Attention capabilities for AI systems

Helgi Páll Helgason[1], Kristinn R. Thórisson[1,2]
[1]*Center for Analysis & Design of Intelligent Agents / School of Computer Science, Venus 2nd floor, Reykjavik University,Menntavegur 1, 101 Reykjavik, Iceland*
[2]*Icelandic Institute for Intelligent Machines, 2. h. Uranus, Menntavegur 1, 101 Reykjavik, Iceland*
helgih09@ru.is, thorisson@ru.is

Keywords: Artificial intelligence, attention, resource management

Abstract: Much of present AI research is based on the assumption of computational systems with infinite resources, an assumption that is either explicitly stated or implicit in the work as researchers ignore the fact that most real-world tasks must be finished within certain time limits, and it is the role of intelligence to effectively deal with such limitations. Expecting AI systems to give equal treatment to every piece of data they encounter is not appropriate in most real-world cases; available resources are likely to be insufficient for keeping up with available data in even moderately complex environments. Even if sufficient resources are available, they might possibly be put to better use than blindly applying them to every possible piece of data. Finding inspiration for more intelligent resource management schemes is not hard, we need to look no further than ourselves. This paper explores what human attention has to offer in terms of ideas and concepts for implementing intelligent resource management and how the resulting principles can be extended to levels beyond human attention. We also discuss some ideas for the principles behind attention mechanisms for artificial (general) intelligences.

1 INTRODUCTION

The field of AI has a long history of targeting isolated, well-defined problems to demonstrate intelligent capabilities. While useful, many of these problems (and especially their task environments, as perceived by the system) are extremely simple compared to the problem of learning how to solve novel tasks and adapting to changes in real-world environments - a problem which must be addressed and solved in order for AI systems to approach human-level intelligence. Given the nature of this prior work, it is not surprising that limited focus has been given to real-time processing and resource management. However, the design of any AI system expected to learn and perform a range of tasks in everyday environments needs to face these realities:

- The real world is highly dynamic and complex and can provide an abundance of information at any given moment.
- Resources of any intelligent system are not only limited, but insufficient in light of the massive amount of information available from the environment.
- A range of time constraints, many of which are dictated by the environment, must be satisfied in order to ensure safe and successful operation of the system.

Much of existing work in the field of AI is also based on greatly simplified operating assumptions - a case in point being the practically impossible (but surprisingly common) assumption of infinite resources, often in terms of storage but particularly in terms of processing: A system based on this assumption will fail to perform and potentially crash in real world operation when fed with information at a greater rate than it is capable of processing. To find inspiration for implementing intelligent resource management we need not look far, nature has provided us with a prime example in human attention; a cognitive function that enables us to focus our limited resources selectively on information that is most important to us at any given moment as we perform various tasks while

remaining reactive to unexpected but important events in the environment. Consider that while reading this chapter you have effectively ignored more than 99.9% of the numerous things that your mind could have spent time and resources on doing. Perhaps not surprisingly, it turns out that this is exactly the kind of resource management that is required to enable AI systems to approach human-level intelligence in real-world environments. Thus, it makes perfect sense to investigate how AI systems can be endowed with this cognitive function for the purpose of improving their operation and making them applicable to more open-ended and complex tasks and environments. The goal need not be to replicate any biological function in detail, but rather to extract useful concepts and methods from the biological side while leaving undesirable limitations behind in order to facilitate the creation of AI systems that can successfully operate in real-world environments in realtime using limited resources.

While attention has been largely ignored in the field to-date, there are notable exceptions. These include cognitive architectures such as NARS (Wang, 1995), LIDA (Baars, 2009) and Clarion (Sun, 2006). However, the attentional functionality implemented in these systems is incomplete in various ways, such as focusing solely on data-filtering (ignoring control issues, e.g. how prioritization affects processing of selected data) and external environmental information (ignoring internal system states). The ASMO framework (Novianto, 2009) is somewhat unique as it assumes a tight coupling between attention and self-awareness and includes focus on internal states. However, none of this work addresses realtime processing, which is one of the major reasons we desire attentional functionality, in a vigorous fashion. Attention has also been studied in relation to AI within the limited scope of working memory (c.f. Phillips 2005 and Skubic 2004). While attention and working memory are closely related, this is a restrictive context to study attention within as working memory can in most cases be modelled as a cognitive function rather than an architectural component.

This paper starts with a brief overview of human attention and subsequently attempts to extract principles that may be useful for AI systems. This is followed by a discussion of how these principles might be extended to levels beyond human attention for meta-reasoning and introspection. We then present a high-level design of an attention mechanism intended for AI architectures.

2 HUMAN ATTENTION

Research of human attention has a long history dating back to the beginnings of psychology. Back in 1890, the American psychologist William James wrote the following (James 1890):

"*Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatterbrained state which in French is called distraction, and Zerstreutheit in German.*"

- William James

This elegant description indicates that the importance of attention for the human mind was identified as early as the 18th century. The beginning of modern attention research is commonly tied to Colin Cherry's work on what has been called the "cocktail party effect" (Cherry 1953), which addresses how we are able to focus on particular sensory data in the presence of distracting information and noise, such as following and participating in a conversation at a cocktail party in the presence of many other conversations and background noise, and still be able to catch when someone calls our name in the background. The ability to be in a focused state of attention while remaining reactive to unexpected events, seems to call for a selective filtering mechanism of some sort while at the same time requiring deliberate steering of cognitive resources. The cocktail party scenario is a good illustration of the dual nature of attention: We will refer to the deliberate, goal-driven side as *top-down* attention and the reactive, stimulus-driven side as *bottom-up* attention.

A number of models for attention were subsequently proposed, some of which were considered *early selection models* as selection of sensory information is assumed to occur early in the sensory pipeline based on primitive physical features of the information. This implies that the determination of what is important and should be selected is based on shallow, primitive processing with very limited or non-existent analysis of meaning. The Broadbent filter model (Broadbent 1958) is the most prominent of these. A number of *late selection models* have also been proposed, that

assume further analysis of incoming sensory information must be performed in order to determine its relevance and carry out efficient selection. The Deutsch-Norman (Norman 1969) model is based on the assumption that sensory information is not actually filtered, but processed to the point of activating representations stored in memory. Selection then occurs at the level of representations, where the most active ones are selected for further processing. The model also assumes an *attentional bottleneck* at this point, where only one representation can be selected for processing at a time. These two classes of attention models are referred to as the *early vs. late selection* models, and have resulted in some debate. Shortcomings of many early selection models are obvious, as they fail to account for parts of the cocktail party effect, especially phenomena such as noticing your own name being called from across the room while engaged in conversation. This contradicts the model, as the physical characteristics of the data (our name being called) would not be sufficient to attract our attention and pass through the filter; some analysis of meaning must be involved.

Some more recent theories and models of attention focus on the interaction between top-down and bottom-up attention. In (Knudsen 2007), an attention framework is presented based on four fundamental processes: working memory, top-down sensitivity control, competitive selection and bottom-up filtering for salient stimuli. The first three processes work in a recurrent loop to implement top-down control attention. Working memory is intimately linked to attention as its contents are determined by attention. This framework seems to capture most of the essential components of attention and is a promising candidate for inspiration with regards to attention for AI.

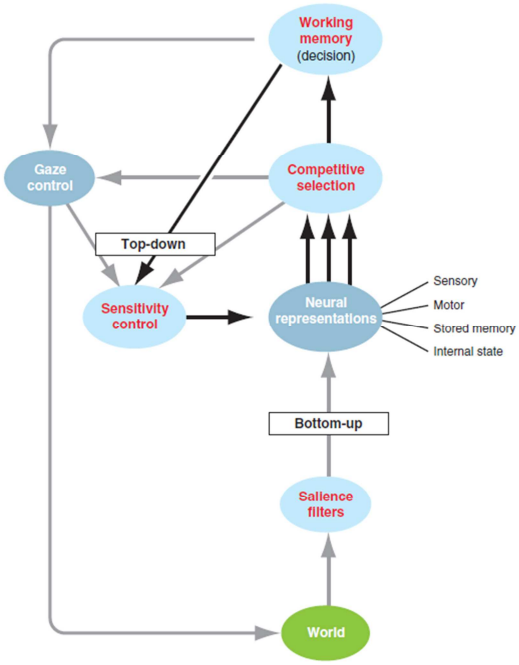

Figure 1: The Knudsen attention framework (from Knudsen 2007). Data flows up from the environment, passes through salience filters (which detect infrequent or important stimuli) and activates neural representations, which encode various types of knowledge. The activation of neural representations is also influenced by working memory via the sensitivity control of top-down attention that adjusts activation thresholds of individual representations. Representations compete for access to working memory with only the most active ones being admitted. Gaze is controlled by working memory and the selection process.

3 ATTENTION AND AI

Let us now consider how the previous chapter can inspire implementation of attentional capabilities for AI systems. As suggested in the introduction, we specifically target general AI systems designed to operate in complex environments under real-time constraints with limited resources. These systems are expected to perform various tasks while being reactive to events in the environment, a requirement that maps neatly to the top-down and bottom-up a workings of attention mentioned earlier. Both of these are necessary for a complete system; those that implement only top-down down attention will continue to work on tasks without being able to react to unexpected or novel events in the environment – events that may be relevant to the current task or necessary triggers for generation of new ones. Conversely, systems implementing only bottom-up attention cannot perform tasks beyond those that are simple and reactive; tasks consisting of multiple steps are not possible. However, when these two types of attention are properly combined, the result is a flexible system capable of performing complex tasks while being faced with interruptions and unexpected events. Part of the role of attention therefore, is to manage the balance between these two at every point in time.

The early vs. late selection debate mentioned in the previous chapter is also relevant here. It is possible to implement attention mechanisms for AI systems that perform selection early in the sensory

pipeline based on primitive features of the data. This approach is adopted in some of the best known existing cognitive architectures, such as SOAR (Laird 2008), where attention is viewed as a perceptual process rather than a cognitive one. Early selection unavoidably means that some data is (partly or fully) ignored without being processed for meaning; ignoring data that is not understood by the system introduces considerable risk as its relevance for the system is not known. This may be acceptable for narrow AI systems designed for specific tasks in specific environments as it may be possible to create shortcuts to understand the nature of incoming information in such cases. However, for general AI (AGI) systems designed for tasks and environments not specified at implementation time, this is highly problematic. Early stages of the sensory pipeline can contribute to attention in useful ways, such as performing biasing as opposed to absolute selection. For example, such biasing might be based on novelty or unexpectedness of the data as these properties may give rough clues to the importance of the information without requiring the information to be processed for meaning. Furthermore, this is a reasonable way to implement bottom-up attention, as suggested by the Knudsen model in Figure 1. As shallow processing at early stages of the sensory pipeline seems unlikely to provide a reliable measure of the importance of information, the late selection paradigm seems more promising than early selection in terms of AI and attention.

Top-down attention may be viewed as a goal-driven process as it is intimately related to current goals of the system. For goals to direct top-down attention, their level of specification is critical. In a system where goals are fully specified in terms of operation, the goal definition will be extremely useful in adjusting attention to elements that are relevant to the goal. A top-down attention mechanism based on pattern matching could generate partially specified patterns from goal specifications and attempt to find matches in sensory information. Predictions and expectations may also be expected to be necessary control input for top-down attention in systems that explicitly implement predictive capabilities – and there is good reason to believe that this is necessary in order to approach human-level intelligence. In terms of top-down attention, predictions may be treated in virtually the same fashion as goals (with level of specification being equally important as for goals).

4 AI ATTENTION: BEYOND THE HUMAN LEVEL

AI systems have an interesting advantage over human minds; they are based on software rather than hardware ("wetware"). While neurons of our brains can adaptively wire up to encode skills, knowledge and experiences the core mechanisms of these processes are fixed. For example, humans cannot easily acquire dramatically better ways of learning or remembering. This limitation does not apply to software AI systems; their potential for flexibility and reconfiguration are only limited by their architectural design. The same can be said for their level of introspection; our introspective capabilities are greatly limited - we only have a very vague sense of what is going on in our minds. On the other hand, there are much weaker limitations on self-observation in software AI systems, which again are limited only by architectural design.

A case for flexible architectures capable of autonomous self-reconfiguration is made in (Thórisson 2009). There are limitations on the complexity of manually built software systems and it is not unreasonable to assume that more complex software systems than exist today are needed in order to approach human-like AI. If our chances of manually building such systems are low, having the systems build themselves (in a sense) from experience is not an unreasonable line of research.

In order to perform deep levels of introspection in complex AI systems, attention is equally useful as for information originating outside the system; the sum of activity within such a system can be considered to be a vast stream of information and system resources remain limited. Determining which parts of this stream are worth processing in order to achieve meta-cognitive goals may be considered as the role of attention, in much the same way as attention operates on environmental information. The main purpose of introspection is to provide information to direct self-reconfiguration of the system. For example, an observation that system process *P* fails repeatedly in certain contexts can be used by the system to shut down process *P* and activate a different process (which may exist or need to be created/learned, generating a new meta-cognitive goal) when such contexts occur in the future.

5 AN ATTENTION MECHANISM FOR AI SYSTEMS

This section presents a design of one possible attention mechanism for AI systems which addresses the concepts related to attention discussed previously. The implementation and evaluation of this mechanism is upcoming future work. The approach taken adopts the theoretical and methodological framework presented in Thórisson (2009).

As attention is a ubiquitous cognitive process that cannot be easily separated from the rest of the cognitive architecture, some architectural requirements are unavoidable when tackling the design of an AI attention mechanism. The attention mechanism proposed here rests on the requirements that the underlying cognitive architecture has the following properties:

- **Data-driven**. All processing occurs in reaction to data. Processes are activated only when paired with compatible input data (fitting the input data specification of the process). Absence of fixed control loops allow for greater flexibility and operation on multiple time scales.
- **Fine-grained.** Processing and data elements of the architecture are numerous and small. Complex tasks require collaboration of many such elements. Reasoning about small, simple components and their effects on the system is more practical than attempting to do so for larger components.
- **Predictive capabilities.** Generate predictions with regards to expected events. Expectations are part of the control data of the attention mechanism.
- **Unified sensory pipeline.** Data from the environment and from within the system are treated equally. Enables systems to sense their own operation and potentially allows cognitive functions to be applied equally to task performance in the environment as well as meta-cognitive processing (e.g. self-reconfiguration).

The proposed attention mechanism implements both top-down and bottom-up attention. Top-down attention is based on goals and predictions, which serve as the basis for generation of so called *attentional templates* (AT), which are patterns that target data to various levels of specification. An AT can target general data (such as all data from a single modality, e.g. auditory) or more specific data such as anything directly related to an object or location in the environment and everything in between. As the architecture implements a *unified sensory pipeline*, sensory data and internal data are targeted in an identical fashion by attention. When a data object matches an active AT, it becomes a candidate to serve as input to a process for which it is compatible as input. Data objects that do not match any active AT are not caught by top-down attention and cannot trigger processing (unless caught by bottom-up attention). Each AT is created with an associated priority value, which is used when a match occurs with data, where the matching data item is assigned the same value. This value initially comes from the goal or prediction used to generate the AT. The assignment of priority values to data upon a match with an AT is called biasing. Available resources of the system are allocated to data items in order of their priority; data items with high priority values (greatest bias) will have better chances of receiving processing than those with lower values.

Bottom-up attention is implemented by primitive data selection principles that attempt to quantify the novelty and unexpectedness of input data based on content, temporal factors and operational experience. The novelty of data is based on how similar it is to data the system has previously seen, with higher novelty values being assigned to data that is different from previously seen data. Time also plays a role as data that has not been seen recently (but is not completely new to the system) will receive higher novelty values than those that have occurred recently. For example, if the environment has been silent for a while and sound is suddenly heard, auditory data is considered novel and would be caught by bottom-up attention. If the sound persists for some period of time, auditory data will cease to be novel and require top-down attention in order to be processed. In this way, the bottom-up part of the attention mechanism implements *habituation*.

Finally, a special mapping process is responsible for ensuring processes capable of consuming data caught by attention will be in active states. As the system is expected to contain numerous processes and data objects at any given time, attempting to match every data object to every process to determine if an operational match exists is not practically feasible. The data-to-process mapping component can be viewed as an optimization that reduces the number of data/process matched required.

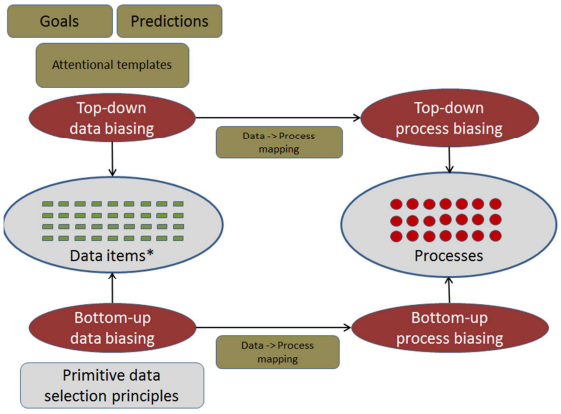

Figure 2: Overview of the proposed attention mechanism.

6 CONCLUSIONS

As has been shown, mapping models and concepts from attention in cognitive psychology to AI systems can be useful and straightforward. Surprisingly limited work has been performed on attention in the field of AI given that it is a field with the ultimate goal of creating human-like intelligence and that attention is clearly a critical cognitive process for humans. The fact that the human mind implements this kind of sophisticated resource management while being orders of magnitude more computationally powerful than existing computer hardware today also hints at the importance of attention for AI.

Furthermore, attention is likely to be equally critical for introspective systems such as those that can manage their own growth and adapt to experience at the architecture level. The internals of the system can be viewed dynamic and complex environment in the same way as the task environment. With a general and flexible attention mechanism, it may be possible to apply the same attention mechanism for both environments simultaneously; giving rise to AI systems that perform tasks *and* improve their own performance while being subject to real-time constraints and resource limitations.

ACKNOWLEDGEMENTS

This work was supported by the European Project HUMANOBS – Humanoids that Learn Socio-Communicative Skills Through Observation (grant number 231453).

REFERENCES

Baars, B. J., Franklin, S. 2009. Consciousness is computational: The LIDA model of Global Workspace Theory. International Journal of Machine Consciousness, 2009, 1(1): p. 23-32.

Broadbent, D. E. 1958. *Perception and Communication*. London: Pergamon.

Cherry, E. C. 1953. Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America*. Pages 975-979.

Knudsen, E. I. 2007. Fundamental components of attention. *Annu Rev Neurosci, volume 30*. Pages 57-78.

James, W. 1890. *The Principles of Psychology*. New York: Henry Holt, Vol.1, pages 403-404.

Norman, D. A. 1969. Memory while shadowing. *Quarterly Journal of Experimental Psychology,* vol. 21, pages 85-93.

Novianto, R., Williams, M.-A. 2009. The Role of Attention in Robot Self-Awareness, *The 18th International Symposium on Robot and Human Interactive Communication*. Pages 1047-1053.

Laird, J. E. 2008. Extending the SOAR cognitive architecture. In *Proceedings of the artificial general intelligence conference*. Memphis. TN: IOS Press.

Phillips, J. L. 2005. A biologically inspired working memory framework for robots. *Proc. 27th Ann. Conf. Congitive Science Society*. Pages 1750-1755.

Skubic, M., Noelle, D., Wilkes, M., Kawamura, K., Keller, J.M. 2004. A biologically inspired adaptive working memory for robots. *AAAI Fall Symp., Workshop on the Intersection of Cognitive Science and Robotics*. Washington D.C. 2004.

Sun, R. 2006. The CLARION cognitive architecture: Extending cognitive modelling to social simulation. In: Ron Sun (ed.), Cognition and Multi-Agent Interaction. Cambridge University Press, New York.

Thórisson, K. R. 2009. From Constructionist to Constructivist A.I. Keynote, Technical Report, FS-90-01, AAAI press, Menlo Park, California.

Wang, P. 1995. Non-Axiomatic Reasoning System: Exploring the Essence of Intelligence. Ph.D. diss., Dept. of Computer Science, Indiana Univ., CITY, Indiana.