MASTERARBEIT

Titel der Masterarbeit

TUNING AND VERIFYING A PSYCHOLOGICALLY PLAUSIBLE COGNITIVE ARCHITECTURE

USING LIDA-BASED COGNITIVE SOFTWARE AGENTS

Verfasser

Tamas Madl, BSc

angestrebter akademischer Grad

Master of Science (MSc)

Wien, Juli 2011

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PUBLICATIONS

Some of the ideas and results in this work have appeared before in the following publication:

Madl, T., Baars, B.J. & Franklin, S., 2011. The Timing of the Cognitive Cycle. PLoS ONE, 6(4), p.e14803.

Sections originally published in this paper are marked with an asterisk (*) at the end of the section title.

ABSTRACT

This thesis describes an approach to provide evidence that the LIDA (Learning Intelligent Distribution Agent) cognitive architecture models basic human cognitive processes. The computational LIDA architecture's processes are correlated with neural processes and its parameters are adjusted using brain-related evidence. The model is also extended by an invariant feature detection-based visual object recognition system, and a novel attentional mechanism based on a model of the locus coeruleus – norepinephrine system in the brain, to facilitate computational modeling of a wider range of cognitive tasks than was possible before. Three LIDA-based cognitive software agents are also described and compared with human behavioral data, modeling three different psychological paradigms: a simple reaction time experiment, a perceptual continuity experiment and a visual attentional blink task. The thesis shows that the LIDA-based agents are capable of accurately reproducing human data, and argues in favor of psychological - and, in most cases, also neuroscientific – plausibility of the LIDA cognitive architecture, based on the correlation of its mechanisms to processes in the human brain, and based on the successfully reproduced behavioral data falling naturally out of the same computational model, using the same parameter set.

1. INTRODUCTION

1.1 Overview

The main goal of this work was the adjustment of the LIDA (Learning Intelligent Distribution Agent) cognitive architecture's parameters to neuroscientifically plausible values using LIDA-based cognitive software agents. These agents perform psychological experiments with results similar to human subjects, using the same parameter set. Thus, they substantiate the LIDA model's hypotheses about human cognition - since reproducing a variety of different human behaviors using the same mechanism and the same parameter set strengthens the plausibility of said mechanism modeling human cognition (Madl et al., 2011).

Novel results of this work include an adjusted parameter set (which results in LIDA's processes working in comparable time frames to their neural counterparts – see Sections 2 and 4), a novel conceptual model of the attentional blink based on the LIDA model (Section 5), the three cognitive software agents – performing simple reaction time, perceptual continuity, and attentional blink experiments – (Section 4), and a robust visual perception module for the LIDA model (Section 3).

This thesis is structured in the following way. Section 1 will introduce the LIDA cognitive architecture and compare it with other biologically inspired approaches to model human cognition. Section 2 will describe some of the areas and processes in the human brain that correspond to LIDA's modules and phases, and will detail the timing of the action-perception cycle in LIDA and in the brain. Sections 3 will describe the two cognitive software agents using simple non-visual (abstract) perception that have been developed in the context of this thesis, detail their implementations, and compare their behavior with human subject data. In Section 4, a novel vision model for LIDA is described and evaluated. Section 5 describes the attentional blink phenomenon, puts forth a novel LIDA-based model to explain it, and details the implementation and the performance of a cognitive software agent performing this experiment. Section 6 will summarize the results and conclude this thesis.

1.2 The LIDA Cognitive Architecture

Cognitive architectures, such as LIDA (or ACT-R, EPIC, etc. – see Section 1.3), attempt to approximate cognitive processes and to mirror the structure of cognitive systems of autonomous agents, such as humans. If implemented computationally, they can act as a control system for software agents - or for robots - and attempt to produce behaviors mirroring those of humans (Franklin, 2008). To put it in the words of (Sun, 2007):

> A cognitive architecture is a broadly-scoped, domain-generic computational cognitive model, capturing the essential structure and process of the mind, to be used for a broad, multiple-level multiple-domain analysis of behavior.

Cognitive architectures play a major role in cognitive science due to their usefulness in providing detailed and verifiable explanations for cognitive processes and in providing hypotheses that can guide ongoing research (Madl et al., 2011; Sun, 2007; Anderson et al., 1997).

LIDA (Learning Intelligent Distribution Agent) is a cognitive architecture based primarily on Baars' (1988) Global Workspace Theory (GWT), which suggests the existence of a fleeting memory capacity that enables access between brain functions that are otherwise separate. The global workspace can be thought of as *"... a theater of mental functioning. Consciousness in this metaphor resembles a bright spot on the stage of immediate memory, directed there by a spotlight of attention under executive guidance. Only the bright spot is conscious, while the rest of the theater is dark and unconscious"* (Baars, 2005). The hypothesized primary functional purpose of functional consciousness is to integrate, provide access, and coordinate the functioning of very large numbers of specialized networks that otherwise operate autonomously.

1. Introduction

Apart from GWT, LIDA is also compatible with a number of theories in psychology and neuroscience (Madl et al., 2011), including situated (embodied) cognition (Varela et al., 1991, Glenberg & Robertson, 2000), perceptual symbol systems (Barsalou, 1999), working memory (Baddeley & Hitch, 1974), memory by affordances (Glenberg, 1997), long-term working memory (Ericsson & Kintsch, 1995), transient episodic memory (Conway, 2002), and Sloman's H-CogAff cognitive architecture (Sloman, 1999).

The LIDA model puts forth a number of hypotheses, which are compatible with current neuroscience research (Franklin, 2008). The following hypotheses are most important for the scope of this thesis:

- *The Cognitive Cycle Hypothesis*
  According to the LIDA model, human cognition is comprised of multiple, cascading cycles of perception, attending and action selection (see Section 1.3). These cycles are comparable to the action-perception cycles in neuroscience (Baars & Franklin, 2007; Freeman, 2002; Fuster, 2002; Halgren et al., 2002) and consist of multiple modules, described in Section 1.5.

- *The Consciousness is Discrete Hypothesis*
  Functional consciousness is hypothesized to be discrete in the LIDA model: *"conscious events occur as a sequence of discrete, coherent episodes separated by quite short periods of no conscious content"* (Franklin et al., 2005). *"Similar to the frames of a movie, the 'frames' of consciousness are discrete but are experienced as being continuous"* (Madl et al., 2011).

- *The Theta-Gamma Coupling Hypothesis*
  Functional consciousness is hypothesized to be facilitated by large-scale theta-gamma synchronization in the brain (see the next section for details). Thus, LIDA's cognitive cycles have to occur approximately at theta rates.

- *The LC-NE Hypothesis*
  Observed attentional deficits that occur short timeframes after conscious processing of a stimulus (e.g. in the Attentional Blink) are hypothesized to be due to a refractory-like period in the locus coeruleus (LC) after a target has been perceived. The LC has been implicated in facilitating attentional enhancement of targets through release of norepinephrine (NE) in widespread cortical projection areas. The LC's refractory-like period is implemented in the Attention Codelet Module activation (see Section 5).

The first three hypotheses have been put forth before (e.g. Franklin et al., 2008), and the agents developed in the scope of this thesis attempt to strengthen them. The last hypothesis (the LC-NE hypothesis), as well as the model of the attentional blink based on it, are put forth in this work for the first time in connection with the LIDA model (although the LC-NE system has been used in modeling attentional phenomena before, these models were significantly less comprehensive than the LIDA model – see Section 5).

Since LIDA is the first comprehensive cognitive architecture claiming to provide neuroscientifically plausible explanations - and partly implemented computational models - of some aspects of consciousness (see Section 1.4), some explanations of which processes in the brain are modeled are necessary. These neuronal processes will be described in Section 2.

When talking about LIDA, the *conceptual* and the *computational* LIDA models have to be distinguished. The *conceptual* LIDA model partially specifies the adaptive algorithms modeling processes in the brain based on the Global Workspace Theory (Baars & Franklin, 2009), and includes explanations for a broader range of cognitive phenomena - such as, for example, a self system (Ramamurthy & Franklin, 2011). The *computational* LIDA model completely specifies and implements these processes and algorithms. In this thesis, we will focus on the *computational* LIDA model.

The conceptual LIDA model can be argued to be biologically plausible, since it is based largely on neuroscientific theories (Baars & Franklin, 2009; Goertzel et al., 2010). However, the biological plausibility of some of the computational parts of the agents described in this work is unclear, or even doubtful:

- Some of the mechanisms of the first two cognitive software agents, especially their low-level perception and action mechanisms, have not been implemented in a neuroscientifically realistic way, since their respective environments are extremely simple. This is especially true of the sensation of the environment, which instead of using a model of human vision only uses an abstract feature detector (See Section 3).
- The third cognitive software agent does use an invariant feature detection-based vision model. However, even though the feature detection mechanism and the perceptual representations are based on neuroscientific theories, the features used are not extracted in a biologically plausible way; scale, position, and rotation invariance are ensured based on a computer vision algorithm (Section 4).

For these reasons, I am only aiming for *psychological* plausibility instead of full neuroscientific plausibility in the agent implementations. See Section 1.5 for the internal structure of the LIDA cognitive cycle, and Section 3 for a description of the computational implementation.

1.3 Comparison with other Cognitive Architectures

The number of proposed cognitive architectures is too large to exhaustively describe in the scope of this thesis. For this comparison, I will focus on well-known and biologically plausible architectures that have computational implementations (based on Goertzel et al. (2010)), and will provide only short descriptions for the purpose of conceptual comparison with the LIDA model. Models that are neurally implausible or have not (yet) shown evidence of being capable of higher-level cognitive capabilities are described only superficially in order to not exceed the scope of this thesis. More comprehensive reviews can be found in (Goertzel et al., 2010) and in (Duch et al., 2008); and (Gluck & Pew, 2005) as well as (Vernon et al., 2010) contain descriptions and evaluations of a large number of cognitive architectures.

Figure 1 below provides an overview about the biologically inspired cognitive architectures which will be summarized here, grouped into symbolic (a paradigm stating that minds manipulate symbols representing aspects of the world or themselves), emergent (a subsymbolic paradigm simulating aspects of brain function, such as, for example, neural networks) and hybrid approaches (combining subsystems operating according to these two paradigms) (Goertzel et al., 2010).
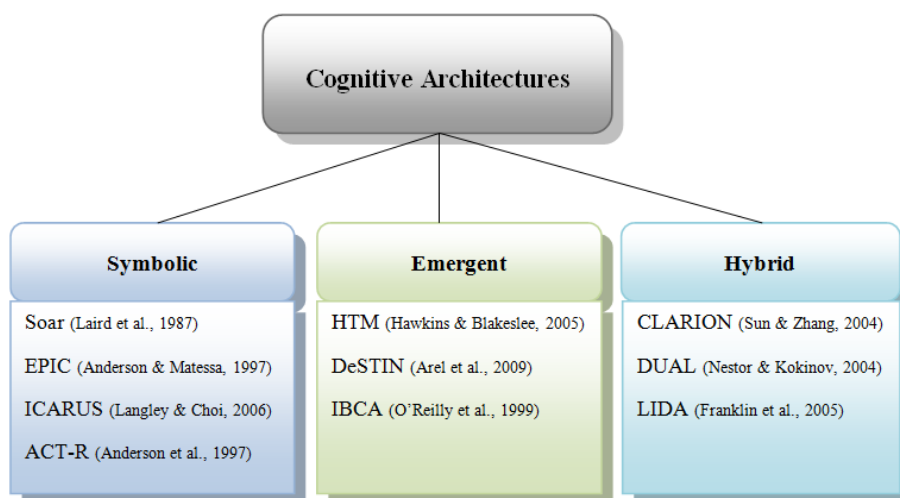
Figure 1. Biologically-inspired cognitive architectures.

1.3.1 Symbolic Cognitive Architectures

Soar (Laird et al., 1987) and **EPIC** (Executive-Process/Interactive Control - Anderson & Matessa (1997)) are symbolic and production-rule based cognitive architectures inspired by human cognitive mechanisms but lacking biological plausibility. The symbolic architecture **ICARUS** (Langley & Choi, 2006) contains mental structures grounded in perception and action; knowledge is specified in reactive skills. However, it does not address biological plausibility either; and some of its cognitive mechanisms are not sufficiently implemented (concurrent processing, attention, dealing with uncertainty) (Goertzel et al., 2010).

ACT-R (Adaptive Control of Thought – Rational, Anderson et al. (1997)), similarly to Soar, follows a production-rule based approach and is primarily symbolic. It utilizes two types of memory: declarative memory, encoding factual knowledge about the world, and procedural memory, containing procedural knowledge in the form of IF-THEN rules. The general usefulness of these chunks and production rules is stored in connectionist structures reflecting previous usage (which has led some researchers to categorize ACT-R as a hybrid cognitive architecture, despite it being primarily symbolic – Duch et al. (2008)).

Apart from memory, the central components of ACT-R are perceptual-motor modules interfacing with the environment, buffers, and a central pattern matcher for productions (matching, selecting and executing production rules). This central module is hypothesized to correspond to the basal ganglia in the brain. Other tentatively proposed neural correlates include the occipital visual cortices for the visual module, the motor cortex and the cerebellum for the motor module, the temporal lobe and the hippocampus for the declarative memory module and various prefrontal and parietal areas for the buffers. ACT-R has been used to replicate a large number of psychological experiments (Anderson et al., 2004).

1.3.2 Emergent Cognitive Architectures

A problem common to most purely emergent approaches is that while they seem to perform well in pattern recognition and associative learning, no one has yet been able to show that they are capable of higher-level cognitive functions (e.g. reasoning, language) (Goertzel et al. 2010). For this reason, these approaches will only be briefly summarized here.

HTM (Hierarchical Temporal Memory) (Hawkins & Blakeslee, 2005) is a pattern recognition approach based on a combination of a model of the cortex and an artificial intelligence algorithm. **DeSTIN** (Deep SpatioTemporal Inference Network) (Arel et al., 2009) is a similar pattern recognition mechanism, which also contains hierarchical networks dealing with action and reinforcement. Both have so far only been used for vision processing.

IBCA (Integrated Biologically based Cognitive Architecture) (O'Reilly et al., 1999) is a neural network-based model of distributed information processing in the brain, aiming to model especially the posterior and frontal cortices and the hippocampus. It has been used in simple psychophysical experiments but has not (yet) been shown to be able to model higher level cognitive processes like reasoning or subgoaling (Goertzel et al., 2010).

1.3.3 Hybrid Cognitive Architectures

Hybrid architectures attempt to make use of the advantages of both the symbolic and the emergent paradigm by combining subsystems of both paradigms.

CLARION (Sun & Zhang, 2004) incorporates explicit (symbolic) as well as implicit (subsymbolic) knowledge through its four memory modules: the action-centered subsystem (regulating actions), non-action-centered subsystem (maintaining general declarative knowledge), motivational subsystem (providing motivation for action), and metacognitive subsystem (monitoring and directing the operations of the other subsystems). Each module has a localist-distributed representation (explicit knowledge) and a connectionist distributed section

(implicit knowledge). Implicit knowledge in the connectionist parts (e.g. multi-layer perceptron networks) is learned via reinforced or supervised learning, while explicit knowledge is acquired from these networks in a bottom-up fashion – although top-down learning is also possible (Duch et al., 2008). CLARION was shown to be able to reproduce large amounts of psychological data (Sun & Zhang, 2004).

DUAL (Nestor & Kokinov, 2004) is a cognitive architecture inspired by Minsky's (1988) Society of Mind and acts upon perceived stimuli using a large number of micro-agents. These agents represent facts in a symbolic way, using micro-frames. However, they also have activation levels, and they spread activation to neighboring agents – thus, they are components of a connectionist network. Agents communicate via learned links and can thus form coalitions representing concepts or episodes. The DUAL framework has been used to model cognitive functions such as perception or analogy (Goertzel et al., 2010); however, it is not clear whether and how the system will be able to model high-level functions such as deliberation or reasoning.

Finally, the **LIDA** model is also a hybrid system as it combines connectionist networks (e.g. the Perceptual Associative Memory or the Procedural Memory – which however operate on a higher than neuronal level) with symbolic processing throughout the cognitive cycle; explained in more detail in the following sections; see also (Franklin et al., 2005; Franklin et al., 2007; Franklin & Patterson, 2006; or Baars & Franklin, 2009).

The greatest advantages of the LIDA cognitive architecture in comparison to the previously described approaches are its comprehensive account for a wide range of cognitive processes, a formal explanation for functional consciousness based on modern neuroscience, and its high biological plausibility (Baars & Franklin, 2009). Other advantages include a large variety of learning mechanisms, including constructivist procedural learning (Franklin & Patterson, 2006) as well as support for higher-level, multi-cyclic cognitive processes such as deliberation, voluntary action, and automatization (Franklin et al. 2007) and non-routine problem solving (Franklin et al. 2007; Negatu et al., in press).

Disadvantages of the LIDA model compared to some of the more developed architectures (e.g. ACT-R) include that the computational implementation is still in a very early stage (the first beta version was released in June 2011), thus not all parts of the conceptual model are incorporated yet; also, no appropriate biologically plausible visual system has been implemented (although a simplified object recognition system based on invariant feature detection was developed in the scope of this work – see Section 4). Another aspect that has been criticized is that some higher-level cognitive processes, while conceptually defined (see e.g. Franklin (2008)), are not exactly specified nor implemented yet (e.g. language or reasoning).

1.4 Consciousness in the LIDA Model and in the Brain[*]

Unlike implementations of other cognitive architectures described in the previous section, LIDA agents can be said to be functionally conscious, as argued below. However, the LIDA model does not account for, explain, or implement phenomenal consciousness or qualia (the current scientific understanding thereof does not even suffice for a complete conceptual explanation). Self-consciousness is not computationally implemented either, although the model includes conceptual approaches for this aspect (Ramamurthy & Franklin, 2011). Thus we shall focus on functional consciousness in this work. For a discussion of various types of consciousness, see (Block, 1995).

As mentioned above, the LIDA model is based on the global workspace theory of consciousness (Baars, 1988), which suggests the existence of a fleeting memory capacity that enables access between brain functions that are otherwise separate. The global workspace theory (GWT) can be thought of as "… a theater of mental functioning. Consciousness in this metaphor resembles a bright spot on the stage of immediate memory, directed there by a

---

[*] *Sections marked with an asterisk have been published before in (Madl et al., 2011)*

spotlight of attention under executive guidance. Only the bright spot is conscious, while the rest of the theater is dark and unconscious" (Baars, 2005). In case of sensory consciousness, the stage corresponds to the sensory projection areas of the cortex, its activation coming either from senses or from internal sources. After a conscious sensory content is established, it is distributed to a decentralized "audience" of expert networks sitting in the darkened theater. Thus, the primary functional purpose of consciousness is to integrate, provide access, and coordinate the functioning of very large numbers of specialized networks that otherwise operate autonomously. In the neuroscientific study of consciousness, this idea of consciousness having an integrative function has proven very useful, and is supported by much recent evidence (Baars, 2005; Tononi, 2004; Dehaene, 2008) (see also the Results section).
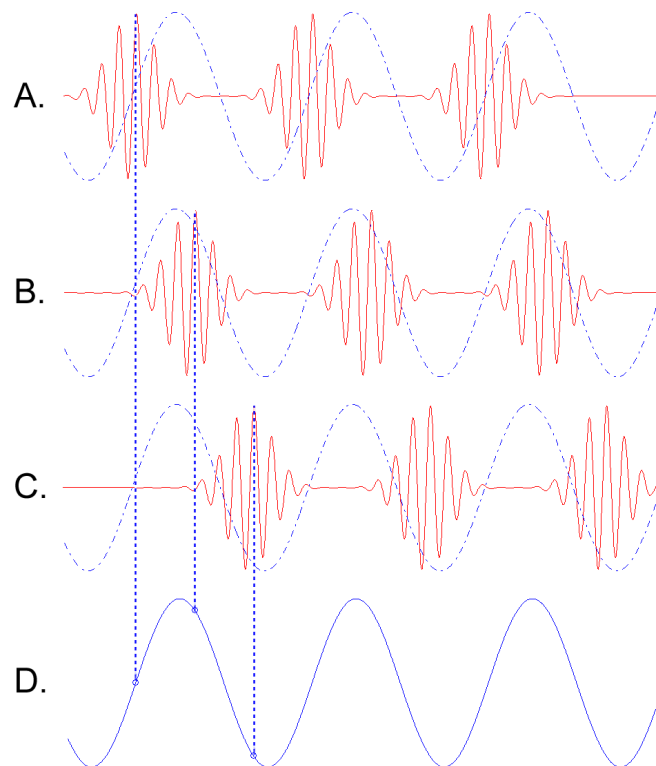
*Figure 2. **Theta-gamma coupling.** Three gamma cycles are sequentially "embedded" in a theta cycle. (A),(B), and (C) depict the temporal activity pattern of three different neuronal assemblies oscillating in the gamma range. Each is phase-locked to the underlying theta rhythm with a different phase offset, as indicated by the dashed lines. This type of coupling is known as phase-amplitude coupling, because the amplitude modulation of each gamma pattern is locked to a particular phase of the theta pattern (S). From (Strain et al., 2010)*

In LIDA, every cognitive cycle can have only a single conscious "frame" (content) at a time, a hypothesis compatible with recent neuroscientific publications which view consciousness as large-scale phase synchronization of neuronal activity (Doesburg et al, 2009; Strain et al., 2010; Buzsáki, 2006; Varela et al., 2001). In this view, the complex rearrangement of neural populations across widespread and diverse cortical regions, which is required for consciousness, is accomplished by oscillatory dynamics; specifically, by theta-gamma coupling between the neural populations (see Figure 2).

Performing cognitive tasks modulates oscillatory brain activity in various frequency bands, including both the theta (4-7 Hz) and gamma (30-150Hz) bands. Gamma-band phase synchrony (Figure 3) has been associated with

perceptual binding and awareness. Numerous studies have observed the occurrence of gamma activity coherence with perceptual (Freeman, 1999; Gray & Singer, 1989) as well as long-term (Osipova et al., 2006) and working-memory-related (Buzsáki, 2006) object representations. Synchronized gamma-band oscillatory activity has also been shown to play an important role in the coding of short-term memory information (Siegel et al., 2009; Schack et al., 2002; Lisman & Idiart, 1995). Moreover, modulation of gamma activity has been demonstrated in attentional selection (Jensen et al., 2007; Tallon-Baudry et al., 2005; Tiitinen et al., 1993), and phase-locked gamma synchrony between ascending and descending systems in a sensorimotor task (Buzsáki, 2006). Many of these studies have observed that activity across different cortical columns representing the percept of an object is gamma synchronized (e.g. (Gray & Singer, 1989)). Thus, the neuronal ensembles responsible for various cognitive processes involved in the processing of a percept, taking place during a cognitive cycle, operate at and are integrated by an internal oscillation frequency in the gamma band.
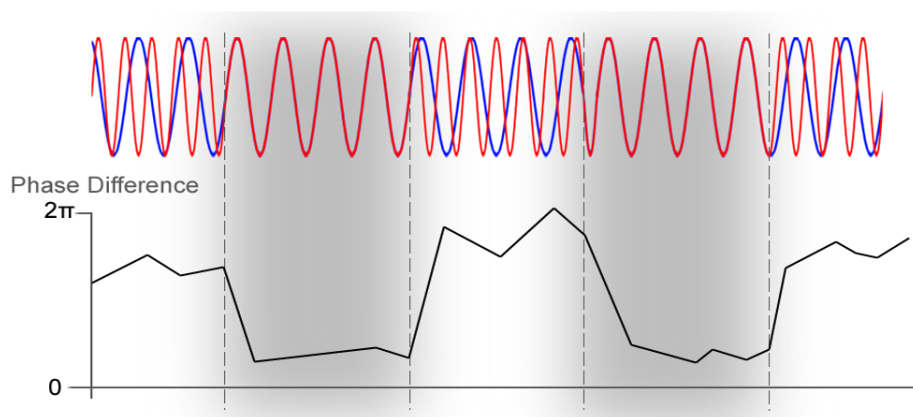
Figure 3. Phase synchrony between two oscillations.
The upper part shows two oscillations (in red and blue), and the lower part their phase-differences. In the two gray areas framed by dotted lines the oscillations are highly phase synchronous and the phase differences are low. Such phase-synchrony in the gamma band has been proposed to be responsible for perceptual binding (for example, cortical columns representing the same object are gamma synchronized)

The construction of such gamma-synchronous neural ensembles has been claimed to be governed by theta-rhythms (Doesburg et al., 2009; Canolty et al., 2006). This might be the integration mechanism required for consciousness: in this view, consciousness emerges from large-scale functional integration of these gamma-synchronous ensembles that form and dissolve at the theta frequency band (Doesburg et al., 2009).

Only one perceptual experience can be contained in a single phase of theta-modulated gamma-synchrony (Doesburg et al., 2009), consistently with the attentional blink ((Shapiro et al., 1997), see also Results section) and other studies of perceptual synchrony (Rodriguez et al., 1999). This indicates that these phases of synchrony define discrete 'frames' of consciousness, which, in the LIDA model, correspond to cognitive cycles (Baars, 2009; Strain et al., 2010). An approximate lower time limit for a single cognitive cycle can already be deduced from this hypothesis. Since each cycle is concerned with a single conscious content, and a new conscious content requires theta-gamma synchronization, conscious processing in the cognitive cycles has to occur at theta rates (4-7 Hz). Therefore cognitive cycles have to take at least 140-250ms. However, since cognitive cycles can cascade as long as they preserve the seriality of consciousness, they could take longer than that (see Results section).

An important hypothesis of the LIDA model is the discreteness of consciousness. Humans can only have a single conscious content at a time, and there are short breaks between these periods of consciousness. In the words of Franklin et al. (2005), "conscious events occur as a sequence of discrete, coherent episodes separated by quite

short periods of no conscious content" (see also (Van Rullen & Koch, 2003)) - similar to the frames of a movie, the 'frames' of consciousness are discrete but are experienced as being continuous (although this analogy is not entirely accurate).

This view is consistent with the idea of consciousness emerging from theta-gamma coupling. Gamma-oscillatory neural ensembles are synchronized as well as desynchronized at theta rates. The transient periods of desynchronization, also called phase scattering, reflect unconscious processing in the brain, thus "ending each 'frame' of (conscious) perceptual experience" (Doesburg et al., 2009). These periods of desynchronization have also been observed, and pointed out, to play a role in the transition from one cognitive content to another by (Rodriguez et al., 1999; Van Rullen & Thorpe, 2001; Thompson & Varela, 2001; Doesburg et al., 2008; Raffone et al., 2009). (For more neuroscientific results about consciousness see Section 2 below). In psychology, Stroud (1967) was one of the first authors to propose the idea of discrete frames or 'moments' underlying consciousness. His 'Discrete Moment Hypothesis' included two important underlying assumptions: a) a complete loss of time-order information within one conscious 'moment', and b) a distinct and non-overlapping set of percepts for each 'moment'. This strict view of discrete consciousness has been regarded with some skepticism. Allport (1968), for instance, has conducted experiments on phenomenal simultaneity, which seem to contradict the Discrete Moment Hypothesis – they are, however, compatible with LIDAs consciousness model, as can be seen from the Results section, in which we replicated the data from Allport's experiment using a LIDA-based agent.

In the LIDA model, single conscious episodes are discrete but, contrary to Stroud's (1967) view, not necessarily distinct – a current conscious 'moment' can contain percepts from a previous moment. Whether or not an older percept remains conscious depends on how long in the past it has been perceived, and on attentional modulation – percepts that are subjectively important and attended to can persist longer in consciousness.
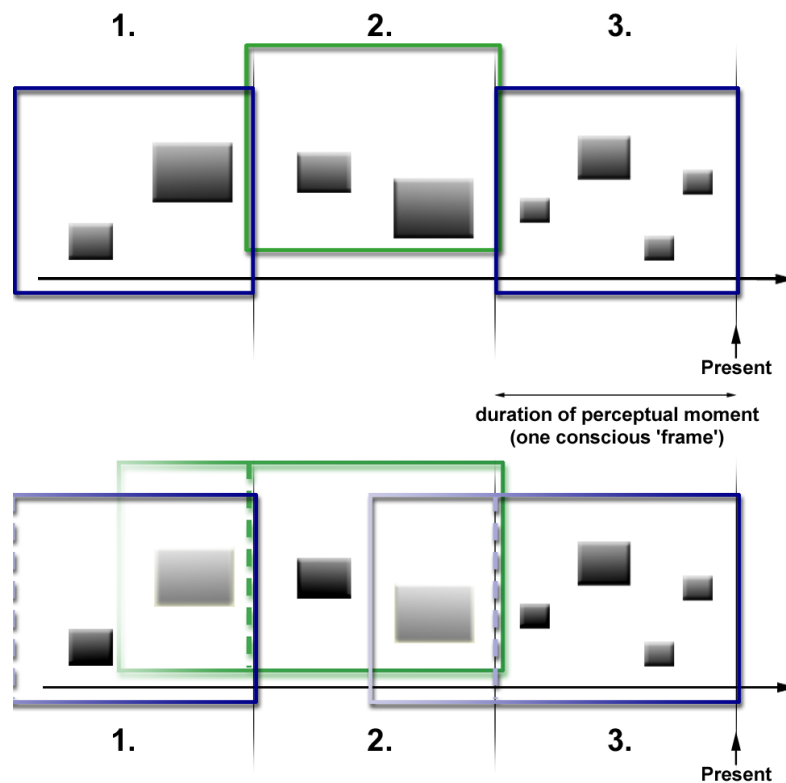

Figure 4. Schematic comparison of the Discrete Moment Hypothesis (top) and LIDA's discrete consciousness hypothesis (bottom). The colored frames represent the temporal constraints of a perceptual moment or conscious

'frame', and the black rectangles symbolize incoming percepts. In LIDA, important percepts from previous conscious 'frames' can remain conscious (rectangles left of the dashed lines in the colored frames in the bottom picture).

To improve our earlier movie analogy, the 'frames' of consciousness in the LIDA model could be compared to a movie shown on a phosphor-based electronic display (CRT): although the frames are discrete, new images on the screen contain past information (see Figure 4). As we will see in Section 3.3, this approach resolves the empirical contradictions of the Discrete Moment Hypothesis.

Since our timing model was largely derived from neuroscientific experiments, some tools and techniques these experiments might use, and the reasons we preferred to use the results of some experiments over others, should be described.

Electroencephalography (EEG) records electrical activity from neural field generators using several electrodes placed on the scalp surface. Recent research concentrates on aspects of this electrical activity time-locked to events, i.e. event-related potentials (ERP), which occur in preparation of or in response to discrete (internal or external) events. We have used EEG experimental results because EEG has great temporal resolution (on the order of milliseconds), and a large number of EEG results are available. Disadvantages of EEG are its low spatial resolution (typically 2 – 3cm in surface tangential directions) and the fact that it only measures synaptic activity from superficial cortical layers (Nunez & Srinivsasan, 2005).

Transcranial magnetic stimulation (TMS) experiments involve stimulating the brain using induced electric currents, which trigger action potentials in the neurons in the current field, disrupting ongoing brain activity (causing temporary "virtual lesions"). We also used TMS experiments because TMS resolutions are very good (temporal resolution on the order of milliseconds, spatial resolution on the order of a few millimeters, depending on the coil shape). Disadvantages of TMS are the impossibility to determine exactly how much area is affected by these induced currents. Also, TMS cannot stimulate regions deeper than the cortex without stimulating the cortex.

The most exact technique measuring brain activity is using depth electrode and subdural grid recordings. Depth electrode recordings are mostly performed on animals and clinical patients. Subdural grid recordings (also called electrocorticograms or ECoG), involving the placing of electrodes directly on the brain surface, are less invasive and have spatial resolution somewhere between depth electrodes and EEG. These techniques provide the most exact and reliable data, but they require surgery and cannot be used in healthy humans (Buzsáki, 2006).

The reason we have not used experiments relying solely on functional magnetic resonance imaging (fMRI) data is that this technique measures blood oxygen levels, and it takes several minutes for the bloodstream in active brain areas to become oxygenated (Van der Zwaag et al., 2009), which is well outside our time scale.

A more complete and detailed review of non-invasive brain imaging techniques can be found in (Shibasaki, 2008).

1.5 The LIDA Cognitive Cycle[*]

Autonomous agents (Franklin & Graesser, 1997) cope with their changing environment by their continuous, cyclic chores of 'perceive-understand-act'. LIDA's cognitive cycle (Franklin et al., 2005) is the cycle of refined cognitive processes (starting after sensation and ending with action) that bring about the appropriate action for specific situation. As Franklin and Baars (2009) put it: *"A cognitive cycle can be thought of as a moment of cognition - a cognitive moment; higher-level cognitive processes are composed of many of these cognitive cycles, each a cognitive atom."* This metaphor is to say that the steps in a cognitive cycle correspond to the various sub-atomic particles in an atom.

1. Introduction

Since the LIDA architecture is composed of several specialized mechanisms, a continual process that causes the functional interaction among the various components is essential. The cognitive cycle as such is an iterative, cyclical, continually active process that brings about the interplay among the various components of the architecture. The steps of cognitive cycle are shown in Figure 5 and will be described below. It is important to point out the asynchrony of the LIDA cognitive cycle. Cycles can cascade as long as they preserve the seriality of consciousness. Furthermore, the components of the cognitive cycle described below should not be seen as serial stages of information processing. The components operate asynchronously - although coordinated, each component has its own internal mechanism and agenda. Components receiving inputs from others are not triggered by those inputs, but rather run continuously at their specified frequencies of operation (See Methods section).
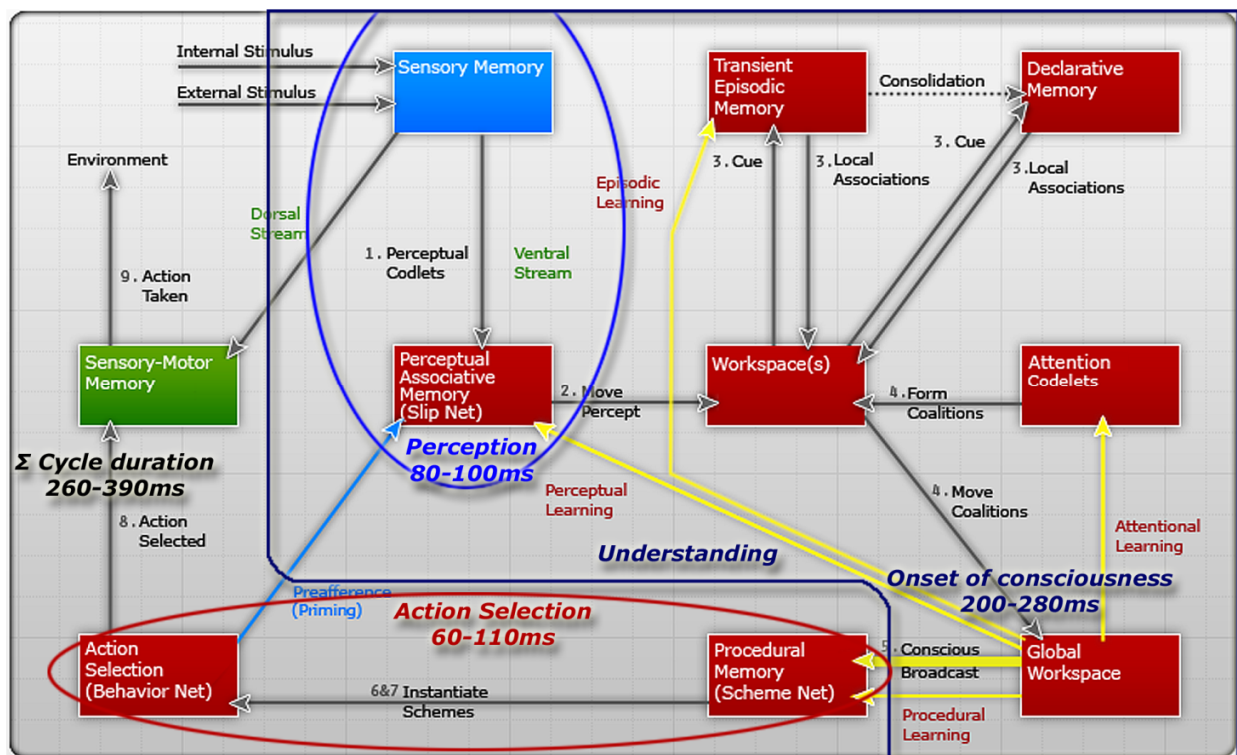
Figure 5. The LIDA cognitive cycle, and the durations of the perception, understanding and action phases. *(Modified from Baars & Franklin (2009))*

During each cognitive cycle the LIDA agent, be it human, animal or artificial, first senses its environment and tries to recognize familiar objects, individuals, etc (perception phase). It then associates percepts with memories and other percepts and decides what portion of this situation is most in need of attention (understanding phase). Broadcasting this portion (bringing it to consciousness) enables the agent to choose a number of actions applicable for the current situation and to select the action best serving its goals (action selection phase), and to finally execute the selected action. The cognitive cycle has the following components:

1) **Perception**. Sensory stimuli, external or internal, are received and interpreted by perception producing the beginnings of meaning.

2) **Percept to preconscious buffer**. The percept, including some of the data plus the meaning, as well as possible relational structures, is stored in the preconscious buffers of LIDA's working memory (workspace). Temporary structures are built.

3) **Local associations.** Using the incoming percept and the residual contents of working memory, including emotional content, as cues, local associations are automatically retrieved from transient episodic memory and from declarative memory, and stored in long-term working memory.

4) **Competition for consciousness**. Attention codelets view long-term working memory, and bring novel, relevant, urgent, or insistent events to consciousness.

5) **Conscious broadcast**. A coalition of codelets, typically an attention codelet and its covey of related informational content, gains access to the global workspace and has its content broadcast consciously. Thus consciousness solves the relevancy problem in recruiting resources.

6) **Recruitment of resources**. Relevant schemes in Procedural Memory respond to the conscious broadcast. These are typically schemes (underlain by behavior codelets) whose context is relevant to information in the conscious broadcast. Thus consciousness solves the relevancy problem in recruiting resources.

7) **Setting goal context hierarchy**. The recruited schemes use the contents of consciousness, including feelings/emotions, to instantiate new goal context hierarchies (copies of themselves) into the Action Selection system), bind their variables, and increase their activation. Other, environmental, conditions determine which of the earlier behaviors (goal contexts) also receive variable binding and/or additional activation.

8) **Action chosen**. The Action Selection module chooses a single behavior (scheme, goal context), from a just instantiated behavior stream or possibly from a previously active stream. Each selection of a behavior includes the generation of an expectation codelet (see the next step).

9) **Action taken**. The execution of a behavior (goal context) results in the behavior codelets performing their specialized tasks, having external or internal consequences, or both. LIDA is taking an action. The acting codelets also include at least one expectation codelet whose task it is to monitor the action, bringing to consciousness any failure in the expected results.

As shown in Figure 5, multiple learning mechanisms are initiated following the broadcast of conscious content. In the perceptual associative memory learning of new entities and associations, and the reinforcement of old ones occur, events are encoded in the Transient Episodic Memory, and new schemes may be learned and old schemes reinforced in Procedural Memory; in all of the learning processes, the conscious content determines what is to be learned. LIDA's modules are described in more detail in Section 3. For more information about the LIDA model and its cognitive cycle see (Franklin et al., 2005; Baars & Franklin, 2009).

2. THE TIMING OF THE COGNITIVE CYCLE[*]

2.1 Overview[*]

As mentioned above, cognition in autonomous agents (Franklin & Graesser, 1997), whether artificial, animal or human, can be thought of as consisting of repeated perception-understanding-action cycles. In these cycles, actions can be external (effecting changes in the environment) or internal (effecting changes in internal representations or processes). Similarly, perceptual information can come from external (from senses sensing the environment) or internal sources. Complex tasks may require many of these cycles before an external action can be taken.

Figure 6 below shows such a cognitive cycle, including its three sub-processes. For the durations of these sub-processes, see Figure 7.

2. The Timing of the Cognitive Cycle*
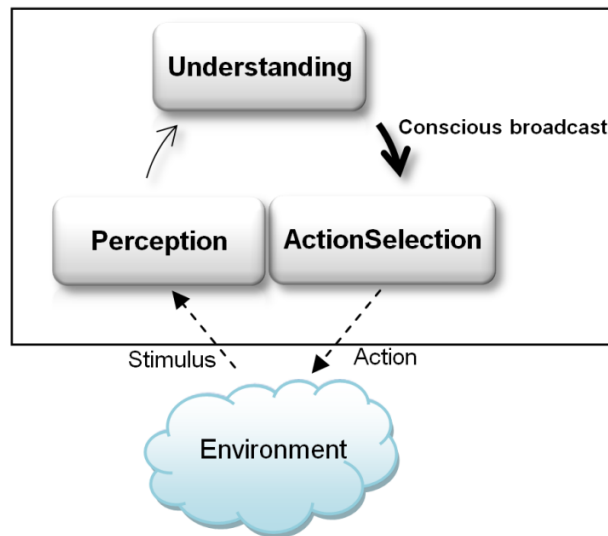
Figure 6. The three phases of the LIDA cognitive cycle.

A stimulus comes in from the environment via the senses. The perception sub-process includes obtaining this data, detecting features, and recognizing objects, categories and events. The understanding sub-process includes making sense of the perceived information and selecting the most relevant, urgent or novel information, which is included in the conscious broadcast (the agent is only consciously aware of the contents of this broadcast). Finally, the action selection sub-process selects the action best serving the agent's goals, based on the conscious broadcast contents.

The understanding phase in this cognitive cycle is frequently called 'cognition' in other cognitive models (e.g. (Anderson et al., 2004; Meyer & Kieras, 1997)). In LIDA, the term 'understanding' is more appropriate because the integration of percepts, the building of associations (with memories and with other percepts) and assessments of subjective significance that take place during this phase all contribute to a representation or situational model (stored in temporary memory, the workspace) which is best described as the agents current understanding of its immediately perceived environment (see Introduction). In other cognitive models, such as ACT-R or EPIC, the cognition phase includes the matching, selection and execution of production rules (Anderson et al., 2004; Meyer & Kieras, 1997).

Figure 7 shows our hypothesized durations for the sub-processes of the cognitive cycle in humans. The next subsections will describe neural equivalents of these sub-processes and provide supporting evidence for the indicated durations. The indicated ranges should not be taken as precise and definite values; rather, they are working estimates derived from recent evidence.

It should be pointed out that the experiments on which these durations are based used very simple settings and stimuli, and in most cases, they did not involve memory recall. For tasks involving the use of memory, the time from stimulus presentation to action execution can be significantly longer than the times indicated here (Healy et al., 2003). However, for most simple tasks, due to the large extent of consistency between these results and various psychological and neuroscientific experiments (see below), we believe that the indicated durations of these processes accurately reflect some of the temporal properties of human cognition.
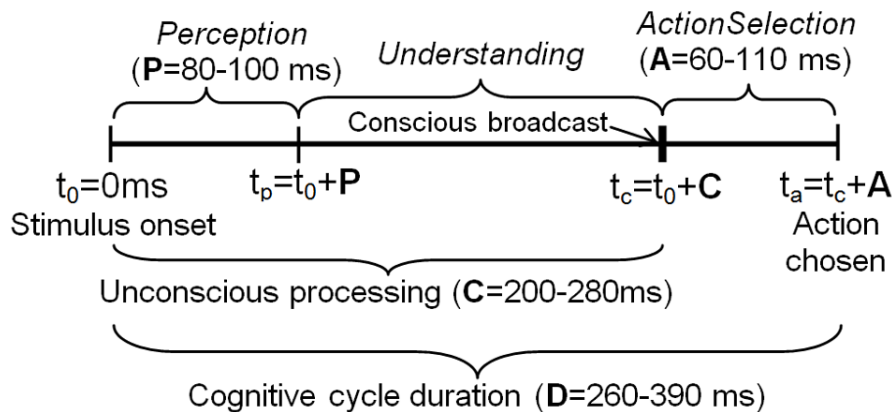
Figure 7. The timing of a single cognitive cycle. *The perception sub-process is estimated to take P=80-100ms, the time until conscious processing C=200-280ms, the action selection sub-process A=60-110ms, and the entire cognitive cycle is hypothesized to take D=260-390ms.*

2.2 Perception[*]

The perception process includes obtaining data from the environment via sensors, detecting features, and recognizing more abstract entities such as objects, events and categories.

In humans, perceptual information can come from different sensory modalities. The most researched and perhaps most complex modality (judging from the size of cortical areas associated with its processing) is visual perception (Bear et al., 2007).

Visual perception starts with an image of the environment on the photoreceptive cells of the retina, which produces neural impulses that are transmitted along the retinofugal projection to the visual cortex, which is located in the occipital lobe, where most of the processing of visual information takes place (Bear et al., 2007).

We have estimated the duration of the perception process in humans for simple tasks to be approximately in the range of P = 80 – 100ms (see Figure 6). For instance, an experiment by Liu et al. (2009), performed using intracranial electrodes in epilepsy patients, has shown that object category information can be decoded from neural activity in the occipital lobe as early as 100ms poststimulus. This is consistent with EEG experiments trying to temporally localize object-selective brain activity, most of which found that the P100 ERP component (90 – 115ms post stimulus) is already associated with object information (Schendan & Lucia, 2010). It is also consistent with the result of various studies of visual processing which have determined that a stimulus presentation time of 100ms is sufficient for recognizing traits and properties (Willis & Todorov, 2006; Agam et al., 2010). Finally, this duration was also indicated by TMS experiments investigating in which time range TMS interferences with the visual system can impede vision. Such experiments found that the range of greatest impairment was between 80 and 100ms, and that TMS interference after 100ms had little to no effect on visual perception (Stewart et al., 2001; Walsh & Cowey, 1998).

This perceptual duration seems to provide an appropriate upper limit for the perception process in general, since information from other modalities is processed in this range or even faster in the human brain. For example, auditory (and somatosensory) event related responses in the sensory cortices can commence in less than 50ms (Lakatos et al., 2007), and the entire auditory neural representation can be built during the N1 stage in ~100ms (Besle et al., 2008; Näätänen & Winkler, 1999).

2.3 Cognitive Processing and Consciousness[*]

According to the LIDA model and GWT (see Section 1), a major functional role of consciousness is to distribute important perceptual information to different, specialized brain areas. (Novel Hypothesis 5 in (Franklin et al., 2005)) It is possible to derive a way to measure the elapsed time between the sensing of a stimulus and its becoming a conscious event from this hypothesis. Unconscious processing of the stimulus appears to be more localized in sensory areas (e.g. the visual cortex for visual stimuli), meaning that these areas have the highest activity in the unconscious processing stage. Conscious processing can be said to start at the moment other brain areas, for example those involved in decision making / action selection (e.g. pre-frontal areas, see next section), become highly active – this information can be derived from fast brain imaging techniques.

There are experimental indications that this distribution of information, termed the conscious broadcast (Baars & Franklin, 2009) commences about 200-280ms post stimulus (Figure 6).

For substantiating the claim of when conscious processing starts, comparisons of conscious and non-conscious processing of the same stimulus are sometimes used. There are a number of such neuroscientific experiments that yield useful timing results from this point of view. Gaillard et al. (2009) have conducted an intracranial iEEG experiment using a visual masking procedure, performing trials with and without conscious visibility of masked words (with and without showing a mask very shortly after presenting the word), concluding that conscious processing takes place 200 – 300ms post stimulus. Other studies using EEG and also using a masked visual paradigm indicated conscious processing to commence at 270ms (Del Cul et al., 2007; Sergent et al., 2005) (see also the survey about conscious and unconscious processing in (Dehaene, 2008)). An MEG study using a different visual paradigm (subjects had to decide whether a cue – a faint circular grating – has been present or absent during stimulus presentation) concluded 240ms post stimulus as the onset of awareness-related activity (Wyart & Tallon-Baudry, 2008). A different MEG study yielded similar results, for both auditory and visual conscious perception of novel words (Marinkovic et al., 2003).

Another approach to determining the onset of conscious processing is by calculating the amount of theta-gamma phase synchrony from brain oscillatory data (see Section 1).

A binocular rivalry experiment using EEG recordings conducted by Doesburg et al. (2009) provides supporting evidence for this hypothesis. Doesburg et al. found that gamma-oscillatory networks across the brain, formed and dissolved at the theta frequency band, are time-locked to perceptual switching (they are time-locked to which of the two stimuli the subject is aware of). On a spectral diagram of their results they could identify the times in which the subject was aware of one or the other stimulus, signified by high levels of theta-gamma phase synchronization. The resulting time until one of the stimuli became conscious was 260-380ms (the temporal distance between the subject being consciously aware of the first and then the second stimulus).The lower time limit is consistent with a previous experiment by the same authors (Doesburg et al., 2008), which observed maximal phase synchrony 220-280ms post stimulus. It is also consistent with the iEEG, EEG and MEG studies described above.

The so called "Visual Awareness Negativity" (VAN), an ERP component defined by the difference between ERPs to conscious versus unconscious stimuli, also fits well into these time ranges, since the part of VAN that is affected by attentional selection occurs at 200 – 260ms (Koivisto et al., 2009).

Finally, all the results above are to some extent consistent with the time frame of the attentional blink (Sergent et al., 2005; Shapiro et al., 1997). In attentional blink experiments, two masked visual stimuli are presented in short succession. For short stimulus onset asynchronies, the identification of the first target hinders the detection of the second target (although the second target is easily seen if the temporal distance between the two targets is

increased). The worst identification performance of the second stimulus has been observed at delays of about 225ms between the onsets of the two stimuli (Shapiro et al., 1997), which is consistent with the LIDA hypothesis that there can be only one conscious content in one cognitive cycle (Franklin et al., 2005; Baars & Franklin, 2009). This idea is also described by Doesburg et al., who write that after one period of phase synchronization (of the subject being conscious of a stimulus), desynchronization is required before the next period of synchronization; and that during one period of synchronization the subject can be conscious of only one stimulus (Doesburg et al., 2009).

It should be pointed out that for determining the time of the conscious broadcast, only the lower limits of the times determined by these experiments are relevant. Cognitive processes after the times indicated by the upper limits in these experimental results presumably include action selection processes (see next section). Therefore, the time range of the conscious broadcast indicated in Figure 7 has been determined by taking into account only the lower limits of these results: the smallest and the greatest lower limit.

Summarizing, consciousness seems to involve large-scale integration of different brain areas through phase coupling, and widespread distribution of sensory information. In simple trials, conscious processing has been estimated to commence C = 200-280ms post stimulus (see Figure 7).

2.4 Action Selection / Decision Making[*]

There are several brain circuits involved in action selection, the most relevant being the prefrontal cortex, the pre-supplementary motor area (preSMA), the supplementary motor area (SMA) and the primary motor cortex (M1). Information from the first three areas converges on the primary motor cortex (see Figure 8), which executes motor commands by transmitting them to the spinal cord and muscles (Haggard, 2008). There can be two classes of inputs to M1, voluntary and stimulus-driven inputs.
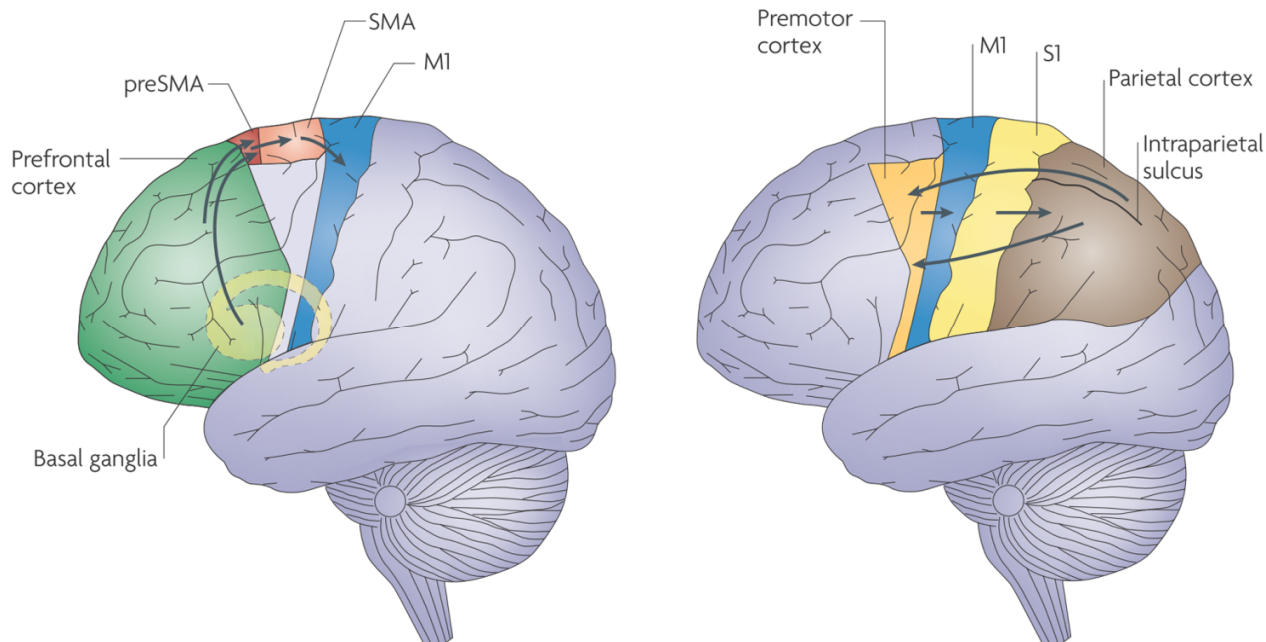

Figure 8. Major brain areas involved in action selection. *The left panel shows the brain areas involved when making voluntary actions; the right panel, object-oriented (stimulus driven) actions. From (Haggard, 2008)*

2. The Timing of the Cognitive Cycle*

The first key input comes into the M1 from the prefrontal cortex by way of the basal ganglia and the preSMA - see the left panel in Figure 8. This circuit is used when making voluntary actions (preSMA activations are stronger for voluntary actions than for stimulus-driven actions).

The second input plays a role in the immediate stimulus-dependent guidance of actions and is projected to M1 from the lateral part of the premotor cortex, which receives its input from the internal representations in the parietal lobe, which in turn are built from information from the sensory cortices (although this circuit also contributes to voluntary behavior) (Haggard, 2008) – see the right panel in Figure 8.

The action selection process begins with receiving the conscious broadcast (Figures 5 and 6), and involves two stages:

- the selection of a number of actions that are applicable, depending on the current situation, i.e. the content of the conscious broadcast (represented by the Procedural Memory module in LIDA) and
- the selection of the best available action, i.e. the action that best serves the goals of the agent (represented by the Action Selection module in LIDA).

This separation of action selection into two stages has also been observed in the brain. The brain begins to prepare several actions in parallel while collecting evidence for selecting between them (Cisek & Kalaska, 2010; Shadlen et al., 2008). For example, in visually guided movement, the first stage involves a reciprocally interconnected network of areas in the posterior parietal and caudal frontal cortex, converting sensory information into parameters of potential actions. Each area can represent information that is simultaneously pertinent to several potential actions. There is a competition between these potential actions, corresponding to stage two mentioned above, which is influenced by a variety of brain areas, most importantly the basal ganglia and the prefrontal cortex (for more details see (Cisek & Kalaska, 2010)).

There are few experimental results concerning the duration of the action selection process; some of them shall be reviewed below.

In an experiment conducted by Nachev et al. (2005), subjects were asked to either follow a specific movement plan or to choose freely between two alternatives in an oculomotor change-of-plan task. After free choice, subjects could be asked to continue their plan or to rapidly change it. Directed trials in which subjects failed to change their planned saccade had latencies 107ms (median) shorter than trials where the plan change was successful, indicating that the process of selecting a different action took 107ms.

Taylor et al. (2007) have used TMS to interfere with preSMA activity, which disrupted subjects' decision whether they should respond with their left or right hand, if applied in the time window between 180 and 300ms. Since awareness of a stimulus is a prerequisite of making a conscious decision, the time until the conscious broadcast (200-270ms, see previous section) can be subtracted from this window, yielding 20-90ms as the duration of the action selection process.

Philiastides et al. (2006) conducted an EEG experiment where subjects had to do a perceptual decision making task, deciding whether there was a face in the shown stimulus (faces in the stimuli had different coherence levels). They found brain activity strongly correlated with the subjects' decision 300ms post stimulus. They also identified a component at 220ms the strength of which systematically increased with task difficulty, to which they have assigned the top-down influence of attention (which is consistent with other experiments dealing with attention and consciousness). Subtracting these two times yields an action selection duration of 80ms.

Van Rullen and Thorpe (2001) have also conducted an EEG experiment involving a go / no go task with presented visual stimuli (depicting vehicles or animals). Resulting median reaction times were around 350ms, but

they also showed that categorization could be performed above chance after 250ms (which therefore constitutes the start of the decision process) – implying a duration of ~100ms for decision making (action selection).

An MEG experiment by Bauer et al. (2009), requiring subjects to perform a simple reaction time task, found high gamma band activity between 200 and 250ms poststimulus and suggested a role of this oscillatory activity in crossmodal integration, consistently with the conscious broadcast times described in Section 1.4. In this experiment, average reaction times were 279.1ms. Subtracting the lower bound of high gamma activity from the reaction time yields 79.1ms required for both the selection of an action and its execution. It is important to point out that reaction time experiments measuring actual motor responses include both the times of the cognitive cycle sub-processes, and the time for motor execution (which is not included in the described cognitive cycle). The time of the propagation of action potentials, from the motor cortex to evoking hand muscle responses, takes about 20ms (motor response was evoked 19-24ms after TMS stimulation of the motor cortex in an experiment by Capaday et al. (1991); which is consistent with the axonal conduction delays of motor neurons (Swadlow & Waxman, 2010)). Motor execution can therefore be said to take around 20ms. This time has to be subtracted from the results of these mechanical reaction time experiments to obtain the cognitive cycle duration. Thus, the action selection part in the experiment of Bauer et al. can be said to take approximately 60ms.

In the neural action selection circuit described above, we have included not only the selection of an action, but also the selection of the appropriate motor command executed by the motor cortex. These low-level motor commands –information about which muscles or actuators have to be used to implement a specific action – are stored in the Sensory-Motor Memory component in the LIDA model and are chosen after the action selection process. Choosing the exact low-level motor command to use takes a short amount of time in addition to the time taken for action selection. For example, when a person in a restaurant is faced with the decision whether to reach for a glass of wine or a glass of water, his or her brain needs to decide first (select the action) and then choose a low-level motor command (i.e. choose which muscles have to be flexed to reach and grasp the correct glass). The Sensory-Motor Memory has not yet been computationally implemented in LIDA; however, for the simple agents described below, this does not make a difference.

Summarizing, the process of action selection or decision making has been indicated to take 60-110ms.These times constitute a lower range for the action selection duration in humans, since they were obtained in studies using very simple settings – action selection may very well take longer if the task is more complex. (The 20ms lower boundary that has been deducted from the Taylor study (Taylor et al., 2007) has been disregarded because it is an outlier compared to the results of other studies).

2.5 Comparison with Psychological Reaction Time[*]

Adding up the durations of the cognitive processes mentioned above yields a total duration of 260-390ms for a single cognitive cycle (Figure 7). This is on the order of most reaction time experiments from psychology (although slightly longer than most simple reaction time experiments and slightly shorter than most choice task experiments).

The reaction times of young adults has been proposed to be in the range of 190-220ms (Kosinski & Cummings, 1999). Results from this and other reaction time experiments include the time taken for motor execution, which was not included in our discussion of the cognitive cycle above, and can be said to be around 20ms (see previous section). The time of the propagation of action potentials, from the motor cortex to evoking hand muscle responses, takes about 20ms (motor response was evoked 19-24ms after TMS stimulation of the motor cortex in an experiment by Capaday et al. (1991); which is consistent with the axonal conduction delays of motor neurons (Swadlow & Waxman, 2010)). Subtracting this delay, the cognitive cycle duration in these experiments can be inferred to be around 170-200ms, which is comparable to the lower limit of the cognitive cycle duration

described. For choice tasks, reaction times are in the range 356-400ms if there are two choices (Ratcliff & Rouder, 1998), which is very close to the upper limit of the proposed cognitive cycle duration.

For more substantial reaction time data, and a more complete survey of reaction time experiments, see (Healy et al., 2003).

2.6 Comparison of the Cognitive Cycle's Timing with other Cognitive Models[*]

Adding up the durations of the cognitive processes mentioned above yields a total duration of 260-390ms for a single cognitive cycle (Figure 7) (see also the data obtained from the LIDA Reaction Time Agent in Section 3.2). This is on the order of most reaction time experiments from psychology (although slightly longer than most simple reaction time experiments and slightly shorter than most choice task experiments).

The reaction times of young adults has been proposed to be in the range of 190-220ms (Kosinski & Cummings, 1999). Results from this and other reaction time experiments include the time taken for motor execution, which was not included in our discussion of the cognitive cycle above, and can be said to be around 20ms (see previous section). The time of the propagation of action potentials, from the motor cortex to evoking hand muscle responses, takes about 20ms (motor response was evoked 19-24ms after TMS stimulation of the motor cortex in an experiment by Capaday et al. (1991); which is consistent with the axonal conduction delays of motor neurons (Swadlow & Waxman, 2010)). Subtracting this delay, the cognitive cycle duration in these experiments can be inferred to be around 170-200ms, which is comparable to the lower limit of the cognitive cycle duration described. For choice tasks, reaction times are in the range 356-400ms if there are two choices (Ratcliff & Rouder, 1998), which is very close to the upper limit of the proposed cognitive cycle duration.

For more substantial reaction time data, and a more complete survey of reaction time experiments, see (Healy et al., 2003).

Figure 9. A comparison of the phase timings in LIDA, ACT-R and MHP

3. NON-VISUAL COGNITIVE SOFTWARE AGENTS[*]

3.1 Overview[*]

A computational framework of the cognitive cycle described in the introduction has been partially implemented (CCRG, 2011).

We have developed two autonomous software agents based on this framework, the LIDA Reaction Time (LRT) agent, performing a simple reaction time experiment; and the LIDA Allport Agent, replicating a psychological experiment regarding the continuity of conscious 'moments' (see Section 3).

3.2 The LIDA Reaction Time Agent[*]

The first implementation, the LRT agent, repeatedly performs a reaction time experiment in a simple environment consisting of a light (which can be red or green), and a button (which the agent has to press as quickly as possible when the light turns green). Figure 10 below contains a screenshot of the LRT agent. A description of how the LIDA computational model was adjusted for this specific task, as well as a list of parameters tuned to fit the described empirical data, can be found in Section 3.
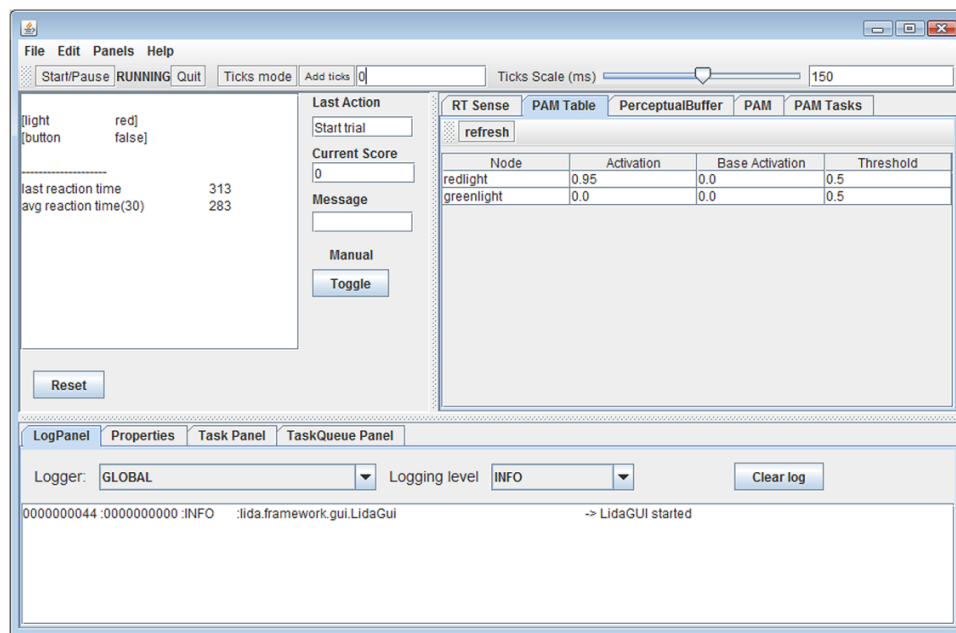
***Figure 10. A screenshot of the LIDA Reaction Time Agent.** The left top panel contains information about the environment (whether the light is red or green and whether the button is pressed) and statistics about the agent's performance (the last and the average reaction time). The right top panel contains internal information (shown here: the contents of PAM, i.e. the PAM nodes for the red and the green light, and their activations).*

Figure 11 below shows the LRT agent's performance at the simple reaction time task over 30 trials. As can be seen from this figure, the cognitive cycle durations of the LRT agent (283 ms) are comparable to the cycle durations inferred from the reaction times of adult humans (200ms according to (Kosinski & Cummings, 1999); see also discussion in the Decision Making / Action Selection subsection), although slightly larger. The main reason for humans being faster at such experiments is the effects of temporal expectation (which has not yet been implemented in LIDA). Humans seem to engage cortical action circuits (inferior parietal and premotor areas) prior to perceiving the stimulus (Coull & Nobre, 2008), and can thus reduce the time required for action selection after stimulus presentation. Still, the reaction times of humans and of the LRT agent are comparable (the difference is around 40%).
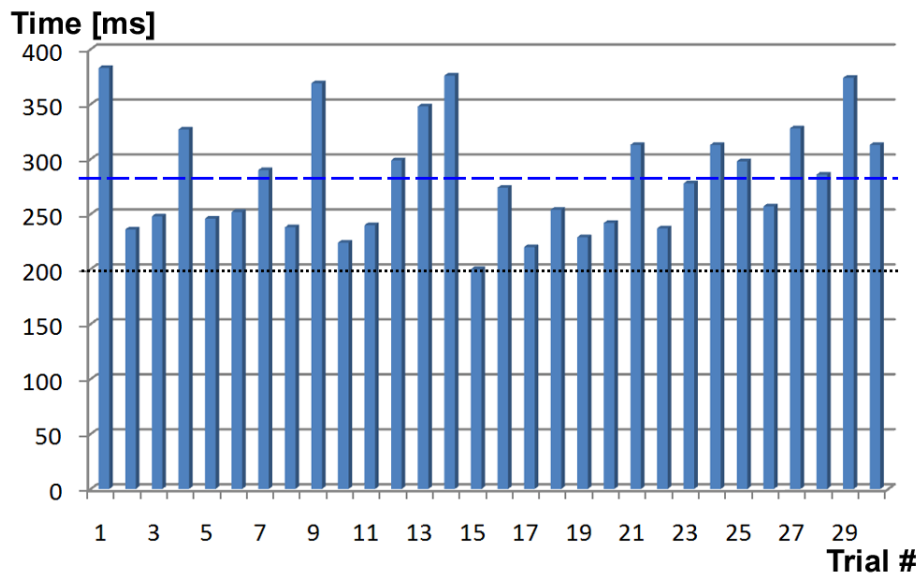
Figure 11. A histogram of the LRT agent's performance at the reaction time task. *The blue bars represent the reaction time in single trials. The figure shows n=30 trials; the average reaction time is 283ms. The dashed blue line is LRT's average reaction time; the dotted black line represents human reaction time (200ms, see Decision Making / Action Selection subsection)*

3.3 The LIDA Allport Agent[*]

Allport (1968) has conducted an experiment comparing two competing consciousness timing models. Stroud's (1967) Discrete Moment Hypothesis, states that consciousness is comprised of distinct and non-overlapping conscious 'moments', within which all time-order information is lost, while the Continuous (Traveling) Moment Hypothesis considers conscious 'moments' to correspond to continuously moving segments of the incoming sensory information.

Allport's results clearly contradict the strict Discrete Moment Hypothesis. LIDA's discrete consciousness mechanism, however, is consistent with this empirical evidence.

We have successfully replicated Allport's experiment computationally with three goals in mind:

- to show that our discrete consciousness model, based on neuroscientific evidence, does not contradict empirical data - unlike the Discrete Moment Hypothesis (see also the section "LIDA and Consciousness" above),

- to strengthen the claim that LIDA's GWT-based consciousness mechanism models human functional consciousness (note: in an artificial agent we refer to functional consciousness (Franklin (2003)), rather than phenomenal consciousness), and

- to substantiate the plausibility of the timing parameters proposed in this paper by showing the similarity of the LIDA Allport agent's behaviour and timing to actual human data.

In Allport's experiment, subjects were seated in front of an oscilloscope screen, which displayed a single horizontal line, appearing in one of 12 positions on the screen. This line rapidly changed position, moving upward. Upon reaching the topmost position, the screen was left blank for the same duration as the line took while traversing all 12 positions, and then the line appeared again on the bottom position – see Figure 12 (the same visual effect could have been achieved if the line had moved over the whole screen in 24 positions, but with the

bottom half of the screen covered). The rate of stepping, and thus the cycle time (τ), was controlled by the subject. At very large cycle times, subjects could see the single line jumping from position to position. Upon decreasing τ, they reported seeing multiple lines, moving together. At a specific cycle time S and below, subjects reported seeing a stationary array of 12 lines flickering in synchrony (see Figure 11 below).
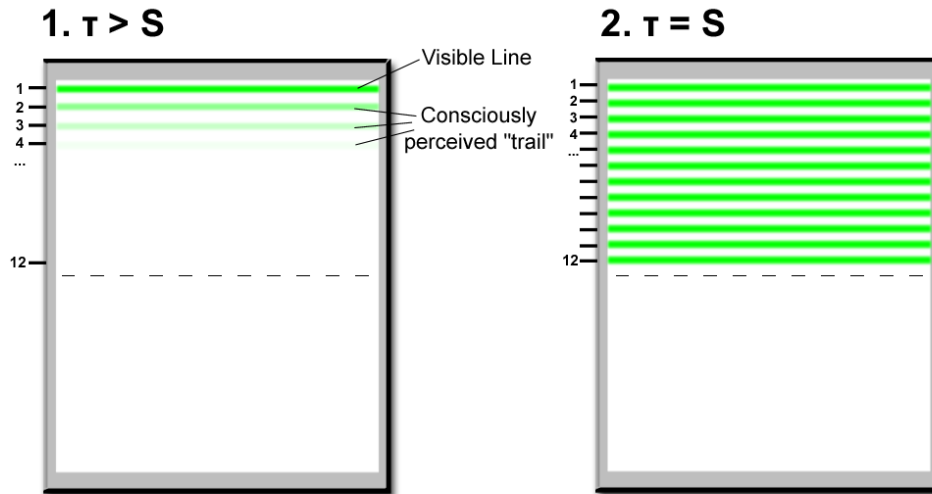
Figure 12. The display and conscious percept in Allport's experiment. *τ denotes the total cycle time. At cycle times $\tau > S$, subjects could see multiple lines moving together (left panel). At $\tau = S$, subjects saw all lines simultaneously and perceived no movement (right panel).*

The subjects had to arrive at the cycle time S, where they did not perceive any movement on the screen. In separate trials subjects first decreased the cycle time from a very high value (slow to fast), and then increased it from a very low value, at which all lines were seen simultaneously (fast to slow). Both times were recorded for each subject. These times were then compared to the predictions of the two hypotheses about consciousness.

According to the Discrete Moment Hypothesis, there are two cycle times at which all 12 lines appear to be on the screen: at $\tau = S$, at which the complete cycle falls within one conscious 'moment', and at $\tau = S/2$, at which conscious 'moments' containing all lines and no lines alternate (and thus the condition of no movement being perceived is met) – see Figure 13 below. The cycle time at which subjects will stop, perceiving no movement, will thus be S when decreasing τ, and S/2 when increasing τ. A significant difference between these two conditions is predicted.

The Continuous Moment Hypothesis predicts that successive events are perceived to be simultaneous whenever, and as long as, they fall within the temporal constraints of the conscious 'moment'. Thus, since the criterion for determining S was not only momentary simultaneity but perpetual absence of perceived movement, there can be only one cycle time S at which this criterion is met (see Figure 13). There should be no difference between trials decreasing or increasing τ.

In (Allport, 1968), twelve subjects performed two versions of this experiment under both conditions:

A) the half screen experiment described above, with

  1. decreasing the cycle time until no movement was perceived

  2. increasing the cycle time; and
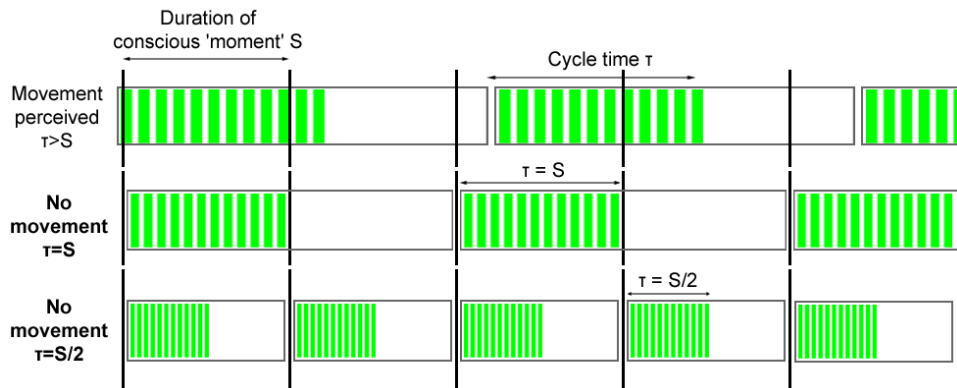
Figure 13. The predictions of Stroud's (1955) Discrete Moment Hypothesis for the Allport experiment. *There are two cycle times at which no movement is perceived (τ=S and τ=S/2). Depending on whether the subjects have to increase or decrease the cycle time, they should encounter one or the other. A difference of S/2 is predicted between the two trial types.*

  B) the full screen version of the experiment (where the 12 positions were distributed over the entire screen and the line immediately appeared again on the bottom of the screen after reaching the end of the cycle, without delay)

    1. decreasing the cycle time and

    2. increasing the cycle time.

Table 1 displays the resulting cycle times averaged over all subjects (*data from Allport, (1968)*). It is clear that the difference between increasing and decreasing trials is not significant (and certainly not close to S/2), which contradicts Stroud's Discrete Moment Hypothesis.

Cycle times τ [ms] Human subjects	1. (decreasing)	2. (increasing)
A (half screen)	95,5 (σ=16,0)	81,4 (σ=14,6)
B (full screen)	86,2 (σ=12,5)	70,7 (σ=8,1)

Table 1. Average cycle times at which subjects did not perceive movement in Allport's experiment *(n=12. σ denotes the standard deviation. Data from Allport (1968))*

The results from the simulation of these experimental conditions by the LIDA Allport agent are shown in Table 2 below. The data matches Allport's results – there is only one cycle time threshold S at which the agent does not perceive any motion. Despite the high standard deviations of Allport's data, and the as yet imprecise estimates of LIDA's internal parameters, it can be seen from this experiment that the timing data of the Allport agent is comparable to human performance.

Cycle times τ [ms] LIDA Allport agent	1. (decreasing)	2. (increasing)
A (half screen)	96	96
B (full screen)	84	84

Table 2. The LIDA Allport agent's cycle times at which the agent did not perceive movement *(n=12)*

3.4 Methods[*]

3.4.1 The Implemented Cognitive Cycle[*]

Both agents are based on the almost completely implemented computational LIDA framework, which provides extendable basic implementations for all modules in the LIDA cognitive cycle (Figure 4). These implementations have been extended to allow the agents to perform their respective experiments; and the default timing parameters have been adjusted to fit the empirical evidence described above.

To illustrate how the agents' implementations work, we shall describe in this section what happens in each of the modules of the LIDA cognitive cycle outlined in the Introduction; specifically:

1. Sensory Memory

2. Perceptual Associative Memory
   *(the 2 modules above are part of the **Perception** phase)*

3. Workspace

4. Attention Codelets

5. Global Workspace
   *(the 3 modules above are part of the **Understanding** phase)*

6. Procedural Memory

7. Action Selection

8. Sensory-Motor Memory
   *(the 3 modules above are part of the **ActionSelection** phase)*

There are two additional modules in LIDA's cognitive cycle (Transient Episodic Memory and Declarative Memory) which will be omitted here since they are not required in these experiments.

For this simple domain, no visual image processing is necessary. The environment class, which contains and controls the sensory stimulus (and the button), is inspected periodically by the **Sensory Memory** module. The LRT agent's sensory stimulus consists of a single red (or green) light, while the Allport agent's has 12 distinct lines which may or may not be alight.

Simple feature detectors monitor their respective fields in the Sensory Memory, and activate relevant **Perceptual Associative Memory** (PAM) nodes if they find corresponding sensory data. This is comparable to the human visual system, which also makes use of feature detectors – for example, V1 contains neurons that are sensitive to features such as orientation, direction and spatial and temporal frequency, and V4 neurons are sensitive to geometric shapes (Bear et al., 2007). In the LRT agent, the single color-sensitive feature detector activates the PAM node representing a red light or a green light, depending on Sensory Memory contents. In the Allport agent, there are 12 feature detectors sensitive to their respective lines, which activate one of the twelve respective PAM nodes upon sensing their line.

Next, the percept (consisting of the identified PAM nodes) is moved into the **Workspace**, which constitutes LIDA's preconscious buffers of working memory. The LRT agent does not use episodic memory, but in the LIDA model, episodic memory contents would be retrieved to the Workspace as well (from the Transient Episodic and Declarative Memories), cued by the percept.

According to Global Workspace Theory, on which LIDA is based, conscious contents reside in a memory capacity that enables access between brain functions that are otherwise separate (see Introduction). In LIDA, this memory capacity is the **Global Workspace**, and its role is enabling the Procedural Memory and the Action

Selection access to the most urgent/novel/relevant Workspace contents. These contents are transferred into the Global Workspace by **Attention Codelets** (codelets are special purpose mini-agents implemented as a small piece of code running on a separate thread). These codelets look for their specific concerns in the Workspace and, upon finding it, copy it to the Global Workspace.

An agent is consciously aware[1] of an object, represented by PAM nodes, the moment these nodes become part of the conscious broadcast (after winning the competition against other contents of the Global Workspace).

Finally, an appropriate action is selected based on the contents in the broadcast. This selection is performed by two components in LIDA. The first component is **Procedural Memory**, from which all behaviours applicable in the current situation are chosen. In the LRT agent, as well as in the Allport agent, there are two possible behaviors (pushing the button, and releasing the button/doing nothing). Note that behaviors could be more complex (they could include many actions) in a more complex domain of application.

The second component is **Action Selection**, in which the action best serving the agent's goal is selected. In the agents described here, this process is trivial – since in all possible states of the environment there is only one applicable action, the Procedural Memory always yields only one action, which only has to be forwarded by the Action Selection component (without competition between actions) to the **Sensory-Motor Memory** for execution. This selected action is then executed in the environment (e.g. the button is pressed). The simple mechanism responsible for this could be called the LRT agent's "actuator".

3.4.2 Parameters*

As do other computational architectures modeling cognition, LIDA contains a multitude of internal parameters that have to be adjusted for a computational agent acting as subject in the replication of an experiment. Such parameters may include decay rates for various types of memory, a threshold above which a perceptual item becomes part of the current percept, or a parameter that makes action selection more goal-oriented rather than opportunistic. The ultimate goal is a tuned set of internal parameters whose values remain constant when a number of disparate datasets are reproduced. Such a tuned parameter set assures the accuracy and usefulness of the model. Inability to find such a tuned parameter set should warn that the model needs revision. The particular parameters that resist such tuning will point researchers to modules and processes within the model that need revision. This parameter tuning provides a metric for assessing the quality of a cognitive model as a basis for understanding the cognitive processes responsible for the behavior of the agent.

Successfully accomplishing this goal will provide substantial evidence of the accuracy and usefulness of the conceptual cognitive model. Cognitive hypotheses from the model can then be tested by experiments with human subjects to see if their data is predicted by running artificial subjects in the same experimental situations. If so, we will have shown the ability of the theoretical model to predict as well as to explain.

The timing parameters described in this section are a first step in the direction of a well-tuned parameter set for the LIDA model.

Each module in LIDA has a specific task (see module descriptions above) that has to be executed at least once every cognitive cycle. The module tasks are run in a parallel and asynchronous fashion - like the human brain, which does not use sequential information processing, but, rather, local neural circuits which run in parallel.

In the computational framework, all of these module tasks are executed periodically to implement the LIDA cognitive cycle. The execution intervals are governed by 'ticks' parameters. These parameters govern in how many 'ticks' (simulated milliseconds) a particular task will be executed.

3. Non-Visual Cognitive Software Agents*

Adjusting these 'ticks' parameters, so that the timings of the resulting LIDA cognitive cycle become comparable with the timings of the human action-perception cycle (and, thus, neuroscientifically plausible) was the main purpose of the development of the LRT agent.

The most important parameters resulting from this parameter adjustment are listed in Table 3 below. It is important to point out that the modules corresponding to these parameters do not run in a serial manner - the LIDA model aims for the highest possible asynchrony. The only points in the cognitive cycle where seriality is enforced are the conscious broadcast and the action selection process (the selection of a behavior can only start when the contents of the global workspace become conscious).

Parameter name	Value [ms]
1. Sensory Memory Ticks	20
2. Feature Detector Ticks	30
3. Attention Codelet Ticks	200
4. NoBroadcastOccurring Trigger	200
5. ProceduralMemory Ticks	110

Table 3. The LRT Agent's most important timing parameters.

The first parameter governs how often the contents of the Sensory Memory are updated, i.e. how often the environment is sampled. This would be a domain specific parameter that must be found anew for each LIDA controlled agent implemented.

The second parameter controls how often feature detector codelets are run, detecting features depending on their specialization. Feature detection is very rapid in the LRT agent, as in humans. V1 neuron response latencies start at 30ms – (Huang & Paradiso, 2008; Kirchner et al., 2009). Also, a presentation time of 20ms is required for simple go/no go classification for visual stimuli – (Thorpe et al., 1996). In other experiments, 30ms was required – (Martinez-Conde et al., 2006). This is also consistent with V1 firing rates, which peak at about 45 spikes per second (Heeger et al., 2000). In the LRT agent, there are only two Feature Detectors, which detect the color of the light stimulus (one for red and one for green). Upon detecting their corresponding light stimulus, these Feature Detectors pass activation to the corresponding nodes in the Perceptual Associative Memory. If the activation of the updated PAM node exceeds a specific threshold, then a copy of this node is instantiated in the Workspace (LIDA's preconscious working memory).

The next important timing parameter (number 3 in Table 3) governs how often the attention codelets are run. Attention codelets are mini-agents that have the purpose of bringing novel, relevant, urgent, or insistent events to consciousness (i.e. bringing instantiations of their corresponding PAM nodes, or other Workspace structures, to the Global Workspace). Since we have argued that the onset of conscious processing in humans starts at about 200ms (see Results), this parameter was set to this value. It is important to point out that the conscious broadcast can have multiple triggers. In more complex domains, the broadcast is triggered whenever the cumulative activations of the coalitions built by Structure Building Codelets exceed a specific threshold. The broadcast can also be triggered if a single coalition exceeds another threshold. Both of these thresholds can be interpreted as contents judged novel or important enough being brought to consciousness. Finally, a broadcast is sent automatically if too much time has passed since the last broadcast has commenced. The idea is to allow the conscious processing of less important information in cases when there is no current novel or vitally important content in the Global Workspace (instead of an extended unconscious period that would last until one or more coalitions exceed the activation threshold again). The time at which this trigger is activated, measured from the

onset of the last conscious broadcast, is controlled by Parameter 4 (NoBroadCastOccuring Trigger) and was set to 200ms, the onset of conscious processing in humans, as well.

In the domain of the LRT agent, there is only a single coalition in the global workspace (containing a PAM node representing a red or a green light). A conscious broadcast is automatically triggered whenever the activation of this coalition exceeds a specific threshold. The timing parameters of the Attention Codelet, and those of the perception process, have been chosen in a way that the broadcast happens in the range of 200 – 280ms (the range for the onset of consciousness in humans – see the Cognitive Processing and Consciousness Section).

The final parameter (number 5 in Table 3) governs the frequency of the process that leads to the selection of an action. The 'ProceduralMemory Ticks' parameter controls how often the set of actions that are applicable in the current situation is retrieved and the actual best action selected. This parameter has been set to 110ms, the upper limit of the duration of action selection (see Results). As in humans, the duration of the action selection phase will depend on task complexity (especially, on the number of available actions). Since the implementation of the Procedural Memory and the Action Selection components in LIDA are still being worked on, the internal timings of this action selection phase have not yet been determined. But both of these processes have to be rescheduled at intervals longer than the internal processing time they require, to avoid bottlenecks, which is why parameter 5 has been set to the upper limit of the action selection duration described in the Results section. In the current LRT agent implementation, these processes take a very short amount of time; and are rescheduled periodically at intervals indicated by parameter 5 in Table 3. For future agents, an improved action selection mechanism based on (Maes, 1989) is in development, which will involve the use of triggers (triggering the selection of the best action, for example, if at least one of the applicable actions has activation above a specific threshold) instead of periodic action selection.

Figure 10 in the previous section shows a diagram of the resulting reaction times of 30 trials performed by the LRT agent. For the results of the Allport agent see Table 4 and the previous section. Although setting these parameters and pointing out consistent results does not prove either the cognitive cycle hypotheses or the correctness of our timings, this parameter adjustment has to be done as a prerequisite of building more complex LIDA agents, because the cognitive cycles will have to run at a speed comparable to human cognitive cycles if we expect them to model human cognition (or an aspect thereof). If a number of such LIDA agents, replicating different psychological experiments and thus focusing on different aspects of human cognition, would operate in time frames consistent with the human brain (without readjustments of internal parameters), this would considerably increase the plausibility of the LIDA architecture as a model of human cognition.

4. VISUAL PERCEPTION FOR LIDA AGENTS

4.1 Vision in LIDA

Visual object recognition in the brain is facilitated by the ventral pathway – area V1 (which contains neurons sensitive to direction, orientation and other features), area V2, area V4 (containing neurons sensitive to color and shape), the inferior temporal cortex or area IT (already containing object category information) (Liu et al., 2009) and the prefrontal cortex (PFC) (Serre et al., 2007) (Bear et al., 2007) (see Section 5.3.1 for a more complete account for the visual system). A majority of visual processing in the brain probably happens in a feedforward fashion, due to the low latencies (Liu et al., 2009) and the neural pathways across the involved stages (Serre et al., 2007). For this reason, the model described here will focus on feedforward object recognition.

The two modules involved in visual perception in the LIDA model are the Sensory Memory and the Perceptual Associative Memory (See Section 1.5). The conceptual LIDA model proposes a feature detection mechanism which is functionally similar to feature sensitive neurons and neuronal ensembles in the visual cortices (Madl et

al., 2011; McCall et al., 2010). Higher-level object recognition capabilities arise from associations and combinations of various low-level feature detectors receptive to simple visual features, which observe a buffer containing raw visual data from the environment – the Sensory Memory – and, upon sensing their respective features, pass activation to connected PAM nodes. PAM nodes above a specific threshold (if enough features associated with that PAM node could be found) can reach working memory and, subsequently, consciousness, and can be considered in action selection (D'Mello et al., 2006).

This theoretical mechanism has been implemented in practice for actual image recognition tasks in the scope of this thesis, since there is no prior implementation of object recognition working on visual environment representations in the computational LIDA model at this time.

4.2 Invariant Feature Extraction

For a robust visual object recognition system working on natural images, scale and position invariance in the detection of image features is vital. In the brain areas concerned with visual processing, there is an increasing amount of such invariance across the ventral pathway from V1 to IT – while V1 feature detector neurons are mostly sensitive to simple features such as line segments in various orientations, V4 neurons can be sensitive to geometric shapes or blobs (Serre et al., 2007; Bear et al., 2007).

In this visual system implementation I have decided to use scale and position invariant features extracted by the SURF (Speeded Up Robust Features) algorithm (Bay et al., 2008). This approach extracts features based on Hessian matrices – basically, the areas in the image with the greatest amount of brightness changes – and it can be said that such an approach is not biologically plausible. However, this approach – implemented in the Sensory Memory, as described below – interfaces with the rest of the LIDA model, which is argued to be biologically plausible elsewhere (Baars & Franklin, 2009; Baars & Franklin, 2007; Franklin et al., 2005; Madl et al., 2011); and has the advantages that it is robust, performant enough to be used in software agents or robots in real time, and last but not least, it was possible to implement in the time constraints on this thesis.

$$\mathbf{v} = (\textstyle\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|)$$
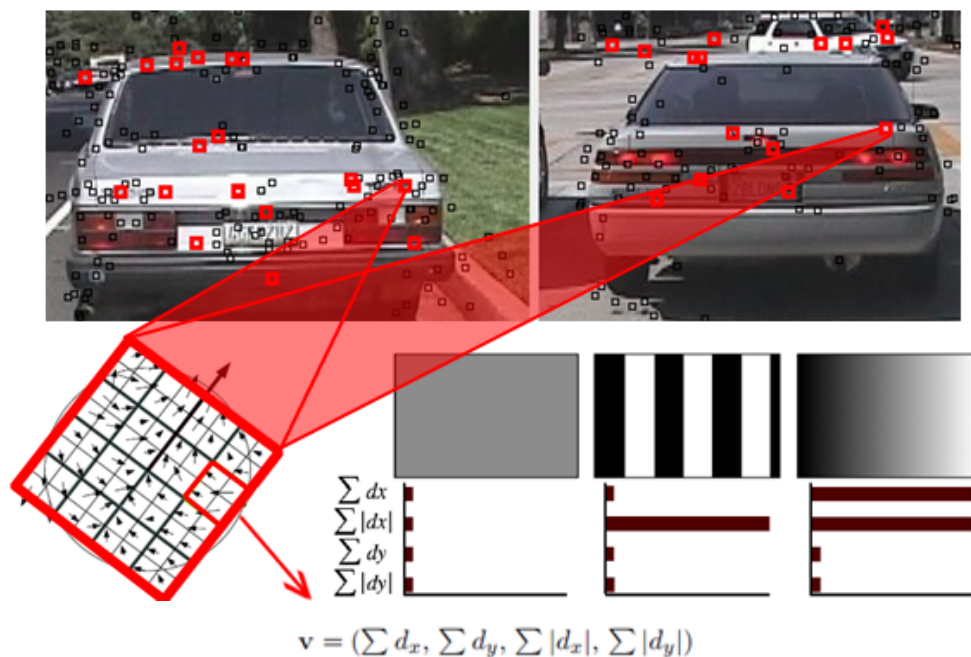
Figure 14. Features extracted by SURF. Top: two images contain small squares marking the locations of SURF features (at areas in the image with maximum Hessian determinants. Note: the features can describe regions of

various sizes, which is not shown here). Matching features are highlighted in red. **Bottom:** *SURF feature descriptors. **v** is the representation of one of the 4x4 subregions of an interest point. The bar charts above visualize the components of **v** (the summed-up Haar wavelet responses) for various types of brightness functions. (The bottom two images were taken from (Bay et al., 2008))*

As stated above, SURF extracts features based on Hessian matrices – SURF features are blob-like structures at locations where the Hessian determinant (a square matrix of second-order partial derivatives of the brightness function which represents the image) is maximum (Bay et al., 2008).

The SURF algorithm computes descriptor vectors for each identified feature in the image. These descriptor vectors describe the distribution of the intensity content in a window around the interest point (feature) with a specific size, which is split into smaller 4x4 subregions. Each subregion is represented by summed up Haar wavelet responses (maximum second order derivatives of the brightness function) of that region – the representation containing of the sum in x and y directions and the sums of the absolutes in x and y directions of the wavelet responses (see Figure 14). Thus, the representation of each subregion contains 4 values. Since each interest point describes 4x4 subregions, the overall descriptor of an interest point is a vector containing 64 values (16 subregions representations) (Bay et al., 2008).

Descriptor vectors can be used to find corresponding interest points in different images. (Bay et al., 2008) propose nearest neighbor distance ratio matching, in which for every point P in image A, the nearest and second nearest point in image B are found (the "nearest" point, N, being the one with the minimum Euclidian distance to the descriptor vector of point P). If the ratio of Euclidian distances of the nearest and second nearest interest points to the interest point P is less than 0.7, then point N of image B is said to match point P of image A.

4.3 SURF-based Visual Object Recognition in LIDA

The first step in the proposed object recognition system is the pre-processing of the input image with SURF, which yields a number of descriptor vectors. This is done in the Sensory Memory, and the extracted vectors are put into an additional layer called the SURF Vector Space (see Figure 15). This layer contains the computational representation of a discrete 64 dimensional space (since the description vectors consist of 64 values), partitioned into a fixed number of hypercubes in which image interest points can be stored. This makes it possible for feature detectors to observe a fixed area, instead of having to do an iterative matching for every input image with $O(c * n * m)$ complexity (n being the number of features that the system is looking out for, m the number of features in the input image, and c the constant steps required to calculate whether two features match).

The original nearest neighbor distance ratio matching is similar to nearest neighbor matching (point P matches point N if N is the nearest neighbor of P and their Euclidian distance is smaller than a threshold T), but additionally penalizes points having many similar matches, which improves precision (Mikolajczyk & Schmid, 2005).

If these points are stored in a data structure where the closest neighbors can be retrieved rapidly, the matching algorithm can be simplified. In the case of nearest neighbor matching, only those points within a hypersphere with radius T have to be iterated through and the nearest point retrieved. This strategy can also be used for nearest neighbor distance ratio matching of interest points in the SURF Vector Space, because this matching strategy is similar to simple nearest neighbor matching (Mikolajczyk & Schmid, 2005) and also because of the distribution of points (the interest points derived from a single image tend to gather closely around a few point clusters in the vector space).

Thus, the feature detectors really have to observe the contents of hyperspheres instead of hypercubes in order to determine whether their region of the Vector Space contains a relevant feature (see the white dotted circles in
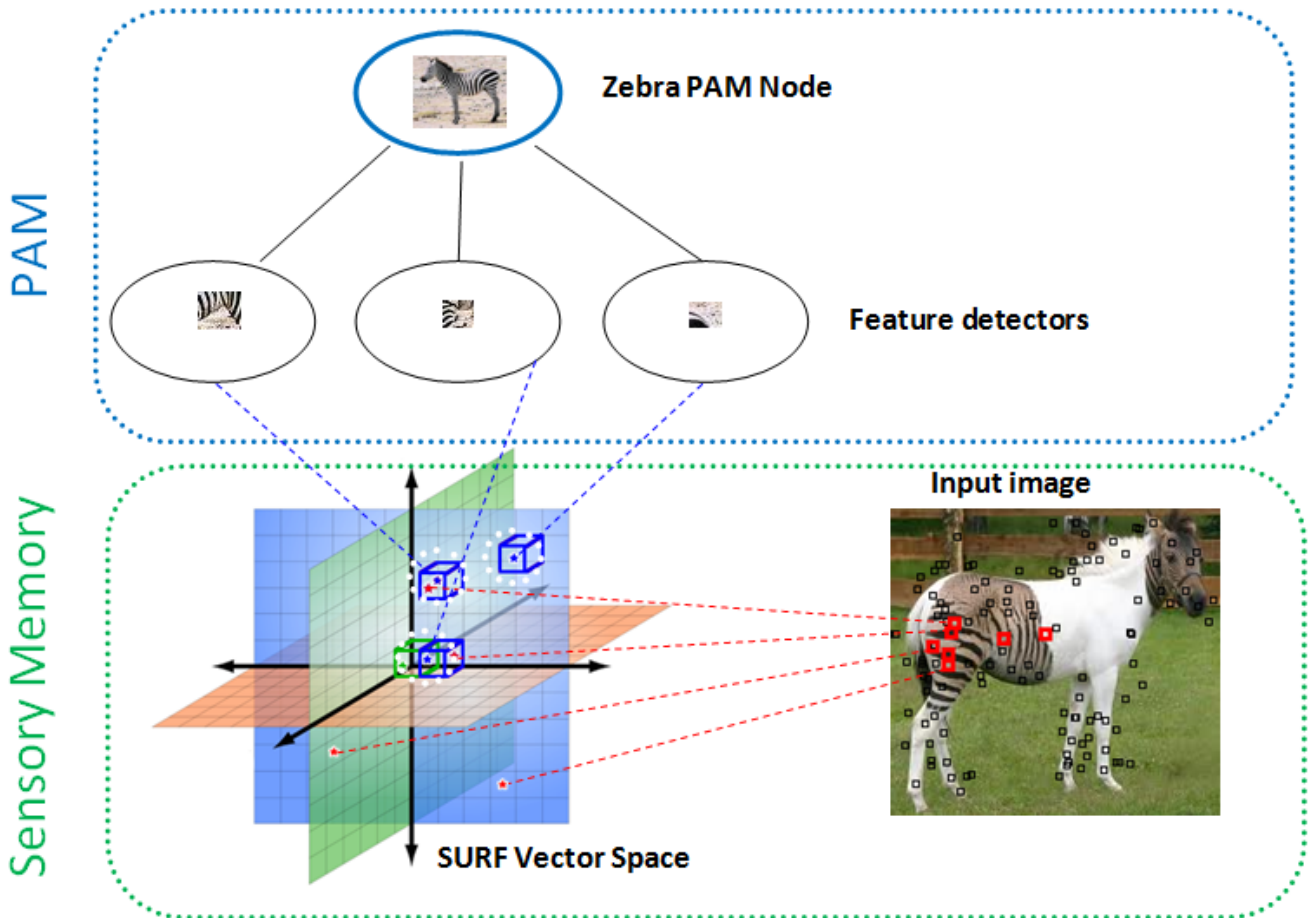
Figure 15. SURF-based visual object recognition using feature detectors. *New images arrive in the Sensory Memory through appropriate visual sensors, are prep-processed by SURF, and the resulting interest points are copied into the appropriate locations in the SURF Vector Space. Feature detectors are periodically checking their corresponding area in the space and pass activation to all PAM nodes (concepts) connected to them if an input image feature is present in that area. (The bottom two features are unobserved by feature detectors.) The small stars represent the locations of features in the Vector Space (red stars correspond to featuresin the input image, blue stars to features that the feature detectors are looking out for). The cubes show partitions in the Vector Space which are observed by feature detectors. The white dotted circles denote the radius around each observed feature that has to be checked by its feature detector for matching interest points.*

Figure 15, which represent these hyperspheres to be checked). However, the computational implementation of a discrete, partitioned hyperspace with non-overlapping partitions which cover the entire space is only possible with hypercubical partitions. For this reason, each feature detector observes not only the hypercube that its corresponding interest point is located in, but also adjacent hypercubes (because of the chosen computational representation, this can be done in computationally cheap and rapid way). The sizes of the partition hypercubes in the SURF Vector Space have been chosen in a way to maximize the probability that matching features are located in the hypercube directly observed by feature detectors, and thus to minimize the time a feature detector needs to check for matches.

All feature detectors finding a matching interest point derived from the input image in the SURF Vector Space pass a specific amount of activation to their corresponding PAM node. This passed activation depends on how

prevalent the feature the detector is looking out for has been in the training set. For features that have been present in every training image (pertaining to the concept represented by the PAM node), the passed activation is high, while for features that were found only in a single image in the training set, this activation is very low.

This mechanism reliably leads to the PAM node representing the object in the input image to receive the highest amount of activation (see next section), provided that the input image is similar enough to the images in the training set.
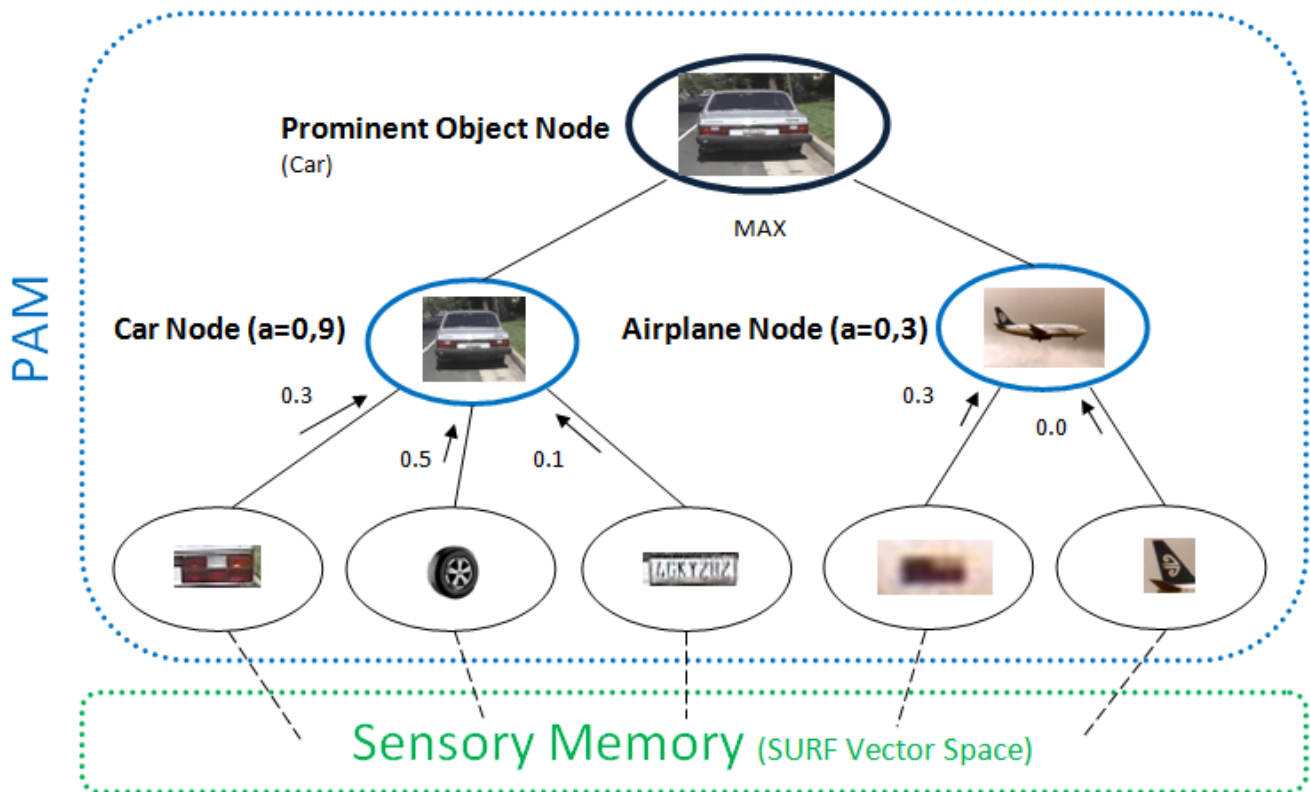
Figure 16. PAM Structure. *The Perceptual Associative Memory contains feature detectors for every feature in the training set images, which observe the SURF Vector Space. If they find matching points there, they pass activation to their parent PAM nodes (here, the Car or Airplane node). The node representing the prominent object in the input image will be the node receiving the highest amount of activation (determined by the MAX node in the top of the figure). Some features are common to multiple objects (which is the reason why the activation of the Airplane node is not zero)*

In the current implementation, the feature detectors are set up during a training phase (but they could as well be learned online). A number of PAM nodes are set up from a pre-defined list of objects, and corresponding training image folders are read and processed by SURF. Detected interest points are cross-compared and their prevalence in the training set determined. Using this prevalence value and the descriptor vector, feature detectors are added to PAM for every interest point, connected to the appropriate hypercube in the vector space using the descriptor vector, and set to pass an appropriate amount of activation using the prevalence.

This addition of PAM nodes and feature detectors could also occur during on-line learning. For example, when the agent is presented with an image of a novel and unrecognized object, it could spawn a new PAM node for the unknown object category and set up feature detectors for it as described above (with the exception of prevalence values – all feature detectors would have to pass the same activation).

4.4 Performance and Comparison

To evaluate the performance of this visual object recognition system, a subset of the CalTech image database (Fergus et al., 2003) was used. This database contains 1074 images of airplanes, 1155 images of cars, and 450 images of faces (amongst others); taken under different lighting conditions, different angles and with different backgrounds (see Figure 17). For each of these three categories, 100 images were selected at random and used as training sets. The remaining images were used for testing.
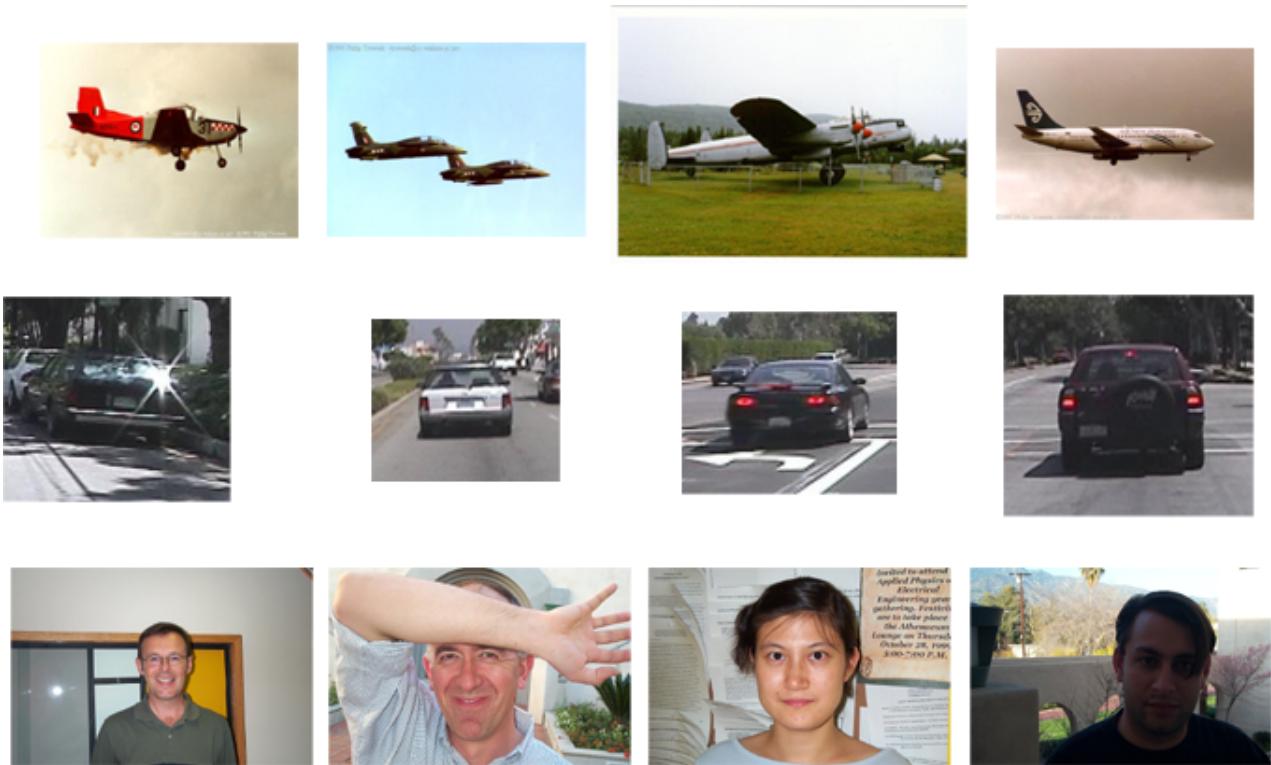
Figure 17. Sample images from the Caltech database

The following table contains the recognition rates of LIDA's SURF-based object recognition system, compared to Fergus et al.'s (2003) system as a benchmark.

Category	SURF-based object recognition	Benchmark
Airplanes	88,8 %	90,2 %
Cars	84,0 %	84,8 %
Faces	87,1 %	96,4 %

Table 4. Recognition rates of the proposed object recognition system. *The benchmark numbers were taken from (Fergus et al., 2003). The displayed percentages are true positive rates (the amounts of correctly recognized images from the total number of images of the respective category).*

As can be seen from Table 4, this SURF-based object recognition system performs comparably to, but slightly worse than, the benchmark. However, unlike the benchmark, these recognition rates were obtained without any pre-processing of the images (for example, Fergus et al. flipped images to ensure that all of them were facing the

same way). It also has to be pointed out that our training set was created entirely at random. It is likely that the system could achieve higher recognition rates with more representative training sets.

The following table contains further information about the system's performance on the Caltech image set.

Category	True positive rate[1] (Sensitivity)	False positive rate	Specificity[1]	Balanced accuracy[1]
Airplanes (n=974)	0.888	0.118	0.881	0.903
Cars (n=1055)	0.840	0.022	0.977	0.909
Faces (n=350)	0.871	0.062	0.937	0.904

Table 5. Recognition performance on the Caltech image database. 100 training images were used for each of the three categories; the rest of the images (n) were used for testing.

5. THE ATTENTIONAL BLINK IN THE LIDA MODEL

5.1 Introduction

The attentional blink (AB) refers to the impairment in consciously perceiving the second of two targets presented in close temporal proximity (200 – 500ms). In a serial presentation of a number of nontargets and the two targets, subjects can almost always accurately report the first target (even is the presentation time is short - usually around 100ms), but often fail to report the second target if it is presented in short succession (see Figure 18).

During the past two decades of research on the AB, a number of related phenomena and effects were found, some of which are hard to explain using traditional models of attention. No theory or model explaining all of the empirical findings related to the AB has yet been found (Dux & Marois, 2009). Here we will introduce a LIDA-based explanation of the AB, and a cognitive software agent reproducing data obtained from human subjects. Our approach aims a) to provide a more comprehensive explanation based on a cognitive architecture, encompassing a wide range of cognitive processes, not just attentional selection; and b) to account for and explain more empirical findings than previously published models.

Before describing the model in Section 5.3, a number of findings and phenomena obtained under different experimental conditions shall be described. They will be used to compare existing models of the AB and the LIDA based model. A good model should account for a wide range of findings.

5.2 Paradigms and Phenomena

The basic and most commonly used attentional blink paradigm consists of the Rapid Visual Serial Presentation (RSVP) of a stream of characters at a rate of about 10 items/second, consisting of digits (distractors or non-targets) and two target letters (called T1 and T2) (Chun & Potter, 1995; Martens & Wyble, 2010). Subjects are instructed to identify the letters in the stream and to report them after the presentation of the stream (see also Section 5.4 for a more detailed description of a visual attentional blink paradigm).

---

[1] $\text{balanced accuracy} = \dfrac{\text{sensitivity} + \text{specificity}}{2} = \dfrac{0.5 * \text{true positives}}{\text{true positives} + \text{false negatives}} + \dfrac{0.5 * \text{true negatives}}{\text{true negatives} + \text{false positives}}$

$\text{sensitivity} = \dfrac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}}$ $\text{specificity} = \dfrac{\text{number of true negatives}}{\text{number of true negatives} + \text{number of false positives}}$
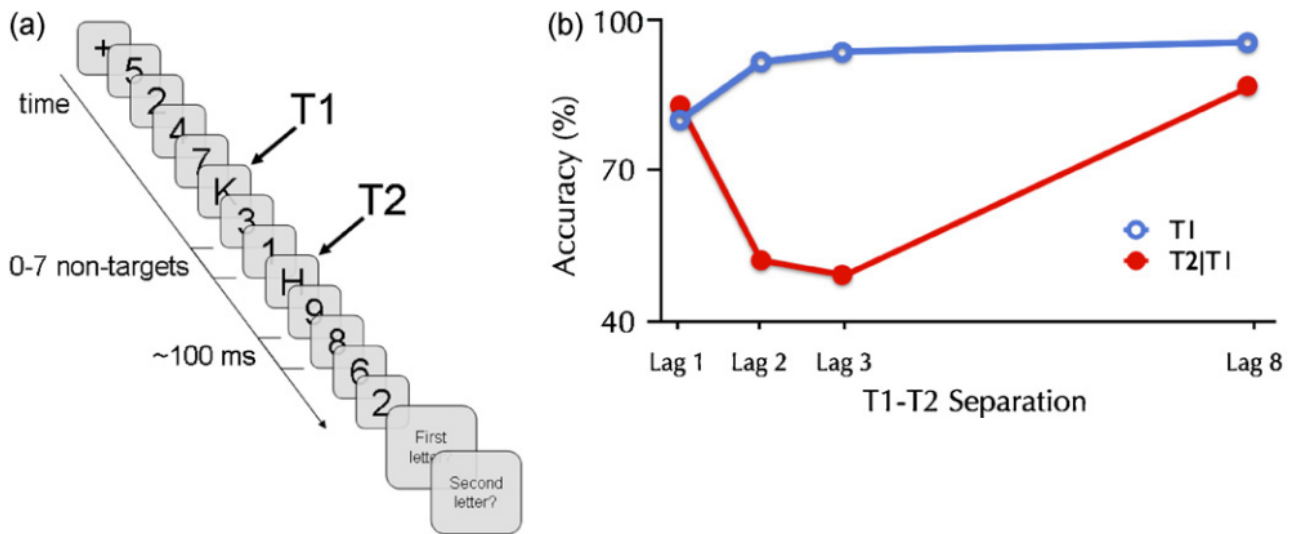
Figure 18. The basic attentional blink paradigm. *a) A series of characters is presented in rapid succession. The subject is instructed to notice and report the two letters (targets) in the stream of digits (non-targets). b) The percentage of correctly reported T1 (blue line), and the percentage of correctly reported T2 from trials where T1 was correctly reported (red line). Subjects often fail to report T2 when it is presented within 200-500ms to T1 (this drop in accuracy is called the attentional blink). If T2 is presented immediately after T1 and inside of about 100ms, both targets can be reported accurately. This paradoxical finding is referred to as lag-1 sparing ("lag" refers to the sequential distance between T1 and T2). (Adapted from Martens & Wyble, 2010)*

The AB seems to constitute a central limitation on the human attentional capacity and can be observed in a great majority of subjects and across a wide range of paradigms, types of target identification, and modalities (Martens & Wyble, 2010) – for example, with symbols (Chun & Potter, 1995), words (Barnard et al., 2004), tactile stimuli (Hillstrom et al., 2002) and pictures (Evans & Treisman, 2005; Potter et al., 2010). Apart from insights into the limits of attention, the latter paradigm also provides information about feature binding in picture detection. Potter et al. (2010) used this picture detection paradigm to show that picture identification must occur rapidly, in the first ~100 ms (or in Stage 1 of two-stage explanations of the AB, see below), and does not occur through the binding of initially unbound features at a later processing stage, as suggested by Evans & Treisman (2005) (this finding is consistent with recent intracranial recordings in occipital and inferior temporal areas, which can be correlated to object categories after about 100ms (Liu et al., 2009)).

Brain related evidence has shown that during an AB task, both targets are processed at least perceptually, regardless of conscious reportability – at least the first 150ms of neural activity exhibits a normal pattern (Martens & Wyble, 2010). EEG studies have revealed the electrophysiological activity that correlates to the AB – the N2pc ERP component (Event-Related Potential, brain activity directly resulting from and time locked to a stimulus), occurring about 200ms poststimulus and associated with the allocation of attention to targets, is suppressed at short temporal distances between T2 and T1. Also, in trials where T2 cannot be perceived because it is presented shortly after T1, the P3 component – associated with working memory consolidation – is not elicited (Martens & Wyble, 2010; Dux & Marois, 2009). Furthermore, an fMRI study conducted by Marois et al. (2004) showed parahippocampal place area activations (associated with high-level scene representations) even in non-conscious T2 targets. The above evidence implies that the AB has to occur at a later stage of processing (later than perceptual recognition, and after 150ms).

Apart from this evidence about attentional processing – which is accounted for by almost all models of the AB (Dux & Marois, 2009) –, a number of other empirical findings will be briefly reviewed here. For more complete reviews of the attentional blink and related findings, see (Martens & Wyble, 2010) and (Dux & Marois, 2009). Section 5.3 will describe how the LIDA-based model can explain these findings, and Section 5.4 will compare this model to previously published models of the AB.

1) **Lag-1 sparing.** Paradoxically, T2 can be reported with high accuracy if presented shortly after T1 (about 100ms after T1) (Martens & Wyble, 2010).

2) **Spread lag-1 sparing.** Multiple targets can be reported as long as they are presented in immediate succession - Olivers et al. (2007) observed that target reports were accurate even for four successive targets.

3) **Posttarget intrusion.** Varying the experimental conditions revealed that the attentional blink only occurs if T2 is backward masked (Giesbrecht & Di Lollo, 1998). Often, this mask or distractor succeeding T2 can be reported even if T2 cannot, implying that the distractor somehow interferes with the reporting of the target (Chun, 1997).

4) **Whole report attenuates the AB.** The accuracy of reporting stimuli is high when subjects are asked to report all stimuli (whole report). However, a significant accuracy drop at lags 2 – 4 (an attentional blink effect, see Figure 18) can be observed for the same stimuli sequence if subjects are required to report only two targets in the sequence (Potter et al., 2008).

5) **Increasing T2 salience attenuates the AB.** If the salience of the second target is increased (e.g. it is presented in a different color), it can be reported more accurately, although an AB effect can still be observed. The same can be observed if salience is increased on both targets (Shih & Reeves, 2007; Martens & Wyble, 2010).

6) **Task-irrelevant cognitive load attenuates the AB.** If the stimuli are presented together with a background field of moving or flickering dots, much smaller drops in accuracy are observed at AB-relevant lags (Arend et al., 2006). The AB is also attenuated if subjects are asked to listen to task-irrelevant music or think about their holiday (Olivers & Nieuwenhuis, 2005).

7) **Using target-similar distractors increases the AB.** The AB magnitude can be increased by making the distractors more similar to the targets (for example, an attentional blink while recognizing colored letter targets in a letter stream is stronger than recognizing the same targets in a stream of digits) (Dux & Coltheart, 2005; Dux & Marois, 2009).

8) **Target Confusion.** The order targets were presented in is often confused for temporally adjacent targets (i.e. during lag-1 or spread lag-1 sparing) (Chun, 1997; Dux & Marois, 2009)

9) **AB without T1 masking.** Although T2 masking is necessary to obtain an attentional blink, recent studies found that there is an (attenuated) blink even if T1 is unmasked (i.e. if there is no distractor between T1 and T2) (Nieuwenstein et al., 2009).

10) **T2 cueing attenuates AB.** Presenting a distractor that is similar to the target category preceding T2 decreases the AB magnitude (e.g. if the targets are red letters in a stream of white digits, presenting a red digit before T2 increased the probability that T2 is perceived consciously) (Nieuwenstein et al., 2005).

5.3 Attention in LIDA

The attentional mechanism is implemented by two major components of the computational LIDA model, the Attention Codelets Module and the Global Workspace Module. Attention Codelets form coalitions using selected portions of the perceptual scene, selecting the most urgent, novel, or important percepts and moving them to the Global Workspace. Each of these coalitions has a stored activation value which depends on bottom-up saliency (coming from early perceptual processes) as well as top-town importance (added by the Attention Codelet and reflecting urgency or importance). The coalitions enter into a competition for consciousness, and the winner – the

one with the highest activation – enters the conscious broadcast, is available for action selection, and can be reported consciously (see also Section 1 for a more detailed description of LIDA's modules, cognitive cycle, and consciousness mechanism).

We will focus on visual attention in this section, since most experiments mentioned above have used visual stimuli, and since the computational agent implementation also uses a visual paradigm. The following subsections will explain which areas and processes are associated with attention in the brain (5.3.1), and detail how LIDA can provide a neuroscientifically plausible account for the attentional blink (5.3.2).

5.3.1 Neuronal Correlates

The human visual system seems to utilize two different processing streams: the *dorsal* pathway (going to parietal areas), called "where/how" stream since it implements spatial vision, and the *ventral* pathway (going to the temporal cortex), called "what" stream since it specializes in object recognition (Itti, 2003). Both of these pathways are implicated in visual attention, contributing to the selection of what and where to attend.
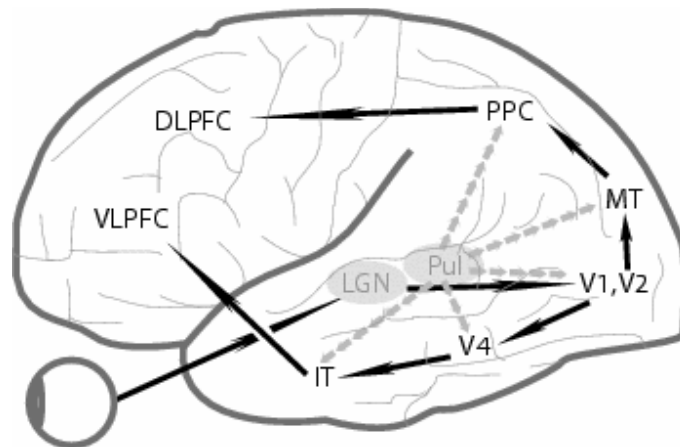
Figure 19. Brain areas involved in selecting "what" and "where" to attend. *The ventral "What" pathway goes from V1 (striate cortex) to VLPFC (ventrolateral prefrontal cortex) via IT and is concerned mainly with object recognition, and the dorsal "where/how" pathway goes from V1 to DLPFC (dorsolateral prefrontal cortex) via PPC (posterior parietal cortex). Arrows represent forward as well as reciprocal connections. From (Itti, 2003)*

Two types of attentional control can be distinguished: bottom-up or stimulus-driven attention, which is largely unconscious and driven by specific attributes and features of the input image; and top-down or voluntary attention, which is largely specified by behavioural goals or task demands.

Bottom-up salience is mainly based on low-level features - more specifically, feature contrast (not local absolute feature strength) (Itti, 2005). This kind of low-level feature selectivity is exhibited by the lateral geniculate nucleus and visual cortices (V1-V4, sensitive to oriented edges, combinations of features, and in the case of V4, color) (Serences & Yantis, 2006). Even in the visual cortices, an effect of voluntary attention has been demonstrated, although it has been found that the relative influence of top-down attention is very low in these areas, and only becomes comparable to the effects of bottom-up attention further along the visual processing pathways (PPC) (Silver et al., 2005).

The following cortical areas play a role in top-down attentional control. The visually selective regions of the PPC (posterior parietal cortex) – the intraparietal sulcus (IPS) in humans and lateral intraparietal area (LIP) in primates -, which contain coarse representations of spatial topography, and are also involved in controlling eye movements (saccades) and directing them towards targets. In this area, neuronal activity correlated with the voluntary

allocation of attention can be observed, and leads to greater target stimulus selectivity based on spatial location and/or salience (Knudsen, 2007; Serences & Yantis, 2006). The frontal eye field (FEF) in the pre-frontal cortex is involved in saccade control as well, but has also been shown to play a role in representing the current locus of attention (Serences & Yantis, 2006) – FEF neurons can be covertly selective to targets even when a saccade is planned in a different direction (Yuan et al., 2004). Finally, when a particular stimulus is attended to and conscious, neurons representing the target in sensory areas, in the PPC and in the PFC exhibit synchronized discharges in the gamma band (Knudsen, 2007; Doesburg et al, 2009 – see also Section 1.3).

Some subcortical structures have also been shown to play a role in top-down attention. The superior colliculus (SC), like FEF, mediates both overt saccades and covert shifts of attention (Serences & Yantis, 2006). More specifically, corollary discharges associated with eye saccades occur in the superior colliculus and propagate to the FEF, via the mediodorsal thalamic nucleus, and shift the locations of visual receptive fields in FEF before each saccade (Knudsen, 2007).

Recently, the locus coeruleus (LC) – norepinephrine (NE) – system has been shown to influence top-down attentional selection (Sara, 2009; Warren et al., 2009; Nieuwenhuis et al., 2005). LC neurons exhibit increased activity during the processing of motivationally salient targets, leading to the release of NE in widespread cortical projection areas, among others in the forebrain. This increased NE presence increases the responsibility of target neurons, enhancing signal detection and stabilizing a neural representation in the face of noise or interference. Thus, LC activity can facilitate the processing of a target stimulus (Cohen et al., 2004; Ashton-Jones & Cohen, 2005; Warren et al., 2009; Nieuwenhuis et al., 2005).
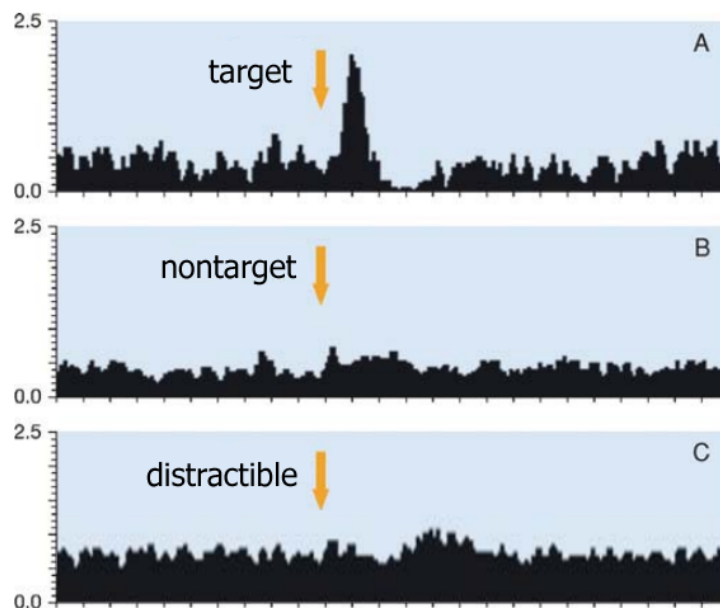
Figure 20. Peri-event time histograms (PETHs) for a typical individual monkey LC neuron in response to various events during performance of the signal detection task, accumulated over 100 sweeps of activity. Note the increased activity during target processing (panel A). Adapted from (Ashton-Jones & Cohen, 2005)

Shortly after target processing, there is a large increase of LC neuronal activity (top panel in Figure 20). The LC is autoinhibitory – increased activity during target processing is followed by a period of reduced activity and thus reduced NE release. Non-target stimuli do not elicit significant changes in LC activity (middle panel). Interestingly, tonic LC activity levels were significantly higher when the animal was less focused on the task

(bottom panel), but there was also a much smaller target-locked phasic response in this case. Behaviorally, these elevated LC activity levels have led to more frequent false alarm errors (Ashton-Jones et al., 1999).

These properties of the LC-NE system play a key role in LIDA's attentional blink model, as can be seen in the next section.

5.3.2 A LIDA-based Attentional Blink Model

As mentioned above, the major components implementing top-down attention in the LIDA model are the Global Workspace Module and the Attention Codelet Module (see Section 1.5 and Figure 5 for a description of the LIDA cognitive cycle). Bottom-up attentional control is less relevant for the modeled RSVP paradigm, but is also accounted for by the LIDA model by earlier processing stages in the cognitive cycle. Feature detectors (corresponding to feature-sensitive neurons in the visual cortices – see also Section 4) pass activation to their corresponding Perceptual Associative Memory (PAM) nodes, which represent objects (or categories, or events) (McCall et al., 2010) and could correspond to neuronal ensembles in the inferior temporal cortex, which contains object category information (Liu et al., 2009). The resulting activation of PAM nodes will depend on the number of relevant and identified features, as well as the salience of those features (e.g. specialized feature detectors could pass large amounts of activation when observing a bright flashing light, such as a fire alarm).

As already mentioned above, the Attenion Codelet Module contains Attention Codelets, which create coalitions from important or relevant percepts in the Global Workspace. The resulting coalition will have an activation value that depends on four factors: a) the activations of the percepts it contains, b) the base level activation of the Attention Codelet, c) the modulatory activation of the Attention Codelet Module and d) a matching factor on how well the percept matches the pattern that the Codelet is looking out for. It should be noted that the computational implementation of the LIDA Attentional Blink agent also contains a fifth factor, e) stochastic noise, which is added to account for extraneous, uncorrelated afferent activity (Knudsen, 2007; Nieuwenhuis et al., 2005).

The first factor corresponds to bottom-up salience in the brain, as described above. The second factor, the base level activation, depends on how useful the Codelet has been in the past and facilitates attentional learning.

The third factor contributing to the activation of coalitions in the Global Workspace is the modulatory activation of the Attention Codelet Module. It has been proposed many times in attention literature that human attentional processing is limited for targets presented in short succession – observable, among others, in an attentional blink paradigm -, presumably because of a suppression of attentional enhancement of subsequent stimuli during the processing of a target (Wyble et al., 2009; Olivers & Meeter, 2008; Nieuwenhuis et al., 2005). The modulatory activation reflects this mechanism, and regulates attentional enhancement of stimuli by increasing or decreasing the activation of coalitions in the Global Workspace. The most probable neural counterpart of this regulatory activity is the locus coeruleus (LC), which can enhance target processing through the release of norepinephrine in the forebrain, as described in the previous section (LC activity was proposed to play a role in the attentional blink by Nieuwenhuis et al. (2005)). Similarly to LC neuron activity, the Attentional Codelet Module activation at first strongly increases upon processing a relevant or important target, which is followed by a period of low activation which is similar to the posttarget refractory-like autoinhibition exhibited by the LC (see Figure 21 for a comparison of LC activity and the Attention Codelet Module activation).

Finally, the fourth parameter influencing coalition activation is a matching factor which is based on how well the percept in a coalition matches the pattern sought by the Attentional Codelet that creates the coalition. This accounts for the finding that in some cases, nontargets are attended to and reported instead of the targets if they are similar or share a common salient feature (Lavie & Cox, 1997; Bichot & Schall, 1999; Martens & Wyble, 2010) although with less probability and less selective neuronal activation (Duncan et al., 1997).
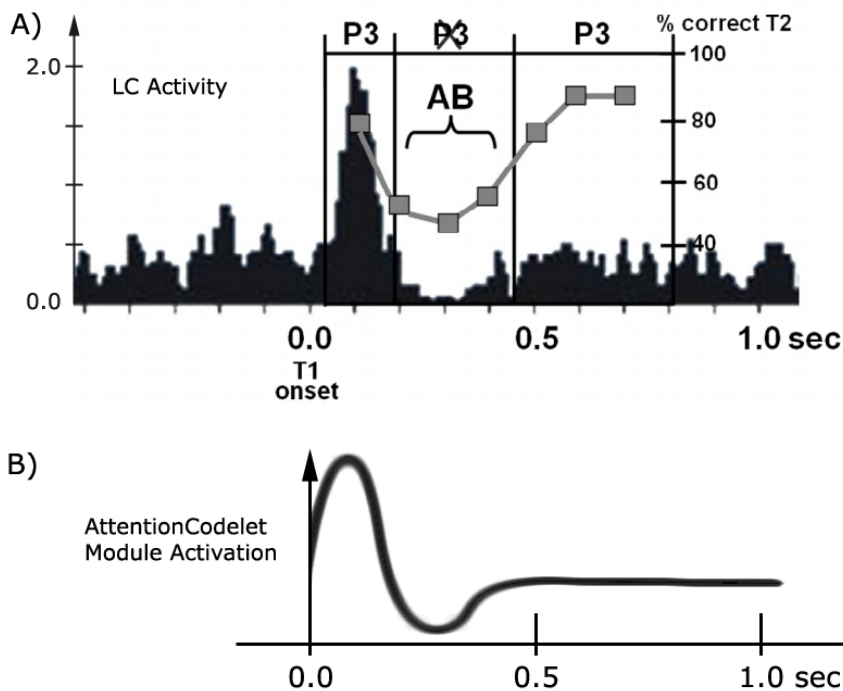
Figure 21. LC activity, Attentional Blink performance, and the Attention Codelet Module activation. A) The background graph shows a PETH of a monkey LC neuron during target processing (see Figure 20). The overlaid gray curve shows attentional blink T2 accuracies at the respective lags (a gray point at 0.1s corresponds to lag-1, etc.). The crossed-out P3 refers to the absence of the P3 ERP during the blink, i.e. during the period between 0.2 and 0.5s where T2 accuracies are the lowest – note that this period also coincides with the minimum LC activity.

LIDA's attentional mechanism can provide a formal (computational, see Section 5.4) explanation for the attentional blink and related findings. Two major reasons are proposed to account for the performance drop at intervals of 200ms – 500ms between the two targets (see Figure 21): a) the posttarget refractory-like period of the Attention Codelet Module activation, which leads to reduced target activations after ~200ms, and b) the discrete, competitive conscious broadcast mechanism (see Section 1.4).

For the current description, an RSVP attentional blink paradigm with images is assumed (similarly to the study conducted on human subjects in Potter et al. (2010)). Stimuli are presented to the LIDA agent at a rate of one image every 107ms. The agent's task is to report target images pertaining to a specific target (in this case, vehicles), which means that there are at least two Attention Codelets, looking out for targets (vehicles) and distractors, respectively. This is also the paradigm used for the implementation of the LIDA Attentional Blink agent (see Section 5.4).

If only a single target is presented, that target is added to a coalition by the Target Attention Codelet, will win the competition for consciousness since there is nothing that could compete with the coalition, and can be consciously reported (see also the description of the LIDA cognitive cycle in Section 1.5).

If two targets are presented in an RSVP of images at lag-1, without a distractor, both targets are perceived in the first 200ms – before the refractory-like period of the Attention Codelet Module – and they are both added to a target coalition by an Attention Codelet looking out for targets. Target Attention Codelets have significantly higher base level activation than Distractor Attention Codelets. Thus they will win the conscious broadcast and can be reported consciously. Possible subsequent targets are also added to the target coalition by the same
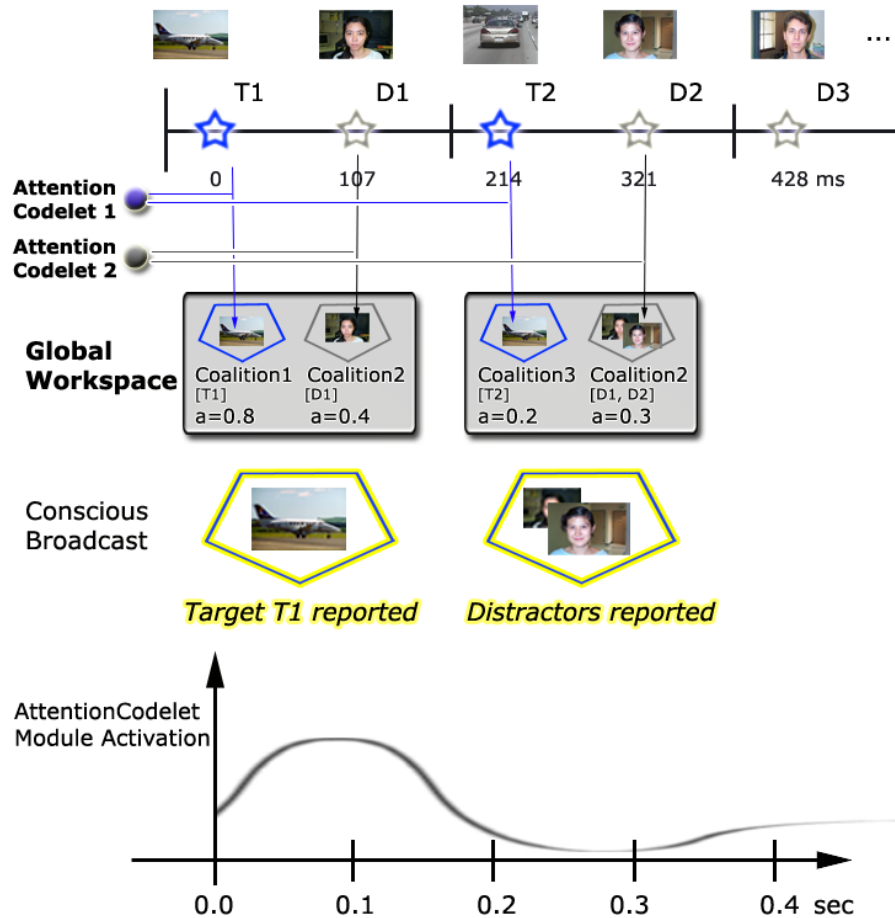
Figure 22. The attentional blink at lag 2. *Tn and Dn refers to targets and distractors, respectively. The vertical black lines intersecting with the timeline on top represent the approximate borders of LIDA cognitive cycles. Attention Codelet 1 is looking out for targets, and Attention Codelet 2 for distractors, adding them to Coalitions in the Global Workspace. The coalitions have to compete for consciousness, and the one with the highest activation is broadcast consciously. The reason the agent fails to report T2 is that in the second cognitive cycle, Coalition 2 (containing the distractors) wins the competition for consciousness.*

Attention Codelet, which adjusts the coalition activation based on the factors described above and on the previous coalition activation – this accounts for the spread lag-1 sparing effect.

At lag 2, the second target cannot be reported consciously because a coalition containing distractors wins the competition for consciousness instead of T2 (see Figure 22). The reason for the low activation of the target coalition is the low Attention Codelet Module activation at this point in time (due to the refractory-like period, see Figure 21). D2 is added to the distractor coalition by the Distractor Attention Codelet, and the coalition activation is updated. The distractor coalition is also modulated with a lower Attention Codelet Module activation, but will come out with a higher activation because a) depending on the timing of the presentation, the Attention Codelet Module activation might be higher at the point the distractor is perceived than at the point when the target is perceived, and b) since the distractor coalition was created upon perceiving D1, at which point the coalition activation was higher (0.4 in Figure 22, due to the high AttentionCodelet Module activation at that point).

The situation is similar at lag 3, although the target coalition will have a higher activation in these cases since the Attention Codelet Module will have regenerated some of its activation (see Figure 21). Distractors will still win the competition for consciousness sometimes – the stochasticity arising from the small random addition to the coalition activation – which causes a low accuracy.

At lag 4, the Attention Codelet Module has been restored to its initial level of activation, and T2 can be reported with a high level of accuracy again (the T2 accuracy at lag 4 approximately equals T2 accuracy at lag 1 in this paradigm, see Potter et al. (2010)).

LIDA's attentional mechanism accounts for all of the attentional blink-related phenomena described in Section 5.1:

1) and 2) **Lag-1 sparing and spread lag-1 sparing.** See explanation above.

3) **Posttarget intrusion.** During the blink, the distractor succeeding T2 often can be consciously reported even if T2 itself cannot (see Figure 22).

4) **Whole report attenuates the AB.** In case of an instruction to report the entire RSVP sequence, a different Attentional Codelet would be required, which would move every presented image into the Global Workspace and into the same coalition – every image would be a target. Thus, for short RSVP sequences, every image could be reported and no AB could be observed (if the sequence is too long, activation decay could lead to "forgetting" of the first images. There is also a limit on how much information the Workspace and the Global Workspace can hold, although this limit has not been quantitatively determined yet).

5) **Increasing T2 salience attenuates the AB.** Increased T2 bottom-up salience leads to a higher activation of the PAM node representing T2 and thus to a target coalition with a higher activation, which increases the probability that T2 wins the competition for consciousness.

6) **Task-irrelevant cognitive load attenuates the AB.** Subject less focused on a task exhibit higher levels of tonic LC activity (see lowest panel of Figure 20), which can explain this phenomenon. In experimental conditions in which moving dots are presented around the target, and in conditions where the subject is instructed to think about something else (see Section 5.2), subjects are less focused on the AB task – therefore their AttentionCodelet Module Activation (modeling LC activity) is higher before a target, and there is a much smaller post-target activation drop, which makes it possible to almost always report T2 accurately – in accordance with behavioral AB experiments with distracted subjects (Arend et al., 2006; Olivers & Nieuwenhuis, 2005).

7) **Using target-similar distractors increases the AB** (for example, an attentional blink while recognizing colored letter targets in a letter stream is stronger than recognizing the same targets in a stream of digits)**.** T2 performance in case of target-similar distractors depends on the Target Attention Codelet; specifically, how target percepts are selected based on the Codelet's "sought content" template. In the example, there could be a Distractor Attention Codelet and two Target Attention Codelets, one looking out for colored stimuli and one looking out for letters. Both of these codelets would add to the activation of a coalition containing real targets (stimuli that are letters and are colored), but stimuli only possessing one of these properties would also be added to a coalition. Such a coalition would have stronger activation than distractors (since Target Attention Codelets have more base activation than Distractor Codelets) and have an increased probability to win the competition for consciousness against real targets.

8) **Target Confusion.** Targets presented temporally adjacently in the same cognitive cycle (e.g. during lag-1 sparing) land in the same coalition. Since coalitions do not contain ordering information, the temporal order of the two targets is unknown to the agent. This could account for the target confusion effect. However, no

exact computational mechanism has been implemented yet for reproducing how human subjects "guess" (often incorrectly, see (Chun, 1997; Dux & Marois, 2009)) the first target in such a case.

9) **AB without T1 masking.** There is still an AB effect without T1 mask since the AB in this model is due to a) the refractory-like period of the Attention Codelet Module and b) intrusion of the post-T2 distractor, and does not depend on the post-T1 distractor (unlike some of the models described in the next section)

10) **T2 cueing attenuates AB.** This has a similar reason to finding 8), Target Confusion. As described above, if targets are defined by multiple properties, there are multiple Attention Codelets watching out for these properties. If a distractor that shares one or more of these properties with the target is perceived, it is put into a coalition and receives activation from one or more of these Target Attention Codelets. It also sets off the temporary increase in Attention Codelet Module activation (see Figure 22), which leads to an increased activation for targets presented in the next ~200ms, leading to an attenuated AB effect.

5.3.3 Comparison with other Attentional Blink Models

This section will briefly summarize other attentional blink models which have been formally implemented (based on (Dux & Marois, 2009), which contains a more detailed review of these models, and other, purely theoretical models not described here). The first six described models are connectionist (neural network-based) approaches (gated auto-associator, CODAM, LC-NE, global workspace model, boost & bounce and eSTST), while the last two are mathematical (attention cascade model) and symbolic (threaded cognition model), respectively.

The **Gated auto-associator model** Chartier et al. (2004) predicts that stimuli are evaluated via two networks, with one performing number identification and feeding this information into a working memory, and another comparing the stimulus color with the target instructions. There is also an attentional gate which determines which stimuli can enter into the working memory. This gate opens when a target is identified, but becomes inhibited during the encoding of a target into working memory and thus does not admit subsequent targets during this period, which leads to the AB.

The **Corollary discharge of attention movement model (CODAM)** (Fragopanagos et al., 2005) proposes that stimuli go through an input and object map module and are subsequently stored in working memory. A so-called inverse model controller (IMC) boosts items in the object map so they can enter the working memory – however, the IMC is suppressed by a monitor module comparing the current item with the current target representation during target item processing. At short lags, the target representation the monitor is comparing with is still T1 even if T2 is presented, therefore T2 doesn't enter working memory, giving rise to the AB.

The **Locus coeruleus–norepinephrine model (LC-NE)** (Nieuwenhuis et al., 2005), similarly to the LIDA-based model detailed in the previous section, proposes that the AB reflects the activation dynamics of the locus coeruleus. Nieuwenhuis et al.'s model consists of a behavior network that is made up of input, detection, and decision layers; and the LC network adjusting the behavior network's gain. Following the LC's phasic response to a target, it is suppressed and unable to facilitate subsequent target detection.

The **global workspace model of the attentional blink** by Dehaene et al. (2003), like LIDA, is based on the Global Workspace Theory (Baars, 2005). The main ideas are an all-or-none mechanic - that only one item can enter consciousness at a time -, and that items (i.e. neuronal ensembles) that are conscious are capable of inhibiting other items, which leads to the AB.

The **Boost and bounce theory** (Olivers & Meeter, 2008), unlike the previously described models, does not depend on capacity limits. It has two stages, sensory processing and working memory. Perceived items' activation strengths are influenced by stimuli appearing before or after them due to backward/forward masking. The working memory maintains task instructions, stores encoded representations where items have been linked to a response. It also employs a filter enhancing items that match the target description and inhibiting items that don't – it inhibits items before T1 and enhances or "boosts" T1, which can thus enter working memory. If T2 is

presented temporally close, it will also receive a boost. Distractors cause a strong but transient suppression of subsequent stimuli, thus causing the AB.

The **Episodic simultaneous type/serial token model (eSTST)** (Bowman & Wyble, 2007) is also a two-stage connectionist model, predicting that all stimuli in an RSVP are identified in a conceptual stage ("type" representation). If a stimulus is also a target, it is attentionally enhanced by a transiently triggered blaster, and subsequently have their identity information bound to a "token" in working memory which contains episodic information (e.g. relative position in the RSVP stream). After a target has been enhanced, the blaster is suppressed until the target is bound to its token – thereby causing the AB.

The attention cascade model (Shih, 2008) is a mathematical model that does not claim neural plausibility but explains the attentional blink with fewer parameters than the models described above. According to this model, stimuli are initially processed among one of two channels: a mandatory pathway and a bottom-up salience pathway. Stimuli in the mandatory pathway activate conceptual representations that are passed into a peripheral sensory buffer, and can trigger an attentional window and be enhanced if they match a target template. Following this enhancement, the target is consolidated and passed into the decision processor (within working memory). Stimuli with strong bottom-up salience can also trigger an attentional window and enter directly into the decision processor along the bottom-up salience pathway.

The threaded cognition model (Taatgen et al., 2009) is loosely based on the ACT-R (Anderson et al., 2004) cognitive architecture, hypothesizing cognition to comprise of multiple processes that are threaded through a single processor. According to this model, target detection is held offline during the encoding of an additional target in working memory. Thus, T2 detection is suppressed at short lags because T1 consolidation is still underway – except for adjacent targets (e.g. lag-1), for which the system overrides this suppression.

↓ AB related phenomenon	LIDA	eSTST (Bowman & Wyble, 2007)	Attention Cascade (Shih, 2008)	Threaded cognition (Taatgen et al., 2009)	Gated Auto-Associator (Chartier et al., 2004)	CODAM (Fragopanagos et al., 2005)	LC-NE (Nieuwenhuis et al., 2005)	Global Workspace Dehaene et al. (2003)	Boost and Bounce (Olivers & Meeter, 2008
Accounts for non-reported stimuli undergoing postperceptual processing	✔	✔	✔	✔	✗	✔	✔	✔	✔
Lag-1 sparing	✔	✔	✔	✔	✔	✔	✔	✔	✔
Spreading the sparing	✔	✔	✔	✔	✗	✗	✗	✗	✔
Posttarget intrusion	✔	✔	✔	✔	✔	✔	✔	✔	✗
Whole report attenuates the AB	✔	✔	✔	✔	✗	✗	✗	✗	✗
Increasing T2 salience attenuates the AB	✔	✗	✔	✗	✗	✗	✗	✔	✗
Task-irrelevant cognitive load attenuates the AB	✔	✗	✗	✗	✗	✗	✔	✗	✗
Using target-similar distractors increases the AB	✔	✔	✔	✔	✗	✔	✗	✗	✔
Target confusion	✔*	✔	✗	✗	✗	✗	✗	✗	✗
AB without T1 masking	✔	✔	✔	✗	✗	✗	✔	✔	✗
T2 cueing attenuates AB	✔	✔	✔	✔	✗	✔	✔	✗	✗

Table 6. A comparison of formal models of the attentional blink, with respect to their ability to account for the AB-related phenomena described above. Green check marks denote phenomena that are accounted for conceptually by the model in that column.

Table 6 above attempts to compare the large number of formal models of the attentional blink by indicating their ability to account for the AB related findings and phenomena described above. The columns are sorted by the number of findings accounted for. As can be seen from this table, the LIDA-based attentional blink model is able to account for the most findings from the models. It should be noted that green check marks denote phenomena that can be conceptually accounted for – some of these have not been explicitly modeled (yet).

5.4 The LIDA Attentional Blink Agent

To show that LIDA's attention model can model human behavior, the attentional blink experiment conducted by Potter et al. (2010) has been reproduced with a LIDA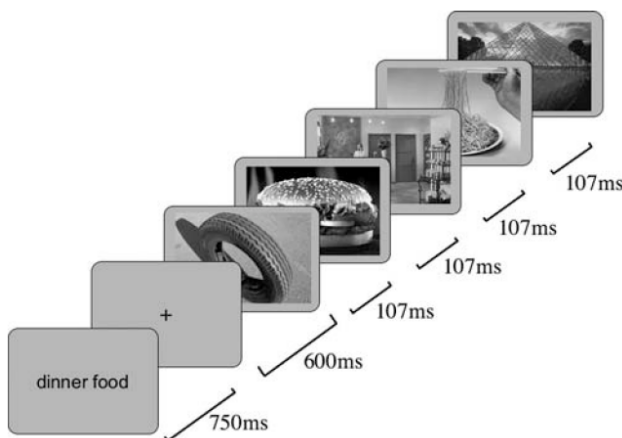-based cognitive software agent. In this paradigm, subjects were shown a written description of the target category (e.g. dinner food, bird, or fruit), and subsequently a rapid serial visual presentation of eight color photographs, consisting of six distractors and two target pictures. The first target, T1 appeared in serial position 2 or 3, and the second target, T2, followed at lag 1 (after 107ms), lag 2 (after 213ms) or lag 4 (after 427ms). Figure 23 illustrates this paradigm. After all images were presented, subjects were instructed to give the names of the target objects (not their category) in the order of presentation in their own time.

Figure 23. The visual attentional blink paradigm reproduced by the LIDA Attentional Blink Agent. (From Potter et al. (2010))

Human performance averaged over the 72 trials that each of the 24 participants performed can be seen in Figure 25.

The LIDA Attentional Blink Agent operates in a very similar environment. Instead of taking random pictures from Google Images, we used the Caltech image database (Fergus et al., 2003) – see Figure 17 in Section 4.4 for a few example images. Vehicles were the target category that the agent had to report; target images could be cars or airplanes. Faces were used as distractor images to allow for easy discrimination. Figure 24 contains a screenshot of the agent and its environment.

The task instructions were predefined in the agent in the form of two Attention Codelets, one of which selected targets, bringing them to the Global Workspace, and the other selecting distractors (see Figure 22 in the previous section). There were also predefined behavior schemes in the Procedural Memory for reporting the targets by pressing buttons (one for each of the target categories, and one for no target perceived, see Figure 24). While it would also be possible in the conceptual model to learn such task settings (D'Mello et al., 2006), no concrete computational implementation for task learning exists at the moment.

The LIDA Attentional Blink (AB) Agent, like the Reaction Time and Allport Agents, is based on the computational LIDA framework – thus, it's cognitive cycle works in the same way, as described in Section 3.4.1. To avoid redundancy, only those processing stages which differ significantly from these agents shall be described here.

The Sensory Memory periodically inspects the Environment, and copies the current image into a visual buffer. It also pre-processes the image data using the SURF algorithm, extracts features, and copies them into the SURF Vector Space layer (see Section 4). Feature Detectors periodically inspect this vector space and pass activation to their parent PAM node. In the example displayed in Figure 24, most feature detectors associated with the airplane node would find their feature, leading to the correct PAM node obtaining the highest activation - although some

Figure 24. The LIDA Allport Agent, and its simple environment. *The top panel contains a screenshot of the agent's graphical user interface, with the sensory memory pixel layer (top left) and the perceptual associative memory node activations (top right). The airplane is perceived since the airplane node has the highest activation (note that this does not automatically mean that it also becomes conscious). The bottom panel illustrates the simple environment that the agent operates in. There is a screen which displays eight images in this rapid serial visual presentation paradigm (bottom left), and three buttons. The agent selects an appropriate action based on the incoming image and presses the "Airplane" button if it consciously perceives an airplane, the "Car" button if it consciously perceives a car and the "Distractor" button if the percept entering consciousness is not part of the target category.*

feature detectors associated with the car and face nodes will also find a feature and pass activation to the wrong object node, since features can be ambiguous.

Upon exceeding its threshold, the identified node is copied into the Workspace (this takes approximately 100ms, see Section 3 for the timing of the processes described here). Attentional Codelets inspecting the Workspace try to locate their sought content and to move it to the Global Workspace as a new coalition – in this example, the Target Attention Codelet would identify the target node, create a new coalition based primarily on the activation

of the node, the modulatory activation of the Attention Codelet Module, and the Codelet's base activation (see Section 5.3.2 and Figure 22), and put this coalition into the Global Workspace. Assuming that this coalition is the one with the highest activation (e.g. if it is the first target presented, or the second but presented at lag-1), it will win the competition for consciousness and be broadcast to the Procedural Memory Module. Of the schemes present in the Procedural Memory, the one representing the behavior to press the "airplane" button will receive the most activation (since it's context or precondition best matches the conscious percepts). This behavior will be instantiated in the Action Selection module, and the action to press the button will be selected. Thus, the agent will indicate that it consciously perceived an airplane.

Figure 25 contains the performance of the LIDA Attentional Blink Agent (reporting accuracies at lags 1, 2 and 4), and compares it to the performance of human subjects.
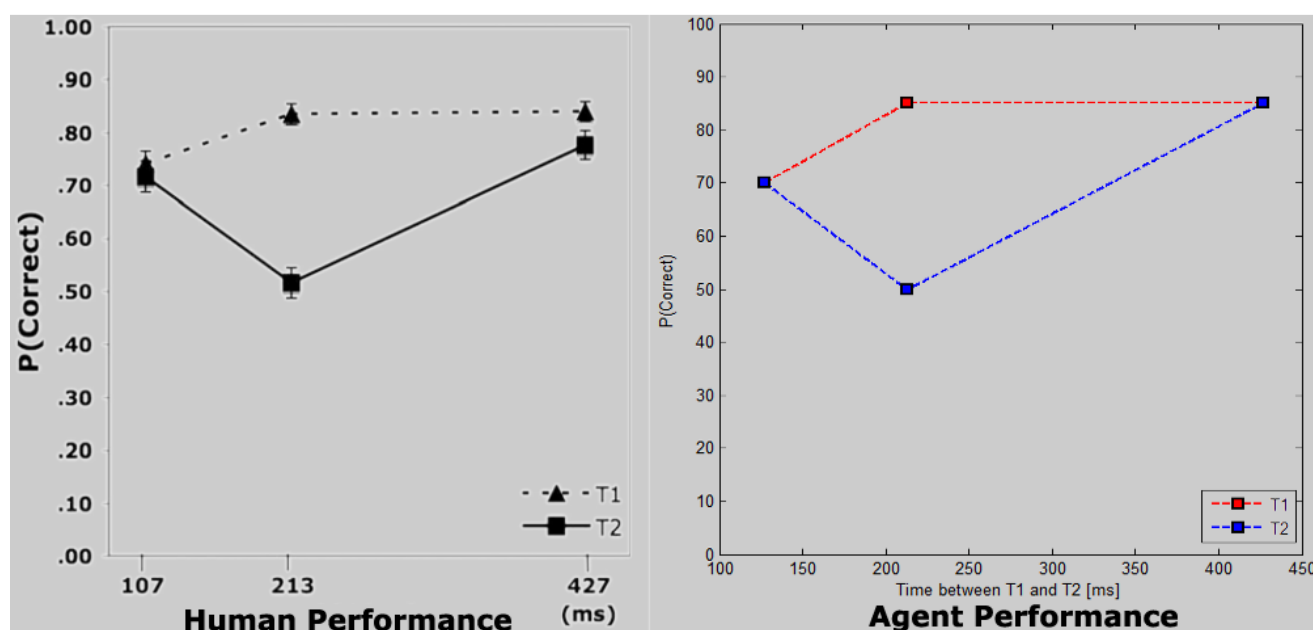
Figure 25. Comparison of human performance and the LIDA Attentional Blink Agent's performance in a visual attentional blink paradigm. The y axis displays the percentage of correctly reported second targets (T2), conditional on the first target (T1) being reported correctly. The x axis represents the time between the two targets. Left panel adapted from (Potter et al., 2010)

6. CONCLUSION

This work presented an attempt to increase the plausibility of the LIDA model, i.e. to substantiate the claim that LIDA accurately models basic human cognitive processes using mechanisms similar to the human brain.

Subprocesses in LIDA's cognitive cycle were correlated with processes in the brain (Section 2), their timings estimated based on recent neuroscientific evidence, and converted into parameters for the LIDA model.

Based on the LIDA computational framework and on these parameters, three cognitive software agents were developed in order to substantiate the LIDA model: a) the LIDA Reaction Time Agent, b) the LIDA Allport Agent, and c) the LIDA Attentional Blink Agent (see Sections 3 and 5). Facilitating the development of these agents, two new models were developed, implemented and added to LIDA's implementation: a) a vision model based on invariant feature detection (Section 4), and b) a model of attention and its deficits based on the locus coeruleus – norepinephrine (LC-NE) system (Section 5).

6. Conclusion

Apart from showing that the mechanisms put forth by LIDA can indeed lead to agents behaving comparably to humans in three different psychological paradigms (using the same framework and the same set of parameters), the developed software agents strengthened LIDA's major hypotheses in the following way (see Section 1.2 for an explanation of these hypotheses):

- *The Cognitive Cycle Hypothesis*
  All three agents contributed to showing that human cognition can be comprised of cognitive cycles – that these cycles might be the building blocks or "atoms" of cognition. The LIDA Reaction Time Agent and the LIDA Attentional Blink Agent both performed psychological experiments in comparable timeframes (and, in the case of the attentional blink, with comparable accuracies) to human subjects, showing that a perception-action cycle in LIDA takes approximately as long as in humans. The LIDA Allport Agent showed that the time boundaries of functional consciousness are very similar in LIDA agents and in humans. Furthermore, the mechanisms the agents use to arrive at these behaviors were detailed and shown to correlate to neural mechanisms (See Section 3 for the LRT and Allport Agents, and Section 5 for the Attentional Blink Agent.)

- *The Consciousness is Discrete Hypothesis*
  The LIDA Allport Agent's behavior in a perceptual continuity task has shown that a discrete functional consciousness paradigm can accurately model human experimental data (Section 3).

- *The Theta-Gamma Coupling Hypothesis*
  This hypothesis was mainly based on consciousness research in neuroscience. The time-frames of the theta-gamma synchronous phases in the brain are consistent with the time frames of conscious broadcasts in the LIDA-based cognitive software agents. However, this is not a hypothesis that can be proven by computational modeling. Section 1.4 presented strong brain related evidence derived from published experiments.

- *The LC-NE Hypothesis* is based on attention research and on a previously published (and much less comprehensive) model of the attentional blink. The LIDA Attentional Blink agent and the fact that it's behavioral performance matches human behavior; as well as the ability of the model to conceptually account for most of the findings associated with the attentional blink (unlike previously published formal models of the attentional blink) as well as neural data provide evidence for its plausibility (see Section 5).

In order to claim that LIDA accurately models the human brain with reasonable plausibility, a much wider range of experimental evidence from cognitive software agents performing different high-level cognitive tasks would be required. This work only presents an initial effort in that direction. Further steps could include modeling a much larger range of psychological experiments, as well as reproducing human behavior in embodied, real world settings. A possible approach for this latter point could be the development of an embodied agent successfully performing the "Cognitive Decathlon" (Mueller et al., 2007), a kind of simplified, behavioral Turing test, and comparing its behavior to human children. In the end, no cognitive model inferior to human cognition can claim it has fully succeeded in accounting for all cognitive processes – at some points, cognitive architectures will need to pass the real Turing test (Turing, 1950) and a Meta-Turing test (Sloman, 2008) to show it is capable of human level cognition and meta-cognition; although even the most modern approaches are still very far away from being able to pass the Turing test (Floridi et al., 2009). However, despite this fact, it is my strong belief that if human-level machine intelligence becomes possible in the future, cognitive modeling will be the most important tool leading science to that achievement.

7. REFERENCES

Agam, Y., H. Liu, A. Papanastassiou, C. Buia, A. J Golby, J. R Madsen, and G. Kreiman. 2010. Robust Selectivity to Two-Object Images in Human Visual Cortex. *Current Biology*.

Albus, J. S. 2001. Engineering of mind: An introduction to the science of intelligent systems.

Allport, D A. 1968. Phenomenal simultaneity and the perceptual moment hypothesis. *British Journal of Psychology* 59 (4): 395-406.

Anderson, J. R, and M. Matessa. 1997. An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction* 12 (4): 391–438.

Anderson, J. R, D. Bothell, M. D Byrne, S. Douglass, C. Lebiere, and Y. Qin. 2004. An integrated theory of the mind. *Psychological Review* 111 (4): 1036–1060.

Anderson, J.R., M. Matessa, and C. Lebiere. 1997. ACT-R: A theory of higher level cognition and its relation to visual attention. *Human Computer Interaction* 12(4): 439-462.

Antwerpen, B. 2009. When the Brain Changes Its Mind: Flexibility of Action Selection in Instructed and Free Choices. *Cerebral Cortex* 19: 2352–2360.

Arel, I., D. Rose, and R. Coop. 2009. Destin: a scalable deep learning architecture with application to high-dimensional robust pattern recognition. In *Proc. of the AAAI 2009 Fall Symposium on Biologically Inspired Cognitive Architectures (BICA)*.

Arend, I., S. Johnston, and K. Shapiro. 2006. Task-irrelevant visual motion and flicker attenuate the attentional blink. *Psychonomic Bulletin & Review* 13 (4): 600–607.

Aston-Jones, G., and J. D Cohen. 2005. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* 28: 403–450.

Aston-Jones, G., J. Rajkowski, and J. Cohen. 1999. Role of locus coeruleus in attention and behavioral flexibility. *Biological Psychiatry* 46 (9): 1309–1320.

Baars, B. J. 1988. *A cognitive theory of consciousness*. Cambridge Univ Pr.

Baars, B. J. 2005. Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Progress in brain research* 150: 45–53.

Baars, B. J, and S. Franklin. 2007. An architectural model of conscious and unconscious brain functions: Global Workspace Theory and IDA. *Neural Networks* 20 (9): 955–961.

Baars, B. J, and S. Franklin. 2009. Consciousness is computational: The LIDA model of global workspace theory.

Baars, Bernard J. 1997. *In the theater of consciousness: the workspace of the mind*. Oxford University Press US.

Baddeley, A. 1992. Working memory. *Science* 255 (5044): 556.

Baddeley, A. D., and G. Hitch. 1974. *Working memory. I G. Bower (Ed.) The Psychology of Learning and Motivation, Vol. 8*. New York: Academic Press.

Barnard, P. J., S. Scott, J. Taylor, J. May, and W. Knightley. 2004. Paying attention to meaning. *Psychological Science* 15 (3): 179.

Barnes, Gareth, Arjan Hillebrand, and Masayuki Hirata. 2010. Magnetoencephalogram. *Scholarpedia* 5 (7): 3172.

Barsalou, L. W. 1999. Perceptual symbol systems. *Behavioral and brain sciences* 22 (04): 577–660.

Bauer, Markus, Robert Oostenveld, and Pascal Fries. 2009. Tactile stimulation accelerates behavioral responses to visual stimuli through enhancement of occipital gamma-band activity. *Vision Research* 49 (9): 931-942. doi:10.1016/j.visres.2009.03.014.

Bay, H., A. Ess, T. Tuytelaars, and L. Van Gool. 2008. Speeded-up robust features (SURF). *Computer Vision and Image Understanding* 110 (3): 346–359.

Bear, Mark F., Barry W. Connors, and Michael A. Paradiso. 2007. *Neuroscience: exploring the brain*. Lippincott Williams & Wilkins.

Berger, H. 1929. Über das Elektroenkephalogramm des Menschen. *European Archives of Psychiatry and Clinical Neuroscience* 87 (1): 527–570.

Van Berkum, J. J. A. Discourse and the brain.

Besle, Julien, Catherine Fischer, Aurelie Bidet-Caulet, Francoise Lecaignard, Olivier Bertrand, and Marie-Helene Giard. 2008. Visual Activation and Audiovisual Interactions in the Auditory Cortex during Speech

Perception: Intracranial Recordings in Humans. *J. Neurosci.* 28 (52): 14301-14310. doi:10.1523/JNEUROSCI.2875-08.2008.

Bichot, N. P, and J. D Schall. 1999. Effects of similarity and history on neural mechanisms of visual selection. *Nature Neuroscience* 2: 549–554.

Block, N. 1995. On a Confusion about a Function of Consciousness. *Behavioral and Brain Sciences* 18: 227–247.

Bolognini, N., and T. Ro. 2010. Transcranial Magnetic Stimulation: Disrupting Neural Activity to Alter and Assess Brain Function. *Journal of Neuroscience* 30 (29) (July): 9647-9650. doi:10.1523/JNEUROSCI.1990-10.2010.

Bowman, H., and B. Wyble. 2007. The simultaneous type, serial token model of temporal attention and working memory. *Psychological Review* 114 (1): 38.

Buzsáki, G. 2006. *Rhythms of the Brain*. Oxford University Press, USA.

Canolty, R. T., E. Edwards, S. S. Dalal, M. Soltani, S. S. Nagarajan, H. E. Kirsch, M. S. Berger, N. M. Barbaro, and R. T. Knight. 2006. High gamma power is phase-locked to theta oscillations in human neocortex. *science* 313 (5793): 1626.

Capaday, C, R Forget, R Fraser, and Y Lamarre. 1991. Evidence for a contribution of the motor cortex to the long-latency stretch reflex of the human thumb. *The Journal of Physiology* 440: 243-255.

CCRG. 2011. The Computational Lida Framework. http://ccrg.cs.memphis.edu/framework.html.

Chartier, S., D. Cousineau, and D. Charbonneau. 2004. A connectionist model of the attentional blink effect during a rapid serial visual task. In *Proceedings of the 6th International Conference on Cognitive Modelling*, 64–69.

Chong, H. Q, A. H Tan, and G. W Ng. 2007. Integrated cognitive architectures: a survey. *Artificial Intelligence Review* 28 (2): 103–130.

Chun, M. M. 1997. Temporal binding errors are redistributed by the attentional blink. *Attention, Perception, & Psychophysics* 59 (8): 1191–1199.

Chun, M. M, and M. C Potter. 1995. A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance* 21 (1): 109.

Cisek, Paul, and John F. Kalaska. 2010. Neural Mechanisms for Interacting with a World Full of Action Choices. *Annual Review of Neuroscience* 33 (1) (June): 269-298. doi:10.1146/annurev.neuro.051508.135409.

Conway, M. A. 2001. Sensory–perceptual episodic memory and its context: Autobiographical memory. *Philosophical Transactions B* 356 (1413): 1375.

Coull, J. T., and A. C. Nobre. 2008. Dissociating explicit timing from temporal expectation with fMRI. *Current opinion in neurobiology* 18 (2): 137–144.

Del Cul, A., S. Baillet, and S. Dehaene. 2007. Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biol* 5 (10): e260.

D'Mello, S. K, S. Franklin, U. Ramamurthy, and B. J Baars. 2006. A cognitive science based machine learning architecture. In *AAAI 2006 Spring Symposium Series Sponsor: American Association for Artificial Intelligence. Stanford University, Palo Alto, California, USA*.

Dehaene, S. 2008. Conscious and nonconscious processes: distinct forms of evidence accumulation. In *Decision making, the human mind, and implications for institutions. Str\üngmann forum reports (Engel C, Singer W, eds)*, 21–49.

Dehaene, S., and L. Naccache. 2001. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *The cognitive neuroscience of consciousness*: 1–37.

Dehaene, S., C. Sergent, and J. P Changeux. 2003. A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Sciences of the United States of America* 100 (14): 8520.

Doesburg, S. M, A. B Roggeveen, K. Kitajo, and L. M Ward. 2008. Large-Scale Gamma-Band Phase Synchronization and Selective Attention. *Cerebral Cortex* 18 (2): 386–396.

Doesburg, S. M., J. J. Green, J. J. McDonald, and L. M. Ward. 2009. Rhythms of consciousness: binocular rivalry reveals large-scale oscillatory network dynamics mediating visual perception. *PloS one* 4 (7): e6142.

7. References

Duch, W., R. J Oentaryo, and M. Pasquier. 2008. Cognitive Architectures: Where do we go from here? In *Proceeding of the 2008 conference on Artificial General Intelligence 2008: Proceedings of the First AGI Conference*, 122–136.

Duncan, J., G. Humphreys, and R. Ward. 1997. Competitive brain activity in visual attention. *Current Opinion in Neurobiology* 7 (2): 255–261.

Dux, P. E, and V. Coltheart. 2005. The meaning of the mask matters. *Psychological Science* 16 (10): 775.

Dux, P. E, and R. Marois. 2009. The attentional blink: A review of data and theory. *Attention, Perception, & Psychophysics* 71 (8): 1683–1700.

Engel, A. K, and W. Singer. 2001. Temporal binding and the neural correlates of sensory awareness. *Trends in cognitive Sciences* 5 (1): 16–25.

Ericsson, K. A, and W. Kintsch. 1995. Long-Term Working Memory. *Psychological Review* 102 (2): 211–245.

Evans, K. K, and A. Treisman. 2005. Perception of Objects in Natural Scenes: Is It Really Attention Free?. *Journal of Experimental Psychology: Human Perception and Performance* 31 (6): 1476.

Fergus, R., P. Perona, and A. Zisserman. 2003. Object class recognition by unsupervised scale-invariant learning.

Floridi, L., M. Taddeo, and M. Turilli. 2009. Turing's Imitation Game: Still an Impossible Challenge for All Machines and Some Judges—-An Evaluation of the 2008 Loebner Contest. *Minds and Machines* 19 (1): 145–150.

Foxe, J. J, and C. E Schroeder. 2005. The case for feedforward multisensory convergence during early cortical processing. *Neuroreport* 16 (5): 419.

Fozard, J. L, M. Vercruyssen, S. L Reynolds, and P. A. Hancock. 1990. Longitudinal analysis of age-related slowing: BLSA reaction time data. In *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 34:163–167.

Fragopanagos, N., S. Kockelkoren, and J. G Taylor. 2005. A neurodynamic model of the attentional blink. *Cognitive Brain Research* 24 (3): 568–586.

Franklin, S. 1995. *Artificial Minds*. MIT Press, September 11.

Franklin, S. 2003. IDA: A conscious artifact? *Journal of Consciousness Studies* 10 (4/5): 47–66.

Franklin, S. 2008. The LIDA model's hypotheses on the cognitive cycle, high-level cognitive processes, and brain rhythms.

Franklin, S., and A. Graesser. 1997. Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents. *Intelligent Agents III Agent Theories, Architectures, and Languages*: 21–35.

Franklin, S., and F. G. Patterson Jr. 2006. The LIDA architecture: Adding new modes of learning to an intelligent, autonomous, software agent. *pat* 703: 764–1004.

Franklin, S., B. J Baars, U. Ramamurthy, and M. Ventura. 2005. The role of consciousness in memory. *Brains, Minds and Media* 1 (1): 38.

Franklin, S., U. Ramamurthy, S. K D'Mello, L. McCauley, A. Negatu, R. Silva, and V. Datla. 2007. LIDA: A computational model of global workspace theory and developmental learning. In *AAAI Fall Symposium on AI and Consciousness: Theoretical Foundations and Current Approaches*, 61–66.

Franklin, Stan, Sidney D'Mello, and Austin Hunter. The LIDA Cognitive Cycle. http://ccrg.cs.memphis.edu/tutorial/synopsis.html.

Freeman, W. J. 1995. *Societies of brains: A study in the neuroscience of love and hate*. Lawrence Erlbaum.

Freeman, W. J. 1999. Consciousness, intentionality and causality. *Journal of Consciousness Studies, 6* 11 (12): 143–172.

Freeman, W. J. 2000. *How brains make up their minds*. Columbia Univ Pr.

Freeman, W. J. 2002. The limbic action-perception cycle controlling goal-directed animal behavior. *Neural Networks* 3: 2249–2254.

Fries, P. 2005. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences* 9 (10): 474–480.

Frund, Ingo, Niko Busch, Jeanette Schadow, Ursula Korner, and Christoph Herrmann. 2007. From perception to action: phase-locked gamma oscillations correlate with reaction times in a speeded response task. *BMC Neuroscience* 8 (1): 27. doi:10.1186/1471-2202-8-27.

7. References

Fuster, J. M. 2002. Physiology of executive functions: The perception-action cycle. *Principles of frontal lobe function*: 96–108.

Gaillard, R., S. Dehaene, C. Adam, S. Clémenceau, D. Hasboun, M. Baulac, L. Cohen, and L. Naccache. 2009. Converging intracranial markers of conscious access. *PLoS Biol* 7 (3).

Giard, M. H., and F. Peronnet. 1999. Auditory-Visual Integration during Multimodal Object Recognition in Humans: A Behavioral and Electrophysiological Study. *Journal of Cognitive Neuroscience* 11 (5): 473-490. doi:10.1162/089892999563544.

Giesbrecht, B., and V. Di Lollo. 1998. Beyond the attentional blink: Visual masking by object substitution. *Journal of Experimental Psychology: Human Perception and Performance* 24 (5): 1454.

Glenberg, A. M. 1997. What memory is for: Creating meaning in the service of action. *Behavioral and brain sciences* 20 (01): 41–50.

Glenberg, A. M, and D. A Robertson. 2000. Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language* 43 (3): 379–401.

Gluck, K. A, and R. W Pew. 2005. *Modeling human behavior with integrated cognitive architectures: Comparison, evaluation, and validation*. Lawrence Erlbaum Assoc Inc.

Goertzel, B., R. Lian, I. Arel, H. de Garis, and S. Chen. 2010. World survey of artificial brains, Part II: Biologically inspired cognitive architectures. *Neurocomputing*.

Gold, J. I, and M. N Shadlen. 2007. The neural basis of decision making.

Gottlieb, J. 2007. From thought to action: the parietal cortex as a bridge between perception, action, and cognition. *Neuron* 53 (1): 9–16.

Gray, C. M, and W. Singer. 1989. Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences of the United States of America* 86 (5): 1698.

Haggard, Patrick. 2008. Human volition: towards a neuroscience of will. *Nat Rev Neurosci* 9 (12): 934-946. doi:10.1038/nrn2497.

Halgren, E., C. Boujon, J. Clarke, C. Wang, and P. Chauvel. 2002. Rapid distributed fronto-parieto-occipital processing stages during working memory in humans. *Cerebral Cortex* 12 (7): 710.

Hawkins, J., and S. Blakeslee. 2005. *On intelligence*. Owl Books.

Healy, Alice F., Robert W. Proctor, and Irving B. Weiner. 2003. *Handbook of Psychology, Experimental Psychology*. Wiley, January 3.

Heeger, David J., Alex C. Huk, Wilson S. Geisler, and Duane G. Albrecht. 2000. Spikes versus BOLD: what does neuroimaging tell us about neuronal activity? *Nat Neurosci* 3 (7) (July): 631-633. doi:10.1038/76572.

Hillstrom, A. P, K. L Shapiro, and C. Spence. 2002. Attentional limitations in processing sequentially presented vibrotactile targets. *Perception & psychophysics* 64 (7): 1068.

Hofstadter, D. R, and M. Mitchell. 1994. The Copycat Project: A model of mental fluidity and analogy-making. *Advances in connectionist and neural computation theory* 2: 31–112.

Huang, Xin, and Michael A. Paradiso. 2008. V1 Response Timing and Surface Filling-In. *J Neurophysiol* 100 (1) (July 1): 539-547. doi:10.1152/jn.00997.2007.

Jensen, A. R, and E. Munro. 1979. Reaction time, movement time, and intelligence. *Intelligence* 3 (2): 121–126.

Jensen, O., J. Kaiser, and J. P Lachaux. 2007. Human gamma-frequency oscillations associated with attention and memory. *Trends in neurosciences* 30 (7): 317–324.

Kirchner, Holle, Emmanuel J. Barbeau, Simon J. Thorpe, Jean Regis, and Catherine Liegeois-Chauvel. 2009. Ultra-Rapid Sensory Responses in the Human Frontal Eye Field Region. *J. Neurosci.* 29 (23) (June 10): 7599-7606. doi:10.1523/JNEUROSCI.1233-09.2009.

Koivisto, M., P. Kainulainen, and A. Revonsuo. 2009. The relationship between awareness and attention: Evidence from ERP responses. *Neuropsychologia* 47 (13): 2891–2899.

Kosinski, B., and J. Cummings. 1999. The Scientific Method: An Introduction Using Reaction Time. *Tested studies for laboratory teaching (ABLE Proceedings)* 25: 219-234.

Laird, John E., A. Newell, and Paul S.P.S. Rosenbloom. 1987. SOAR: An Architecture for General Intelligence. *Artificial Intelligence* 33: 1-64.

Lakatos, P., C. M Chen, M. N O'Connell, A. Mills, and C. E Schroeder. 2007. Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53 (2): 279–292.

7. References

Langley, P., and D. Choi. 2006. A unified cognitive architecture for physical agents. In *Proceedings of the National Conference on Artificial Intelligence*, 21:1469.

Lavie, N., and S. Cox. 1997. On the efficiency of visual selective attention: Efficient visual search leads to inefficient distractor rejection. *Psychological Science* 8 (5): 395.

Libet, B. 1985. Unconscious cerebral initiative and the role of conscious will in the initiation of action. *Behavioral and Brain Sciences* 8: 529–566.

Lisman, J E, and M A Idiart. 1995. Storage of 7 +/- 2 short-term memories in oscillatory subcycles. *Science (New York, N.Y.)* 267 (5203) (March 10): 1512-1515.

Liu, H., Y. Agam, J. R Madsen, and G. Kreiman. 2009. Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron* 62 (2): 281–290.

Madl, T., B. J Baars, and S. Franklin. 2011. The Timing of the Cognitive Cycle. *PLoS ONE* 6 (4): e14803.

Maes, Pattie. 1989. How To Do the Right Thing. *CONNECTION SCIENCE JOURNAL* 1: 291--323.

Marinkovic, K., R. P Dhond, A. M Dale, M. Glessner, V. Carr, and E. Halgren. 2003. Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron* 38 (3): 487–497.

Martens, S., and B. Wyble. 2010. The attentional blink: past, present, and future of a blind spot in perceptual awareness. *Neuroscience & Biobehavioral Reviews* 34 (6): 947–957.

Martin, E. W., and K. L. Shapiro. 2008. Does failure to mask T1 cause lag-1 sparing in the attentional blink? *Perception & Psychophysics* 70 (3) (April): 562-570. doi:10.3758/PP.70.3.562.

Martinez-Conde, Susana, S. Macknik, Maria M. Martinez, Jose-Manuel Alonso, and Peter U. Tse. 2006. *Visual Perception Part 1, Volume 154: Fundamentals of Vision: Low and Mid-Level Processes in Perception.* 1st ed. Elsevier Science, November 27.

Mathewson, K. E, G. Gratton, M. Fabiani, D. M Beck, and T. Ro. 2009. To see or not to see: prestimulus ${$alpha$}$ phase predicts visual awareness. *Journal of Neuroscience* 29 (9): 2725.

McCall, R., J. Snaider, and S. Franklin. 2010. Sensory and Perceptual Scene Representation. *Journal of Cognitive Systems Research*.

Melloni, L., C. Molina, M. Pena, D. Torres, W. Singer, and E. Rodriguez. 2007. Synchronization of neural activity across cortical areas correlates with conscious perception. *Journal of Neuroscience* 27 (11): 2858.

Mikolajczyk, K., and C. Schmid. 2005. A performance evaluation of local descriptors. *IEEE transactions on pattern analysis and machine intelligence*: 1615–1630.

Minsky, M. 1988. *The society of mind.* Simon and Schuster.

Molholm, S., W. Ritter, M. M Murray, D. C Javitt, C. E Schroeder, and J. J Foxe. 2002. Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive Brain Research* 14 (1): 115–128.

Mueller, S. T, M. Jones, B. S Minnery, and J. M.H Hiland. 2007. The BICA cognitive decathlon: A test suite for biologically-inspired cognitive agents. In *Proceedings of Behavior Representation in Modeling and Simulation Conference, Norfolk.*

Näätänen, R., and I. Winkler. 1999. The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin* 125 (6): 826–859.

Nachev, P., G. Rees, A. Parton, C. Kennard, and M. Husain. 2005. Volition and conflict in human medial frontal cortex. *Current Biology* 15 (2): 122–128.

Negatu, Aregahegn, Stan Franklin, and L McCauley. A non-routine problem solving mechanism for a general cognitive agent architecture. In *Problem Solving: Techniques, Steps, and Processes.* Hauppauge, NY: Nova Science Publishers.

Nestor, A., and B. Kokinov. 2004. Towards active vision in the dual cognitive architecture. *INFORMATIONTHEORIES*: 9.

Nieuwenhuis, S., and M. Jepma. 2011. Investigating the role of the noradrenergic system in human cognition. *Decision Making, Affect, and Learning: Attention and Performance XXIII* 23: 367.

Nieuwenhuis, S., M. S Gilzenrat, B. D Holmes, and J. D Cohen. 2005. The role of the locus coeruleus in mediating the attentional blink: A neurocomputational theory. *Journal of Experimental Psychology: General* 134 (3): 291.

7. References

Nieuwenstein, M. R, M. M Chun, R. H.J van der Lubbe, and I. T.C Hooge. 2005. Delayed attentional engagement in the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance* 31 (6): 1463.

Nieuwenstein, M. R, M. C Potter, and J. Theeuwes. 2009. Unmasking the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance* 35 (1): 159.

Noudoost, Behrad, Mindy H Chang, Nicholas A Steinmetz, and Tirin Moore. 2010. Top-down control of visual attention. *Current Opinion in Neurobiology* 20 (2) (April): 183-190. doi:10.1016/j.conb.2010.02.003.

Nunez, Paul L., and Ramesh Srinivasan. 2005. *Electric Fields of the Brain: The Neurophysics of EEG*. 2nd ed. Oxford University Press, USA, December 1.

O'Reilly, R. C, T. S Braver, and J. D Cohen. 1999. A biologically based computational model of working memory. *Models of working memory: Mechanisms of active maintenance and executive control*: 375–411.

Ogawa, Seiji, and Yul-Wan Sung. 2007. Functional magnetic resonance imaging. *Scholarpedia* 2 (10): 3105. doi:10.4249/scholarpedia.3105.

Olivers, C. N.L, and M. Meeter. 2008. A boost and bounce theory of temporal attention. *Psychological Review* 115 (4): 836.

Olivers, C. N.L, and S. Nieuwenhuis. 2005. The beneficial effect of concurrent task-irrelevant mental activity on temporal attention. *Psychological Science* 16 (4): 265.

Olivers, C. N.L, S. van der Stigchel, and J. Hulleman. 2007. Spreading the sparing: Against a limited-capacity account of the attentional blink. *Psychological Research* 71 (2): 126–139.

Opris, I., A. Barborica, and V. P Ferrera. 2005. Microstimulation of the dorsolateral prefrontal cortex biases saccade target selection. *Journal of cognitive neuroscience* 17 (6): 893–904.

Osipova, D., A. Takashima, R. Oostenveld, G. Fernandez, E. Maris, and O. Jensen. 2006. Theta and gamma oscillations predict encoding and retrieval of declarative memory. *Journal of Neuroscience* 26 (28): 7523.

Palva, J. Matias, Satu Palva, and Kai Kaila. 2005. Phase Synchrony among Neuronal Oscillations in the Human Cortex. *J. Neurosci.* 25 (15) (April 13): 3962-3972. doi:10.1523/JNEUROSCI.4250-04.2005.

Pesaran, B., M. J Nelson, and R. A Andersen. 2008. Free choice activates a decision circuit between frontal and parietal cortex. *Nature* 453 (7193): 406–409.

Philiastides, Marios G., Roger Ratcliff, and Paul Sajda. 2006. Neural Representation of Task Difficulty and Decision Making during Perceptual Categorization: A Timing Diagram. *J. Neurosci.* 26 (35) (August 30): 8965-8975. doi:10.1523/JNEUROSCI.1655-06.2006.

Potter, M. C, M. Nieuwenstein, and N. Strohminger. 2008. Whole report versus partial report in RSVP sentences. *Journal of memory and language* 58 (4): 907–915.

Potter, M. C, B. Wyble, R. Pandav, and J. Olejarczyk. 2010. Picture detection in rapid serial visual presentation: Features or identity? *Journal of Experimental Psychology: Human Perception and Performance* 36 (6): 1486.

Raffone, Antonino, and Narayanan Srinivasan. 2009. An adaptive workspace hypothesis about the neural correlates of consciousness: insights from neuroscience and meditation studies. In *Attention*, Volume 176:161-180. Elsevier. http://www.sciencedirect.com/science/article/B7CV6-4X53TTW-J/2/30eac3cf080807d1576ffb5a462eab04.

Ramamurthy, U., and S. Franklin. 2011. Self System in a Model of Cognition. *Proceedings of Machine Consciousness Symposium at the Artificial Intelligence and Simulation of Behavior Convention (AISB'11)*: 51-54.

Rodriguez, E., N. George, J. P Lachaux, J. Martinerie, B. Renault, and F. J Varela. 1999. Perception's shadow: long-distance synchronization of human brain activity. *Nature* 397 (6718): 430–433.

Sage, A. P, and W. B Rouse. 2009. *Handbook of systems engineering and management*. Wiley-Interscience.

Sauseng, Paul, and Wolfgang Klimesch. 2008. What does phase information of oscillatory brain activity tell us about cognitive processes? *Neuroscience & Biobehavioral Reviews* 32 (5) (July): 1001-1013. doi:10.1016/j.neubiorev.2008.03.014.

Schack, B., N. Vath, H. Petsche, H. -G. Geissler, and E. Möller. 2002. Phase-coupling of theta-gamma EEG rhythms during short-term memory processing. *International Journal of Psychophysiology* 44 (2) (May): 143-163. doi:10.1016/S0167-8760(01)00199-4.

7. References

Schendan, Haline E., and Lisa C. Lucia. 2010. Object-sensitive activity reflects earlier perceptual and later cognitive processing of visual objects between 95 and 500ms. *Brain Research* 1329 (May): 124-141. doi:10.1016/j.brainres.2010.01.062.

Searle, J. R. 1995. The problem of consciousness. *Thinking and literacy: The mind at work*: 21–30.

Sergent, C., S. Baillet, and S. Dehaene. 2005. Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience* 8 (10): 1391–1400.

Serre, Thomas, Aude Oliva, and Tomaso Poggio. 2007. A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences* 104 (15) (April 10): 6424 -6429. doi:10.1073/pnas.0700622104.

Shadlen, M. N., R. Kiani, T. D. Hanks, and A. K. Churchland. 2008. Neurobiology of decision making. An intentional framework. In *Strungmann forum report: better than conscious*, 71–101.

Shapiro, K. L, J. E. Raymond, and K. M. Arnell. 1997. The attentional blink. *Trends in Cognitive Sciences* 1 (8): 291–296.

Shibasaki, Hiroshi. 2008. Human brain mapping: Hemodynamic response and electrophysiology. *Clinical Neurophysiology* 119 (4) (April): 731-743. doi:10.1016/j.clinph.2007.10.026.

Shih, S. I. 2008. The attention cascade model and attentional blink. *Cognitive Psychology* 56 (3): 210–236.

Shih, S. I, and A. Reeves. 2007. Attentional capture in rapid serial visual presentation. *Spatial Vision* 20 (4): 301–315.

Siegel, Markus, Melissa R. Warden, and Earl K. Miller. 2009. Phase-dependent neuronal coding of objects in short-term memory. *Proceedings of the National Academy of Sciences* 106 (50) (December 15): 21341 -21346. doi:10.1073/pnas.0908193106.

Sigman, M., and S. Dehaene. 2006. Dynamics of the central bottleneck: Dual-task and task uncertainty. *PLoS Biol* 4 (7): e220.

Sloman, A. 1999. What sort of architecture is required for a human-like agent. *Foundations of Rational Agency*: 35–52.

Sloman, A. 2008. Varieties of Meta-cognition in Natural and Artificial Systems. In *Workshop on Metareasoning, AAAI*. Vol. 8.

Soon, C. S, M. Brass, H. J Heinze, and J. D Haynes. 2008. Unconscious determinants of free decisions in the human brain. *Nature neuroscience* 11 (5): 543–545.

Stewart, L., A. Ellison, V. Walsh, and A. Cowey. 2001. The role of transcranial magnetic stimulation (TMS) in studies of vision, attention and cognition. *Acta Psychologica* 107 (1-3): 275–291.

Strain, Steve, Stan Franklin, and Patrick Jeuniaux. 2010. The LIDA model's hypotheses on the cognitive cycle, high-level cognitive processes, and brain rhythms. about:blank.

Stroud, John M. 1967. The Fine Structure of Psychological Time. *Annals of the New York Academy of Sciences* 138 (2 Interdiscipli) (February): 623-631. doi:10.1111/j.1749-6632.1967.tb55012.x.

Sun, R., and X. Zhang. 2004. Top-down versus bottom-up learning in cognitive skill acquisition. *Cognitive Systems Research* 5 (1): 63–89.

Sun, Ron. 2007. The importance of cognitive architectures: an analysis based on CLARION. *Journal of Experimental & Theoretical Artificial Intelligence* 19 (June): 159–193. doi:10.1080/09528130701191560.

Sun, Ron, and Stan Franklin. 2007. *Computational models of consciousness: A taxonomy and some examples*. Cambridge University Press.

Swadlow, Harvey A., and Stephen G. Waxman. 2010. Axonal Conduction Delays. http://www.scholarpedia.org/article/Axonal_conduction_delays.

Taatgen, N. A, I. Juvina, M. Schipper, J. P Borst, and S. Martens. 2009. Too much control can hurt: A threaded cognition model of the attentional blink. *Cognitive Psychology* 59 (1): 1–29.

Tallon-Baudry, C., O. Bertrand, M. A Henaff, J. Isnard, and C. Fischer. 2005. Attention modulates gamma-band oscillations differently in the human lateral occipital cortex and fusiform gyrus. *Cerebral Cortex* 15 (5): 654.

Taylor, P. C.J, A. C Nobre, and M. F.S Rushworth. 2007. Subsecond changes in top down control exerted by human medial frontal cortex during conflict and action selection: a combined transcranial magnetic stimulation electroencephalography study. *Journal of Neuroscience* 27 (42): 11343.

7. References

Thompson, E., and F. J Varela. 2001. Radical embodiment: neural dynamics and consciousness. *Trends in cognitive sciences* 5 (10): 418–425.

Thorpe, S., D. Fize, and C. Marlot. 1996. Speed of processing in the human visual system. *nature* 381 (6582): 520–522.

Thorpe, SJ, and M Fabre-Thorpe. 2003. Fast Visual Processing. In *The handbook of brain theory and neural networks*, 441-444. MIT Press.

Tiitinen, H. T., J. Sinkkonen, K. Reinikainen, K. Alho, J. Lavikainen, and R. Naatanen. 1993. Selective attention enhances the auditory 40-Hz transient response in humans. *Nature* 364 (6432) (July 1): 59-60. doi:10.1038/364059a0.

Tononi, G. 2004. An information integration theory of consciousness. *BMC neuroscience* 5 (1): 42.

Tononi, G., G. M Edelman, and O. Sporns. 1998. Complexity and coherency: integrating information in the brain. *Trends in cognitive sciences* 2 (12): 474–484.

Turing, A. M. 1950. Computing Machinery and Intelligence. *Mind* LIX (236): 433-460. doi:10.1093/mind/LIX.236.433.

Uchida, N., A. Kepecs, and Z. F Mainen. 2006. Seeing at a glance, smelling in a whiff: rapid forms of perceptual decision making. *Nature Reviews Neuroscience* 7 (6): 485–491.

VanRullen, R., and C. Koch. 2003. Is perception discrete or continuous? *Trends in Cognitive Sciences* 7 (5): 207–213.

VanRullen, R., and S. J Thorpe. 2001. The Time Course of Visual Processing: From Early Perception to Decision-Making. *Journal of Cognitive Neuroscience* 13 (4): 454–461.

Varela, F. J, E. Thompson, and E. Rosch. 1992. *The embodied mind: Cognitive science and human experience*. The MIT Press.

Varela, F., J. P Lachaux, E. Rodriguez, and J. Martinerie. 2001. The brainweb: phase synchronization and large-scale integration. *Nature Reviews Neuroscience* 2 (4): 229–239.

Verbruggen, F., A. R Aron, M. A Stevens, and C. D Chambers. 2010. Theta burst stimulation dissociates attention and action updating in human inferior frontal cortex. *Proceedings of the National Academy of Sciences*.

Vernon, David, Claes von Hofsten, and Luciano Fadiga. 2010. *A Roadmap for Cognitive Development in Humanoid Robots*. 1st ed. Springer, Berlin, November.

Walsh, V., A. Ellison, L. Battelli, and A. Cowey. 1998. Task-specific impairments and enhancements induced by magnetic stimulation of human visual area V5. *Proceedings of the Royal Society B: Biological Sciences* 265 (1395): 537.

Welford, A. T. 1980. *Reaction times*. Academic Pr.

Westendorff, Stephanie, Christian Klaes, and Alexander Gail. 2010. The Cortical Timeline for Deciding on Reach Motor Goals. *J. Neurosci.* 30 (15) (April 14): 5426-5436. doi:10.1523/JNEUROSCI.4628-09.2010.

Willis, J., and A. Todorov. 2006. First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science* 17 (7): 592–598.

Wyart, V., and C. Tallon-Baudry. 2008. Neural dissociation between visual awareness and spatial attention. *Journal of Neuroscience* 28 (10): 2667.

Wyble, B., H. Bowman, and M. Nieuwenstein. 2009. The attentional blink provides episodic distinctiveness: Sparing at a cost. *Journal of Experimental Psychology: Human Perception and Performance* 35 (3): 787.

Zeki, Semir. 2000. *Inner Vision: An Exploration of Art and the Brain*. Oxford University Press, USA, February 17.

van der Zwaag, Wietske, Susan Francis, Kay Head, Andrew Peters, Penny Gowland, Peter Morris, and Richard Bowtell. 2009. fMRI at 1.5, 3 and 7 T: Characterising BOLD signal changes. *NeuroImage* 47 (4) (October): 1425-1434. doi:10.1016/j.neuroimage.2009.05.015.

8. APPENDIX A – GERMAN ABSTRACT / DEUTSCHE ZUSAMMENFASSUNG

Anpassung und Verifizierung einer psychologisch plausiblen kognitiven Architektur

unter Verwendung von LIDA-basierten kognitiven Software-Agenten

LIDA (Learning Intelligent Distribution Agent) ist eine biologisch inspirierte, partiell Computer-implementierte kognitive Architektur mit dem Ziel, menschliche Kognition zu modellieren und zu erklären (Baars & Franklin, 2009). Um einige Vorhersagen von LIDA zu verifizieren, wurden drei auf LIDA basierende Software-Agenten entwickelt, die psychologische Experimente ausführen; ihre interne Parameter wurden angepasst und ihr Verhalten mit dem menschlicher Versuchspersonen verglichen.

Ein visuelles Objekterkennungssystem, das auf dem Erkennen invarianter Merkmale basiert, und ein neuartiger Aufmerksamkeitsmechanismus, der auf dem Locus Coeruleus / Noradrenalin System basiert, wurden zusätzlich entwickelt, um das LIDA-Modell weiterzuführen und eine größere Bandbreite an Experimenten modellieren zu können.

Da LIDA neurowissenschaftliche Prozesse auf hoher Ebene modelliert, war es möglich, LIDA-Module mit neuralen Korrelaten in Beziehung zu setzen. Neurowissenschaftliche Ergebnisse mit hoher zeitlicher Auflösung, verbunden mit dieser Zuordnung zu neuralen Korrelaten, wurden benützt, um die internen Parameter des computergestützten LIDA-Systems anzupassen. Die Software-Agenten, die auf diesem angepassten Modell der Kognition basieren, verhalten sich ähnlich wie menschliche Versuchspersonen in psychologischen Experimenten und liefern vergleichbare Daten. Dabei wurden die Parameter ausschließlich auf Grund von neurowissenschaftlichen Daten bestimmt, ohne das experimentelle Paradigma zu verändern. Dies erhöht die Plausibilität des LIDA-Modells und seiner zugrundeliegenden Hypothesen.

Drei kognitive Software-Agenten, basierend auf dem partiell Computer-implementierten LIDA-System, wurden entwickelt, um LIDAs Hypothesen zu stärken und die Plausibilität zu erhöhen, dass LIDA menschliche Kognition modellieren kann.

Der erste Software-Agent führt einfache Reaktionszeitexperimente aus. Das Hauptziel ist, die Parameter verschiedener Zeitabläufe, die aus neurowissenschaftlichen Daten gewonnen wurden, zu verifizieren. Der Agent operiert in einer simplen Umgebung, die aus einem Lichtreiz (rot oder grün) und einem Schalter (der gedrückt wird, wenn das Licht grün wird) besteht.

Der zweite Agent führt ein psychologisches Experiment durch, das zuerst von Allport (1968) vorgeschlagen wurde. Die Intention dieses Experiments ist es, LIDA's Hypothese zu stärken, dass Bewusstsein diskret (unstetig) ist. Im Besonderen erlaubt dieses Experiment den Vergleich der 'Discrete Moment Hypothesis' – die besagt, dass Bewusstsein aus distinkten und nicht überlappenden bewussten 'Momenten' besteht, in denen keine Zeit- und Ordnungsinformationen vorhanden sind– und der 'Continuous Moment Hypothesis', die bewusste 'Momente' als zusammenhängende und sich in Bewegung befindende Segmente von eintreffender sensorischer Information sieht.

In dem originalen psychologischen Experiment erscheint eine horizontale Linie in einer von 12 Positionen auf einem Bildschirm, auf den die Versuchsperson blickt. Die Linie ändert ihre Position sehr schnell und bewegt sich nach oben. Wenn der obere Bildschirmrand erreicht ist, beginnt die Linie wieder in der untersten Position. Abhängig von der Zyklusdauer (also der Geschwindigkeit, in der sich die Linie bewegt), nehmen die Versuchspersonen bewusst entweder eine einzige oder mehrere Linien wahr; unter einer gewissen Zyklusdauer sogar alle Linien auf einmal. In diesem Fall kann keine Bewegung mehr wahrgenommen werden. Dieses Experiment wurde hier verwendet, um zu zeigen, dass LIDAs Modell von diskreten Bewusstseinsabschnitten plausibel ist und dass die zeitlichen Beschränkungen eines bewussten 'Fensters' der Erfahrung eines LIDA-Software-Agenten jenen von menschlichen Versuchspersonen gleichen.

Der dritte Softwareagent modelliert das Aufmerksamkeitsblinzeln in einem Rapid Serial Visual Presentation (RSVP) – Paradigma, um LIDA's Hypothesen über Aufmerksamkeit zu stärken und außerdem um LIDA in einem komplexeren Experiment, das auch Bilderkennung involviert, zu testen. In diesem Paradigma werden dem Agenten 8 davor nicht gesehen Bilder in schneller Folge präsentiert (für deren Kategorie die Bilderkennung jedoch im Voraus trainiert wurde), in denen er zwei zielrelevante Bilder durch das Drücken eines virtuellen Knopfes identifizieren muss. Bei solchen Experimenten sind menschliche Versuchspersonen des Öfteren unfähig, das zweite zielrelevante Bild bewusst wahrzunehmen, wenn es sehr schnell nach dem ersten präsentiert wird – dieser Effekt wird Aufmerksamkeitsblinzeln genannt. Der Agent, der dieses Experiment reproduziert, tut dies mit einer Erkennungsrate, die mit menschlichen Ergebnissen vergleichbar ist; und ist fähig, im Vordergrund stehende Objekte in Bildern zu erkennen. Der Grund für das Nicht-Bewusstwerden des zweiten zielrelevanten Bildes bei schneller Präsentation ist hauptsächlich ein Aufmerksamkeits-Mechanismus, der auf dem Locus Coeruleus – Noradrenalin System im Gehirn beruht und von dem vermutet wird, bei der Verstärkung von relevanten Perzepten durch die Ausschüttung von Noradrenalin im Gehirn eine Rolle zu spielen; und der nach dem Verarbeiten eines relevanten Perzepts einer Refraktärperiode unterworfen ist. LIDA's existierendes Aufmerksamkeitsmodell wird durch ein Modell von diesem Mechanismus erweitert, damit menschliches Verhalten beim Aufmerksamkeitsblinzeln reproduziert werden kann.

Ein visueller Objekterkennungsalgorithmus wurde auch entwickelt um diese Aufgabe durchführen zu können, da LIDA zuvor keine implementierte visuelle Wahrnehmung hatte. Dieses visuelle System basiert auf invarienter Merkmalsextraktion und hat das Ziel, plausibel zu sein und gleichzeitig eine hohe Erkennungsrate zu erreichen. Nachdem prominente Merkmale im Bild von einem Extraktionsalgorithmus gefunden wurden, wird ein Merkmalsraum erstellt und die gefundenen Merkmale eingefügt. Eine große Anzahl von Detektoren, ähnlich den Merkmalsdetektoren im visuellen Cortex, überprüfen den Merkmalsraum und aktivieren entsprechende Wahrnehmungsrepräsentationen, falls genügend mit dieser Repräsentation verknüpfte Detektoren ihr Merkmal lokalisieren können. Anschließend kann diese Wahrnehmungsrepräsentation ins Bewusstsein eindringen und eine Aktion auslösen, falls die Aufmerksamkeit des Agenten auf sie gerichtet ist und falls sie eine bestimmte Aktivierungsschwelle überschreitet.

Alle drei beschriebenen kognitiven Softwareagenten verhalten sich vergleichbar zu menschlichen Versuchspersonen, und tun dies basierend auf demselben Model und mit denselben Parametern. Dies erhöht die Plausibilität des LIDA-Modells und stärkt seine Hypothesen. Außerdem zeigt es die vielseitige Einsetzbarkeit des Modells, und seine Erklärungskraft. Schließlich kann dieses plausible und detaillierte Computermodell menschlicher Kognition in Zukunft empirisch verifizierbare Hypothesen und Behauptungen hervorbringen und dadurch zu neuen Erkenntnissen in der Kognitionswissenschaft führen.

9. APPENDIX B – CURRICULUM VITAE

General information

Name:	Tamas Madl
Address:	Gärtnergasse 22
	2320 Kledering, Austria
Date of Birth:	27.03.1989
Citizenship:	Hungary
E-Mail:	tamas.madl@gmail.com

Research Experience

September 2010 – Computational modeling of human perceptual decision making
December 2010 using a LIDA-based (Learning Intelligent Distribution Agent)
(half-time) agent. In cooperation with Dr. Stan Franklin and the CCRG group at the University of Memphis. Supervisor: Dr. Igor Farkas

July 2010 – Computational cognitive architectures research at the University
September 2010 of Memphis (development of a LIDA-based agent, adjustment of
(full-time) parameters to neuroscientifically plausible values)
Supervisor: Dr. Stan Franklin. Related paper: *Madl, T., Baars,*
B.J., Franklin, S., (in press). The Timing of the Cognitive Cycle

April 2010 – Computational modeling of the evolution of cooperation in a
June 2010 prehistoric society. Supervisor: Dr. Karl Sigmund
(half-time)

Education

October 2009 MSc MEi:CogSci (Master's in Cognitive Science) at the
- June 2011 University of Vienna
(Most relevant modules: Computational Models of The Mind,
Machine Learning & Neural Computation, Neuroscience, Philosophy, Psychology, Models of Personality & Emotions, Statistics)

September 2008 – BSc (Hons) Computing at the University of Central Lancashire
May 2009 in Preston, UK – first class with honours
Most relevant modules: Advanced Software Engineering, Artificial Neural Networks, Object Oriented Methods, Data Warehousing
(Direct entrance into the 3rd year - the subjects at the Higher Technical Institute Spengergasse corresponding to the first two years at the University of Central Lancashire were recognised)

2003-2008: Higher Technical Institute Spengergasse, Electronic Data Processing Department, Vienna, Austria

January 2000-2003: Secondary school AHS Franklinstraße 26 in Vienna, Austria

1995-2000: Primary school Hriszto Botev in Veszprém, Hungary

Employment History

October 2009 -	Part-time job as a web-programmer for biz:Consult GmbH (PHP, June 2010 JS, AJAX, SQL, HTML, CSS)
September 2008 - April 2009	Part-time job as a web programmer for Skypark Secure (PHP, AJAX)
August 2008:	Internship at the OnTec-Company in Vienna (Java, JSP and JS)
Summer 2007:	Designing the SoftJobs website (www.softjobs.eu)
August 2006:	Internship at the company Qualysoft in Hungary Web project in C#, ASP.NET and Flash ActionScripts
August 2005:	Internship at Schrack Seconet AG in Vienna, Austria
July 2005:	Designing the HungaroImmo website (www.hungaroimmo.com)

Publications

Madl, T., Baars, B.J. & Franklin, S., 2011. The Timing of the Cognitive Cycle. PLoS ONE, 6(4), p.e14803.

Conferences Attended

June 2011	Middle European Interdisciplinary Cognitive Science Conference (Ljubljana, Slovenia): Conference Talk (*Tuning and Verifying a Psychologically Plausible Cognitive Architecture*)
June 2010	Middle European Interdisciplinary Cognitive Science Conference (Dubrovnik, Croatia): Poster presentation (*An Agent-Based Social Simulation of the Evolution of Cooperation*)
March 2010	Interdisciplinary College 2010 (Günne, Germany)

Academic Prizes & Awards

Scholarship of Achievement 2010 (University of Vienna)
2nd, 3rd and 5th place at the Austrian Physics Olympiad (2008, 2007, 2006)

Professional Qualifications

Programming skills:
- C# .NET
- C and C++
- Java
- SQL, PL/SQL
- Web programming: HTML, XHTML, JavaScript+AJAX, CSS, Flash ActionScripts, PHP, JSP

Software experience:
- Microsoft Visual Studio 2008
- Borland C++Builder 6
- Eclipse, NetBeans
- Oracle SQL Developer
- Adobe Photoshop and Flash
- Visual Paradigm for UML
- Microsoft Word, Excel, PowerPoint, Access; OpenOffice.org
- Microchip MPLAB IDE

Certificates:
- IELTS (Academic, 8.0 overall band score)
- Cisco ITEssentials II
- Cisco CCNA 1, 2, and 3

Languages

Native speaker in German and Hungarian, fluent in written and spoken English