Requirements [1]

Version 1.0
5th March, 2022

Adaptron Inc.

Adaptron is an artificial intelligent agent and as such it needs to be able to do things that humans can do.
These requirements describe what it needs to know and what it needs to do.
If Adaptron is to be a successful artificial intelligent agent it makes sense that its requirements should
correspond to those of naturally intelligent agents such as humans.
Because cognitive science describes human behaviour, it is an excellent source of requirements.

Contents

- Purposeful
- Grounded
- General purpose
- Adaptable
- Stages of learning
- Other requirements
- Thinking / reasoning

This presentation elaborates on some of the more important requirements for Adaptron to achieve artificial general intelligence.

Purposeful

- Motivated
  - Intrinsic
    - Seek out novelty / avoid familiarity
    - Maximize reliability / reduce uncertainty
  - Extrinsic
    - Pursue pleasure / avoid unpleasant stimuli

Adaptron needs a purpose. Otherwise it may as well be a rock.

This means that it needs to be motivated to act and achieve one or more goals.

Such goals must be built in by its creators.

The two intrinsic motivations are the pursuit of novelty and reliability.

The first produces exploratory behaviour and the second results in practice.

Extrinsic motivations are all about obtaining pleasure and avoiding unpleasant things

Examples are eating tasty food, pleasant touch and sexual behaviour.

These all produce some form of exploitative behaviour.

Types of Motivation

- Intrinsic [2]
  - Seek novelty
    - Interesting stimuli
  - Avoid familiarity
    - Boring stimuli
  - Produces exploratory behaviour – curiosity
    - Orienting responses
  - Inherent in information processing
    - Wherever there is memory

Since motivation is such an important aspect of autonomous robot behaviour I cover this subject in more detail.

The intrinsic motivation of seeking novelty results in exploratory behaviour such as pursuit of entertainment and playing.

Surprising stimuli are novel. There is an interest in re-experiencing them. We call it curiosity.

Once they are fully explored and familiar they loose their interest and become boring.

Animals are naturally curious at a young age and most of their learning is based on intrinsic motivation.

The orienting response is an example of a reflex that attracts ones attention to novelty.

Unexpected changes in the environment are novel and they interrupt our behaviour.

Intrinsic motivation is an inherent property of information processing. New information is naturally compared to existing information (i.e. memories).

Types of Motivation

- Intrinsic
  - Maximize reliability
    - Predictable acts
  - Reduce uncertainty
    - Unpredictable acts
  - Produces repetitive behaviour
    - Practice and play
  - Reliable acts can be done automatically

The second type of Intrinsic motivation is to achieve reliable behaviour.

The problem is that the results of repeating an action are never quite the same.

Practice repeats an action until all the possible results have been experienced, are known and predictable.

Once they are fully practiced and reliable, actions can be done automatically without conscious involvement.

Types of Motivation

- Extrinsic
  - Pursue pleasure
    - Rewarding stimuli
  - Avoid unpleasant stimuli
    - Punishing / painful stimuli
  - Produces purposeful behaviour
    - Directed action
  - Built into a robot
    - Determined by its creator

Pleasant and unpleasant stimuli provide the extrinsic motivation to perform purposeful behaviour.

In humans they include the pleasure from certain foods, touch and other survival related motivations.

Extrinsic motivations are decided on and built into a robot by its creator.

However they could be left out.

In that case the result would be a robot that is always curious and continuously exploring.

It would always be playing and seeking entertainment.

Autonomous [3]

- Act independently of external control
- Gain knowledge and abilities from experience [4]
  - Interacting with the world
  - Unsupervised learning
- Distinguish between events it caused and those caused by the world
- But dependent of its extrinsic motivations

Adaptron needs to act independently of any external controls. We don't want it to be a puppet.

What it knows and can do must all be gained from its experience interacting with the world.

This is known as unsupervised learning in artificial intelligence.

This means that Adaptron needs to learn, play and be trained like humans.

It also means that it must be able to distinguish between events it caused and those that happen independently in the world.

But it cannot be independent of everything. It will still be dependent on its built in motivations and

the configuration of its senses and action devices.

Grounded [5]

- Symbol grounding problem [6]
  - Concepts related to the world
    - How do they get their meaning?
- Peripatetic axiom
  - Nothing is in the intellect that was not first in the senses
- Convert non-symbolic sensory values into symbolic ones

Given the assumption that an intelligent agent uses symbolic representations for its knowledge and thoughts there is a problem with how are these symbols related to the realities of the world which are mostly analog / non-symbolic. This is known as the symbol grounding problem.

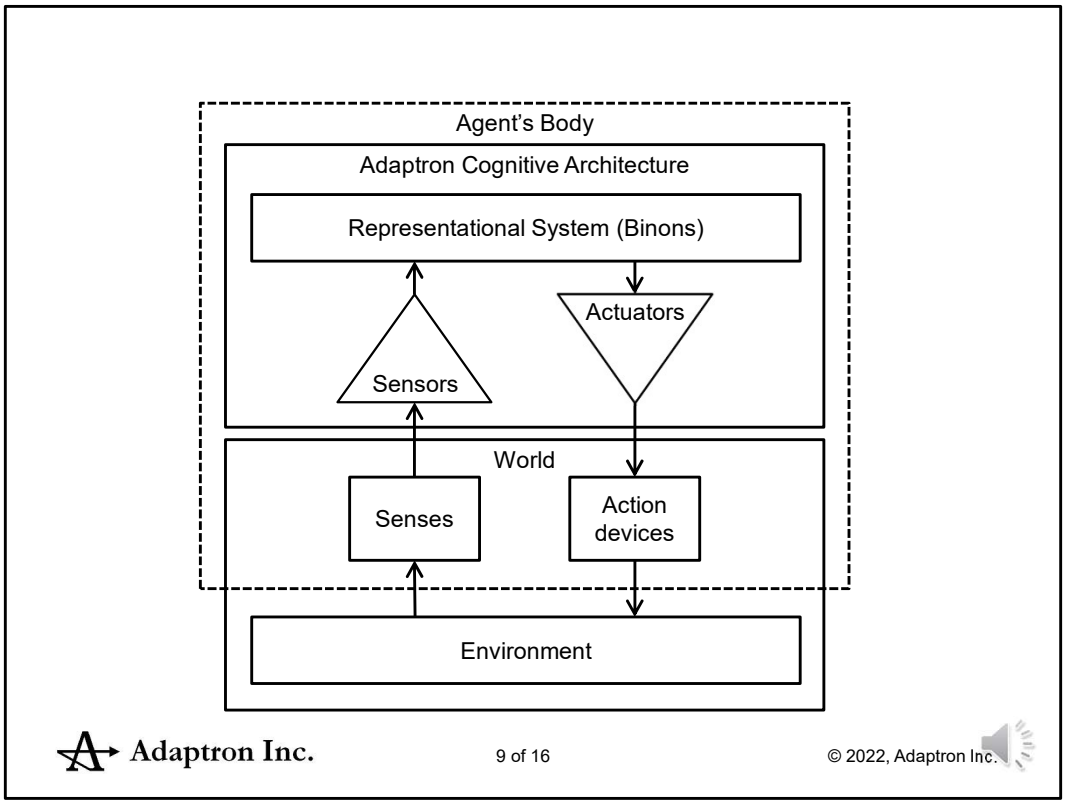
Since Adaptron is going to use symbolic representations, there is a need for it to be grounded.

The solution comes from the Peripatetic axiom, which says that “Nothing is in the intellect that was not first in the senses”.

In other words everything we learn originates in the senses.

This means that Adaptron must obtain all its knowledge of the world via its sensors.

And this means there needs to be a way of converting non-symbolic sensory information into symbolic values.



Adaptron needs to be connected to its environment via a configuration of senses and action devices.

Any number and combination of senses and action devices should be possible.

They are part of an intelligent agent's body but from Adaptron's perspective they are in the world because they can fail.

An important layer inside Adaptron is that of the *sensors* and *actuators*. They allow it to communicate with the world via the senses and action devices. In animals, for example, *hair cells* in the *ear* detect the volume of *sounds*, *cone* and *rod cells* in the *eye* measure *colour* and *brightness* respectively, and *thermoreceptors* in the *skin* detect *heat.* In Adaptron *charge-coupled devices* are the sensors detecting the light intensity in a camera. Actuators work in the opposite direction. For example, *muscle fibers* contract when stimulated by *motor neurons.* In Adaptron, lights shine brighter, heaters burn hotter and motors produce more *torque* if the *voltage* is increased, and the *loudness* of a speaker is determined by the *amplitude* of the electrical *signal*.

General purpose

- Interface devices
- Memory
  - Sensory
  - Working
  - Declarative
    - Semantic
    - Episodic
  - Procedural
- Invariant features for perceptual constancy

If we want Adaptron to do things that humans do it is going to have be general purpose.

This means that it must be able to use all types of senses and action devices, i.e., interface devices.

It must also have a variety of memories that span different time scales and purposes.

Sensory memory for events last for a maximum of a few seconds.

Working memory captures ones last few thoughts.

Declarative memory keeps ones knowledge about things and events that have happened.

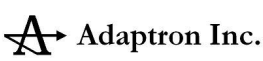
Procedural memory contains the knowledge of how to do things.

And since no two experiences are ever the same it must be able to recognize and represent the invariant features

of objects and events so as to recognize them.

Adaptable

- Learn to
  - Perceive
  - Act
  - Think / reason
- Learn
  - Continuously as the world changes
  - Quickly



As a adaptable agent Adaptron needs to learn all that it knows and can do by interacting with its world.

This means it needs the ability to perceive the world, act in the world and use reasoning to predict what will happen when it performs actions.

For perception it needs to be able to recognize and encode both spatial and temporal representations of objects and events.

For action it needs to be able to control its action devices to make changes in the world.

And it needs to learn how to reason about its world based on its experiences so as to predict the outcomes of its actions and thereby decide on what to do.

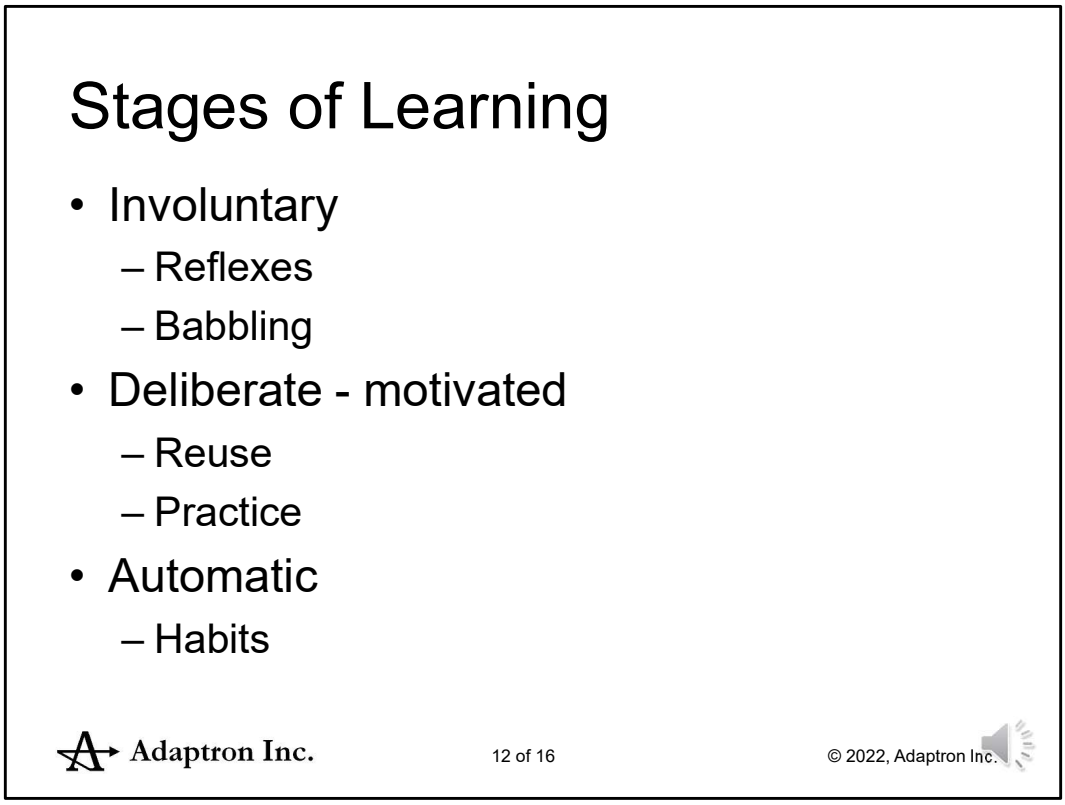
It needs to continuously learn about the real world as it changes.

To accommodate this requirement, its memories must evolve and grow.

Learning also needs to be based on as few experiences as possible.

Adaptron should not have to repeat an experience more than a few times to learn it.

This is only possible by building upon already known less complex behaviours.



In Adaptron there are five stages to learning. They can be categorized into three types, Involuntary, deliberate and automatic.

Reflexes are built-in involuntary actions that are triggered by predefined stimuli.

The orienting response is an example of one that attracts attention to novel stimuli while the

blink reaction is an example of one that is designed to protect the eyes.

Babbling is a way for a young agent to produce random actions to “experiment” with its world.

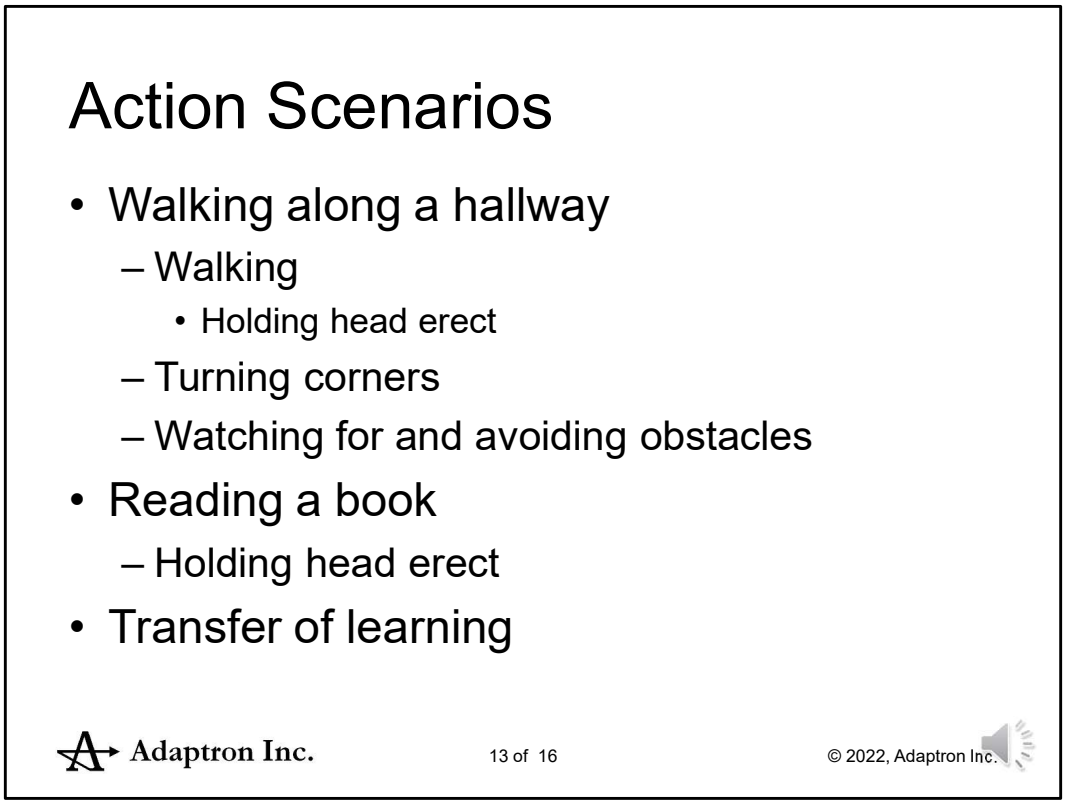
Acts learnt from reflexes and babbling get reused if they produced results that accomplished goals.

But acts don’t produce consistent results so acts are practiced until all the possible results are known.

What is learnt are reliable habits. They can be repeated and be performed automatically.

They are started consciously but are carried out sub-consciously.

These include habits to recognize things, to do things and to think / reason.



If it is going to be useful, Adaptron needs to act in the world.

Controlling actions is quite complex. For example, in the task of walking along a hallway one has to perform many acts in parallel.

They include walking, holding one's head erect, turning corners and watching for and avoiding obstacles.

But it is not just a simple list.

Holding one's head erect is part of walking and is reused in many other actions such as reading a book.

And turning a corner or avoiding obstacles are not done continuously, they are ready to be performed when the right situation is encountered.

Another key action requirement is the transfer of learning. It is the reuse of known actions in new situations.

For example, you have learnt to use a potato peeler on potatoes and carrots and then you try to peel an apple with it and you can.

Other Requirements

- Controlled and safe [7]
- Transparent
- Scalable [8] [9]
- Efficient
- Robust
- Reliable

Adaptron Inc. 14 of 16

Adaptron needs to trustworthy and there needs to be the ability to control it should our safety be compromised.

Especially considering its ability to act autonomously to achieve its goals.

Being transparent means it must be possible to examine Adaptron's processes and understand how it makes decisions.

Scalability is a requirement for a system to handle an increasing number and complexity of its resources and environments.

Adaptron needs to be able to work with a variety of interface devices and operate in many diverse worlds.

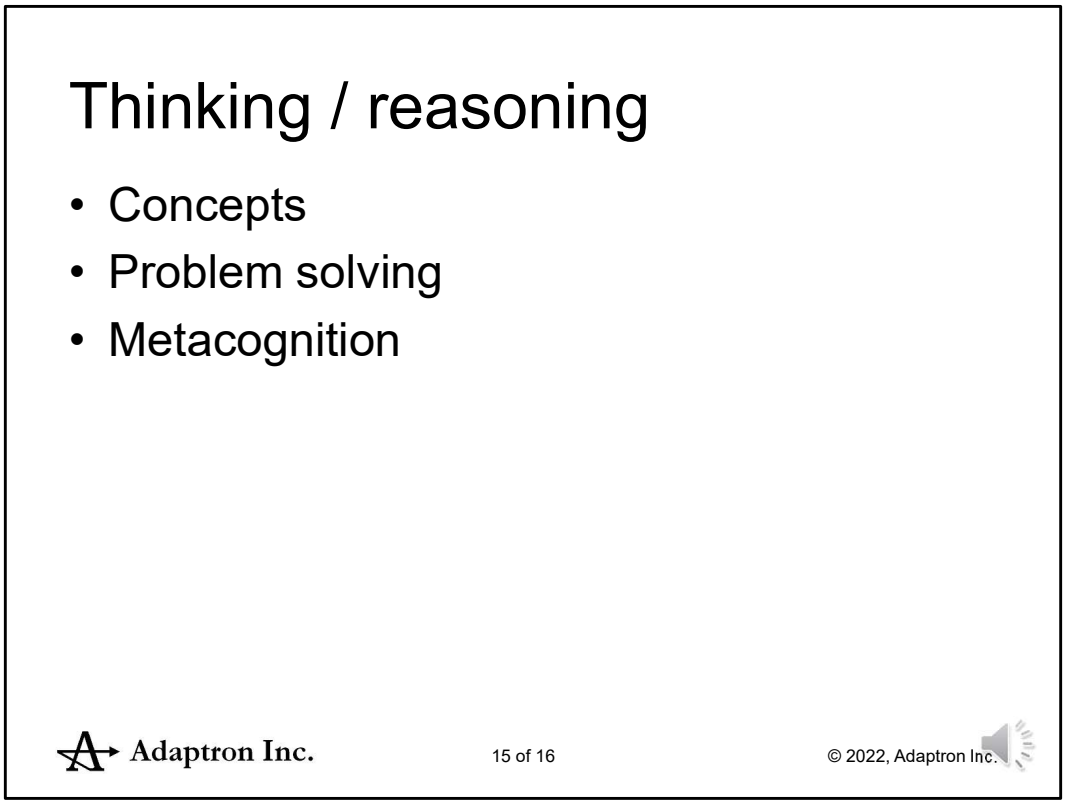
It needs to continuously gain knowledge and abilities throughout its life.

It must be efficient in its mental representations, i.e., no redundant or duplicated information.

It should also be able to learn rapidly from just a few experiences.

To be robust it must recover from failures or problems that occur due to the natural fluctuations in the world.

Adaptron is reliable if it has a low failure rate.



Adaptron needs to be able to reason. This is often described as the ability to simulate its world so as to predict what will happen next either as a result of what it does or what just happens naturally in the world.

Reasoning includes recalling things from its mental model of the world which it has gained from interacting with it.

If the reasoning is directed at achieving a goal it is called problem solving.

Reasoning about thinking is referred to as metacognition.

It is the ability to think about the why, when and what of one's thoughts.

References

[1] Vernon, D., von Hofsten, C., & Fadiga, L. (2016). Desiderata for developmental cognitive architectures. *Biologically Inspired Cognitive Architectures* 18, 116-127.

[2] Burda, Y., Edwards, H., Pathak, D., Storkey, A., Darrell, T., & Efros, A. A. (2018). Large-scale study of curiosity-driven learning. *arXiv:1808.04355v1 [cs.LG].*

[3] Thórisson, K. R. & Helgasson, H. P. (2012). Cognitive architectures and autonomy: a comparative review. *Journal of Artificial General Intelligence,* 3(2), 1-63.

[4] Kugele, S., & Franklin, S. (2020). General intelligence requires autonomous, cognitive, intentional agents. E*ighth Annual Conference on Advances in Cognitive Systems*.

[5] Barsalou, L. W. (2010). Grounded cognition: past, present, and future. *Topics in Cognitive Science* 2(4), 716-724.

[6] Harnad, S. (1990). The symbol grounding problem. *Physica D,* 42, 335-346.

[7] Everitt, T., Lea, G., & Hutter, M. (2018). AGI safety literature review. *arXiv:1805.01109v2 [cs.AI].*

[8] Rutkowska, J. C. (1994). Scaling up sensorimotor systems: constraints from human infancy. *Adaptive Behavior,* 2(4), 349-373.

[9] Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural networks: a review *arXiv:1802.07569v4 [cs.LG]. Neural Networks,* 113, 54-71.

References