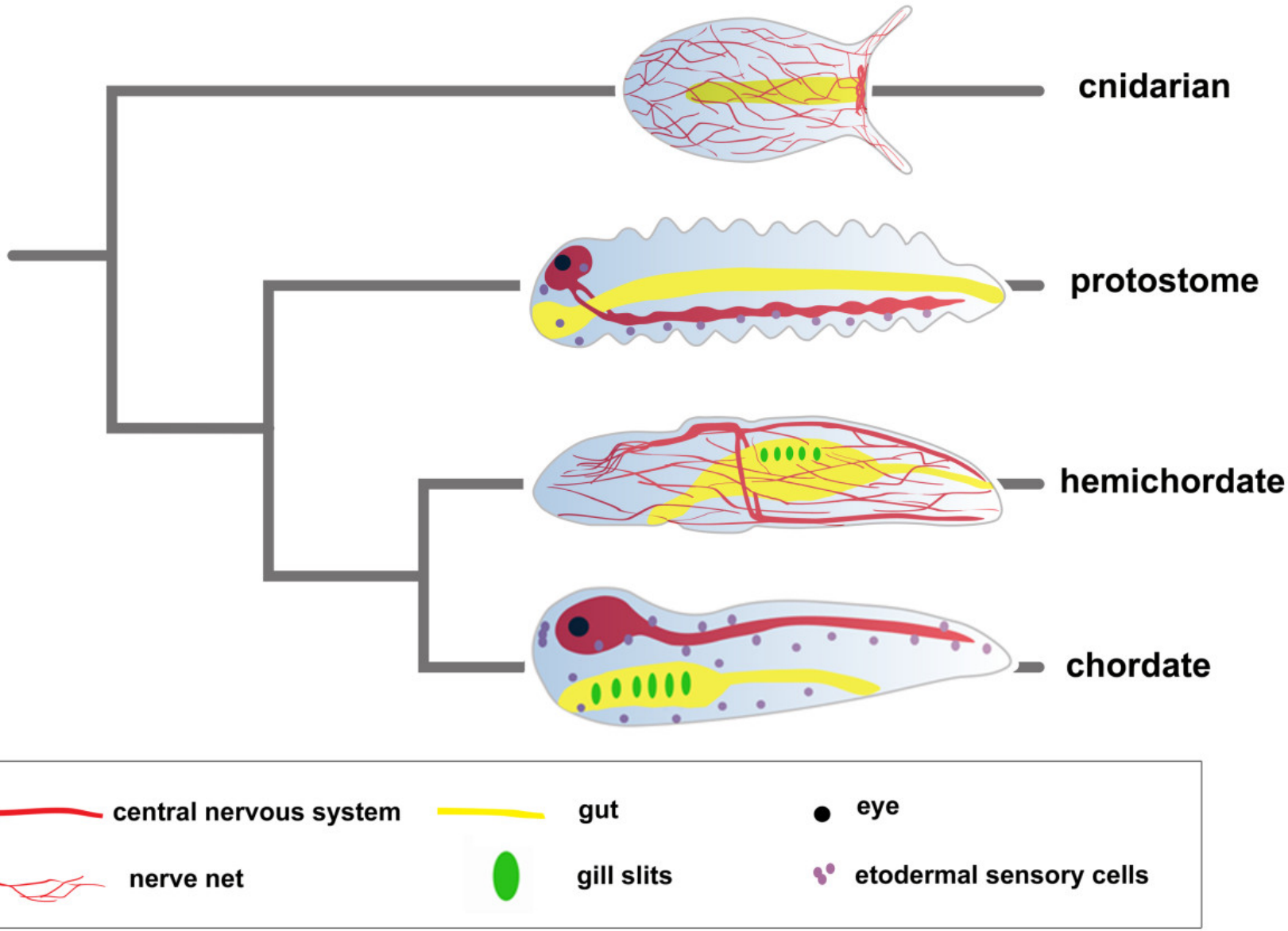
Aaron Sloman, quoted in *The Emotion Machine*, p. 97.

Content

Biological Considerations

- Nervous systems long predate brains.
- Brains themselves evolved over hundreds of millions of years.

Biological Considerations



Biological Considerations

- Neurons function as a signalling mechanism in an organism.

Biological Considerations

- Neurons function as a signalling mechanism in an organism.
- They receive signals from body parts, but also other neurons.
  $\Rightarrow$ Neurons *listen in* on other neurons.

Biological Considerations

- Neurons function as a signalling mechanism in an organism.
- They receive signals from body parts, but also other neurons.
  $\Rightarrow$ Neurons *listen in* on other neurons.
- Functionality was added gradually, with no long-term design.
  $\Rightarrow$ Components have no well-defined interfaces.

Biological Considerations

- Neurons function as a signalling mechanism in an organism.
- They receive signals from body parts, but also other neurons.
  $\Rightarrow$ Neurons *listen in* on other neurons.
- Functionality was added gradually, with no long-term design.
  $\Rightarrow$ Components have no well-defined interfaces.
- Thus our working hypothesis: a *White-box model of cognition*.

Biological Considerations - White-box model

- Traditional programming languages work via *black boxes*:

  A functions are called, but their internals are unobservable.

- We assume that components in the brain are *white boxes*:

  Any component can observe the activity of others.

Biological Considerations - White-box model

- Of course, practical considerations apply.
- We still use regular functions within components,
- but components can publish *messages*.
- These are stored in a central *message space*.
- Any component can read from and write into the message space.
- Components are *loosely coupled*:
  $\Rightarrow$ components don't know who reads their messages;
  $\Rightarrow$ components don't know who wrote the messages.

Content

Architecture

- The implementation consists of a
  - *world simulator* and an
  - *agent architecture*.

Architecture — World Simulator

- Worlds are 2D grids with some inaccessible cells.
- *Entities* are agents or Wumpuses.
- *Wumpuses* roam the world in search for agents.
- *Meat*, *fruits*, and *gold* lie around.
- *Plants* can be harvested for fruit.
- Killed entitites leave behind meat.
- Pits kill whatever falls into them.
- Entities can fight, but this costs health.
- Meat and fruit can be *eaten* to restore health.
- Agents can give each other gifts and communicate via gestures.

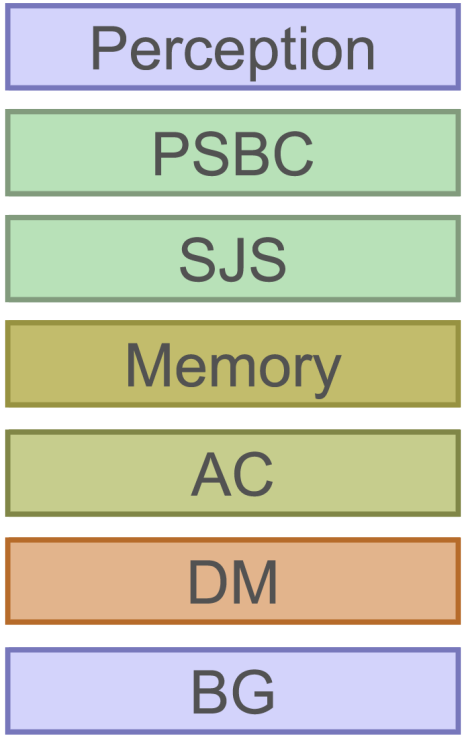
Architecture — World Simulator

- Time moves in rounds.
- In each round, every entity takes one action.
  - The entity gets a slice of the world as *perception* and
  - it has to return its desired action.
- Actions are
  - *move*,
  - *rotate*,
  - *attack*,
  - *gather*,
  - *eat*, etc.

Architecture — Agents

- Based on evolutionary considerations, we designed 7 components.

- Perception,
- Pre-social Behaviour Control (PSBC),
- Social Judgment System (SJS),
- Memory,
- Attention Control (AC),
- Decision Maker (DM), and
- Belief Generator (BG).



Architecture — Agents

- Perception
  - Complex messages from the world simulator are chopped up into atomic pieces.
  - The other components only have to deal with simple facts, e.g.
    - “My Health is 0.7”,
    - “Cell (x,y) has an entity”, or
    - “There is a plant on cell (x,y)”.

Architecture — Agents

- Pre-Social Behaviour Control

Architecture — Agents

- Pre-Social Behaviour Control
  - The agent has four emotions:
    - anger,
    - fear,
    - enthusiasm, and
    - contentment.

Architecture — Agents

- Pre-Social Behaviour Control
  - The agent has four emotions:
    - anger,
    - fear,
    - enthusiasm, and
    - contentment.
  - Each emotion has an associated graph.

Architecture — Agents

- Pre-Social Behaviour Control
  - The agent has four emotions:
    - anger,
    - fear,
    - enthusiasm, and
    - contentment.
  - Each emotion has an associated graph.
  - Each node responds to a specific message (e.g. low health).

Architecture — Agents

- Pre-Social Behaviour Control
  - The agent has four emotions:
    - anger,
    - fear,
    - enthusiasm, and
    - contentment.
  - Each emotion has an associated graph.
  - Each node responds to a specific message (e.g. low health).
  - Nodes may activate neighboring nodes.

Architecture — Agents

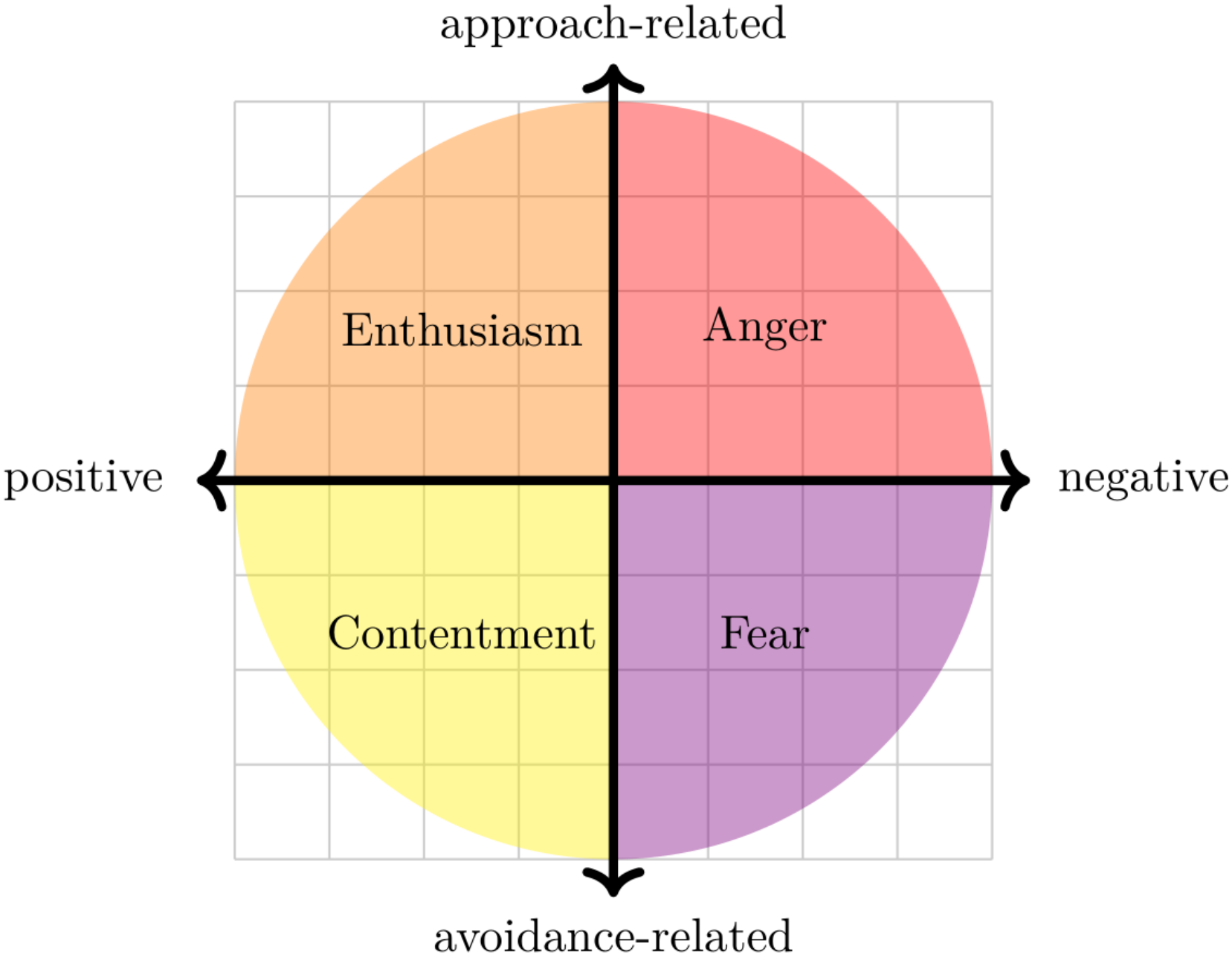
- Pre-Social Behaviour Control
  - The agent has four emotions:
    - anger,
    - fear,
    - enthusiasm, and
    - contentment.
  - Each emotion has an associated graph.
  - Each node responds to a specific message (e.g. low health).
  - Nodes may activate neighboring nodes.
  - The more nodes are activated, the stronger the emotion.

Architecture — Agents

- Pre-Social Behaviour Control
  - The agent has four emotions:
    - anger,
    - fear,
    - enthusiasm, and
    - contentment.
  - Each emotion has an associated graph.
  - Each node responds to a specific message (e.g. low health).
  - Nodes may activate neighboring nodes.
  - The more nodes are activated, the stronger the emotion.

  - “Pre-Social” $\Rightarrow$ anger, fear, etc. are evolutionarily older than social emotions like trust or contempt.

Architecture — Agents



Architecture — Agents

- Social Judgment System
  - Other agents evoke
    - sympathy,
    - trust,
    - respect.

Architecture — Agents

- Social Judgment System
  - Other agents evoke
    - sympathy,
    - trust,
    - respect.
  - Every stranger has its own emotional levels.
  - Agents can become friends (through positive interactions) or enemies.

Architecture — Agents

- Memory
  - We store our perceptions for later recall.
  - Memory can also store *imagined* worlds in a *tree structure*.

Architecture — Agents

- Memory
  - We store our perceptions for later recall.
  - Memory can also store *imagined* worlds in a *tree structure*.
- Attention Control
  - We assist the Decision Maker by selecting important targets.
  - “Important” means ”evokes the strongest emotions”.

Architecture — Agents

- Decision Maker
  - Each emotion has some actions associated with it, e.g.,
    - attacking is associated with anger,
    - eating is associated with enthusiasm.

Architecture — Agents

- Decision Maker
  - Each emotion has some actions associated with it, e.g.,
    - attacking is associated with anger,
    - eating is associated with enthusiasm.
  - The PSBC and the AC guide the planning:
    - We select an action associated with our strongest emotion,
    - and the cell(s) which have the most attention.

Architecture — Agents

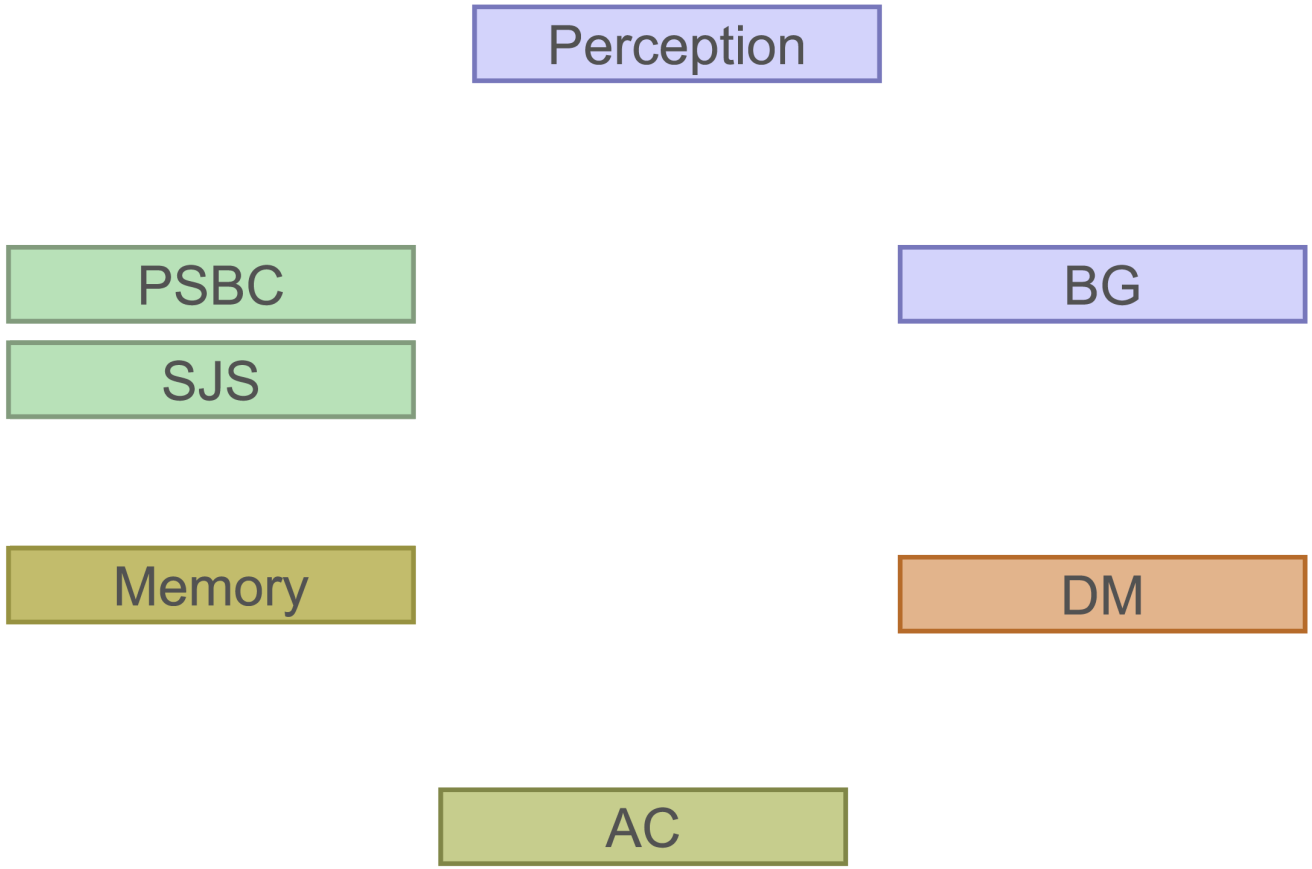
- Decision Maker
  - Each emotion has some actions associated with it, e.g.,
    - attacking is associated with anger,
    - eating is associated with enthusiasm.
  - The PSBC and the AC guide the planning:
    - We select an action associated with our strongest emotion,
    - and the cell(s) which have the most attention.
  - If our emotion is strong enough, we take a *real* action,
  - otherwise, we take a *hypothetical* one.
  - The DM can also abort plans if, e.g., fear begins to override anger.

Architecture — Agents

- Belief Generator
  - If the DM chose a *hypothetical* action, we simulate its consequences.
  - We use the actual world simulator, but we construct the world from memory.
  - The agent is then given *imagined perceptions*.

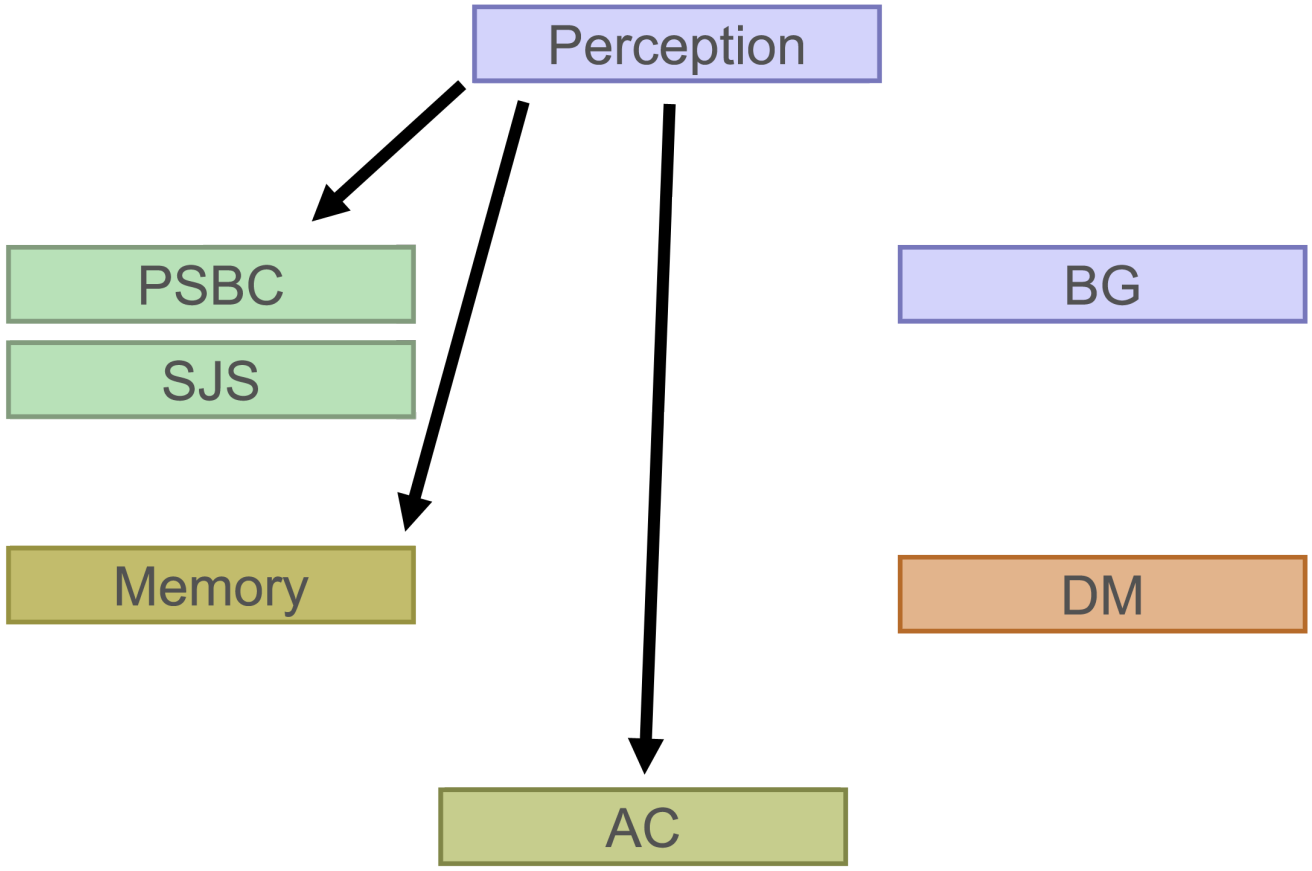
Architecture — Agents

- How does this all fit together?



Architecture — Agents

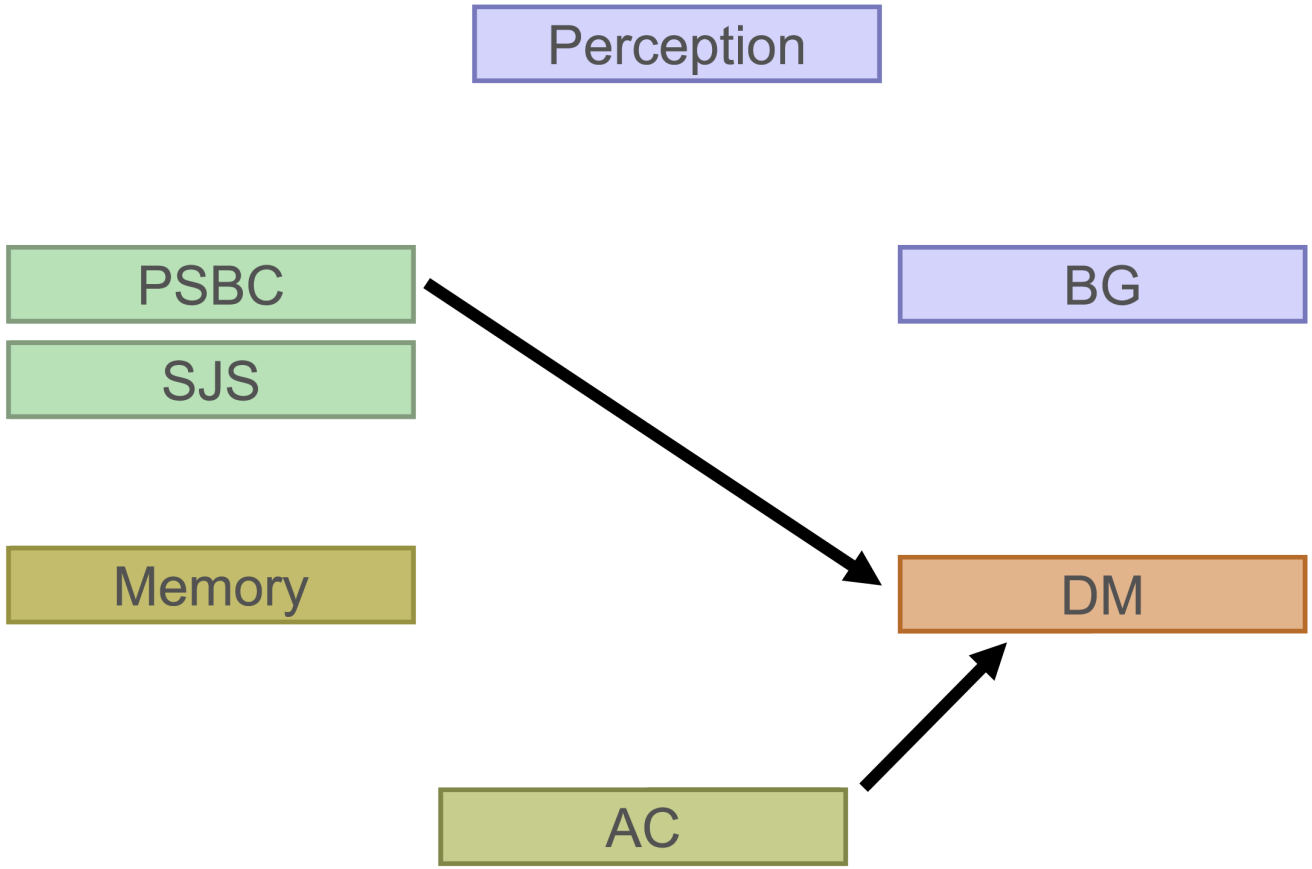
- How does this all fit together?



Perception distributes its messages.

Architecture — Agents

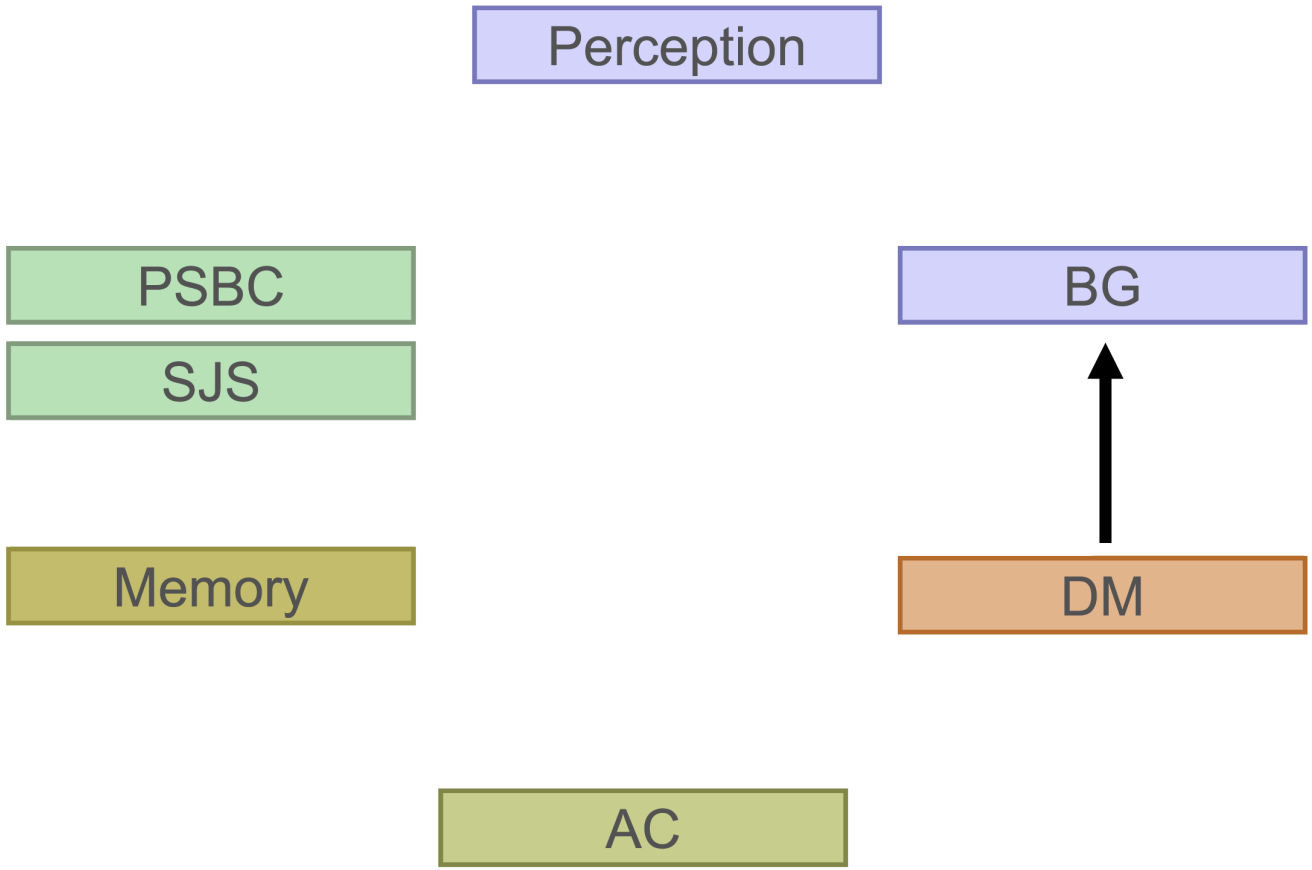
- How does this all fit together?



The *Decision Maker* is informed about the affective reactions.

Architecture — Agents

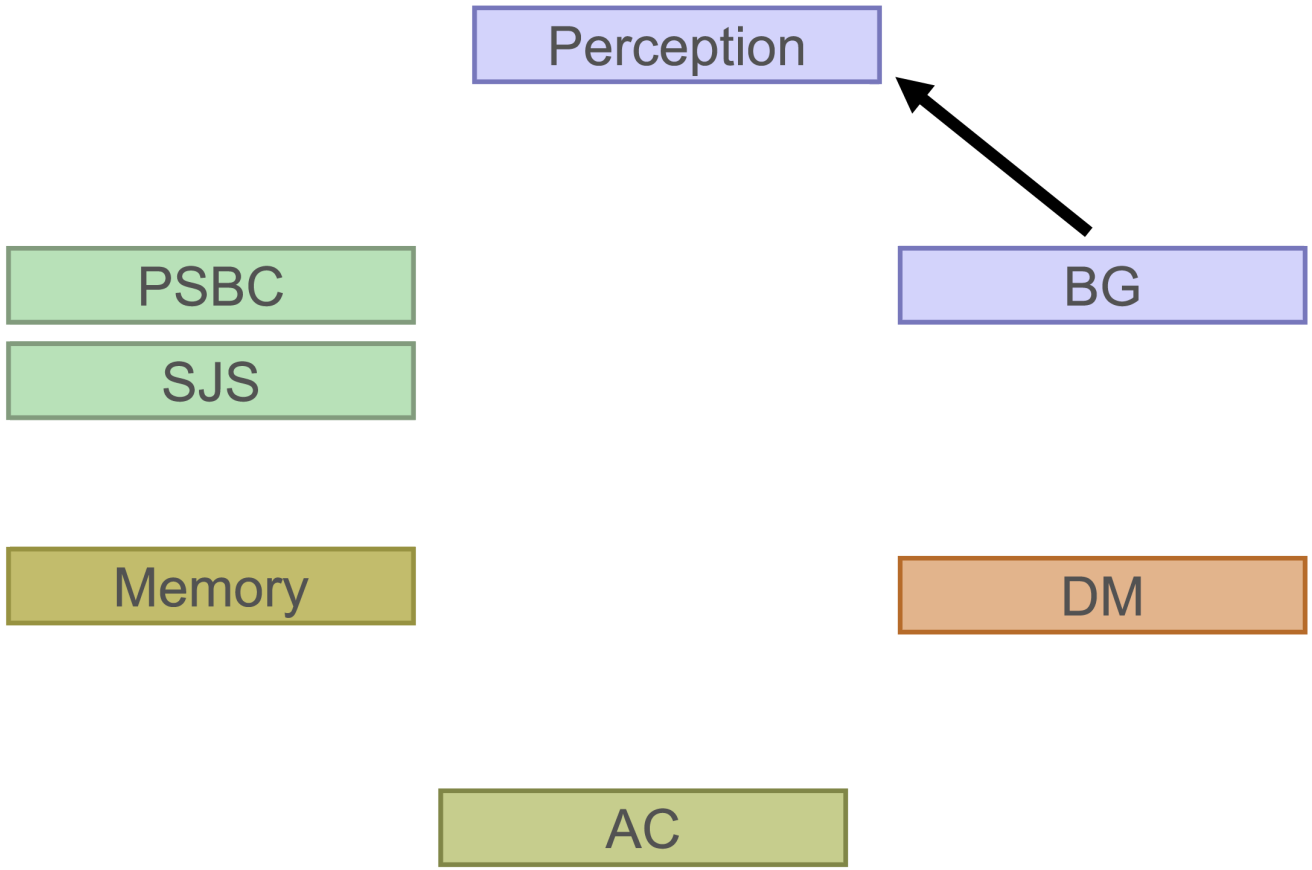
- How does this all fit together?



Option 1: The *Belief Generator* simulates the consequences.

Architecture — Agents

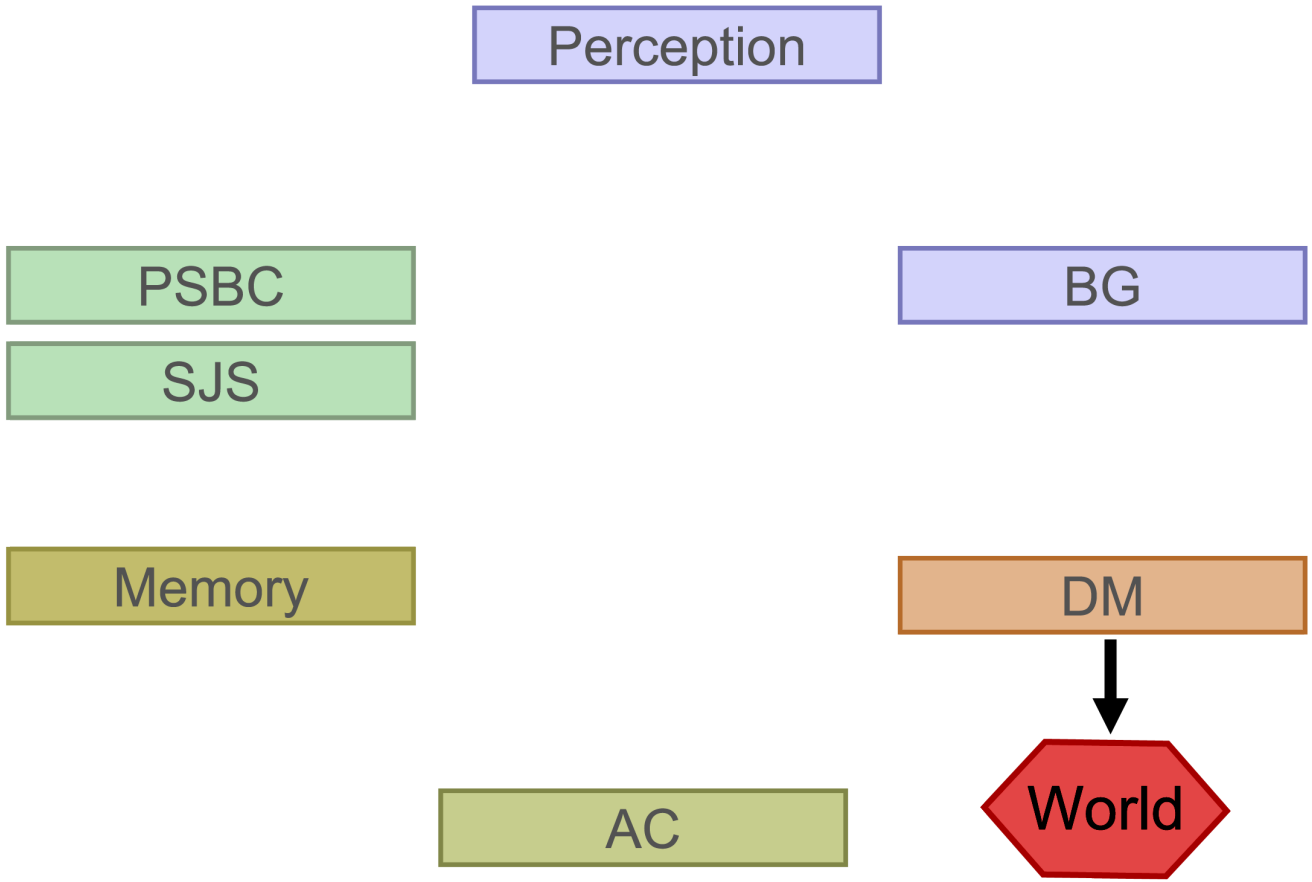
- How does this all fit together?



The loop begins anew, but with *imagined* perceptions.

Architecture — Agents

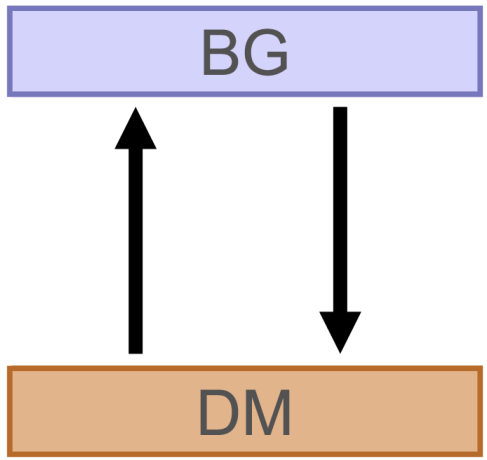
- How does this all fit together?



Option 2: The *Decision Maker* chooses a real action.

Architecture — Agents

- To simplify it: DM and BG are in a loop.



1. We choose a *hypothetical* action,
2. Then we simulate its consequences.

- We repeat this until the DM deems the outcome satisfactory and chooses a *real* action.

Content

Results

- We created 32 populations with different personalities
- and simulated 50 rounds with each population.
- We combined weak/strong anger, fear, enthusiasm, contentment, as well as friendly/hostile demeanor.
- Throughout the simulation, we collected data: the number of
  - gifts given,
  - surviving Wumpuses,
  - surviving agents,
  - plants harvested, etc.

Results

Personality fragment	*weak/hostile*	*strong/friendly*
Anger	24.25	22.313
Fear	22.438	24.125
Enthusiasm	22.75	23.8125
Contentment	23.063	23.5
Hostility	23.625	22.938

Table: Average number of surviving agents, by personality fragment.

Results

Personality fragment	*weak/hostile*	*strong/friendly*
Anger	24.25	22.313
Fear	22.438	24.125
Enthusiasm	22.75	23.8125
Contentment	23.063	23.5
Hostility	23.625	22.938

Table: Average number of surviving agents, by personality fragment.

Results

Personality fragment	*weak/hostile*	*strong/friendly*
Anger	24.25	22.313
Fear	22.438	24.125
Enthusiasm	22.75	23.8125
Contentment	23.063	23.5
Hostility	23.625	22.938

Table: Average number of surviving agents, by personality fragment.

- Weak anger and strong fear are useful!

Results

- Other surprising results:

Personality	*Wumpuses*
$\langle \mathtt{S}, \mathtt{W}, \mathtt{W}, \mathtt{W}, \mathtt{H} \rangle$	0
...	
$\langle \mathtt{W}, \mathtt{W}, \mathtt{W}, \mathtt{W}, \mathtt{H} \rangle$	0
$\langle \mathtt{S}, \mathtt{W}, \mathtt{W}, \mathtt{S}, \mathtt{F} \rangle$	1
$\langle \mathtt{S}, \mathtt{W}, \mathtt{W}, \mathtt{W}, \mathtt{F} \rangle$	2
...	
$\langle \mathtt{S}, \mathtt{W}, \mathtt{S}, \mathtt{W}, \mathtt{F} \rangle$	19
$\langle \mathtt{S}, \mathtt{W}, \mathtt{S}, \mathtt{W}, \mathtt{H} \rangle$	19

Table: Number of meat items given as gifts after 50 rounds.

- $\langle \mathtt{S}, \mathtt{W}, \mathtt{S}, \mathtt{W}, \mathtt{H} \rangle$ means “strong anger, weak fear, strong enthusiasm, weak contentment, hostile demeanor”.
- Agents with strong anger and weak fear killed everything and shared the meat among themselves.

Content

Conclusion

- We designed an agent architecture based on evolutionary considerations.

Conclusion

- We designed an agent architecture based on evolutionary considerations.
- Our aim was to approximate real organisms.

Conclusion

- We designed an agent architecture based on evolutionary considerations.
- Our aim was to approximate real organisms.
- Agents were put into a moderately complex game world

Conclusion

- We designed an agent architecture based on evolutionary considerations.
- Our aim was to approximate real organisms.
- Agents were put into a moderately complex game world
- and performed reasonably well.

Conclusion

- We designed an agent architecture based on evolutionary considerations.
- Our aim was to approximate real organisms.
- Agents were put into a moderately complex game world
- and performed reasonably well.
- Personalities differentiated themselves in interesting ways.

Thank you for your attention!