Integrating perception, narrative, premonition and confabulatory continuation

Ladislau Bölöni
Dept. of EECS
University of Central Florida
4000 Central Florida Blvd, Orlando FL 32816

Abstract

Current state-of-the-art AI algorithms outperform humans on several well delimited tasks but have difficulty emulating general human behavior. One of the reasons for this is that human behavior, even in short scenarios, requires the integration of multiple cognitive mechanisms which are deployed simultaneously and are interacting with each other in complex and subtle ways. In this paper we show how a simple scenario of watching television requires at least four different cognitive mechanisms: perception, narrative, premonition and confabulatory continuation. We describe the general requirements of these mechanisms and outline the techniques through which the Xapagy cognitive architecture implements them.

Introduction

Recent years had seen significant progress in artificial intelligence algorithms. For instance, we have software which can beat humans at chess or Jeopardy, and in future years, a number of other specific domains will likely be conquered. To achieve the original, broader goals of AI however, it is not sufficient to create a series of narrow systems. Human behavior does not consist of moving from the chess table to a game of Jeopardy, and then to a system diagnosis task. Rather, in all the encountered situations, humans deploy a wide range of different cognitive behaviors. Memory, anticipation, emotions, goals, self interest, altruism and humor are applied simultaneously and interwoven in complex and subtle ways.

The field of cognitive architectures have spent the last 30 years working towards Alan Newell's vision of systems which integrate the whole (or at least large parts of) human cognition (Newell 1994). Certainly, researchers in the 1990s underestimated the difficulty of building a human-equivalent cognitive system. However, the fact that we now have AI algorithms which perform spectacularly well in a narrow domain, but fail trivially as soon as they leave it, validates the importance of the integrative vision of AI.

Let us consider how a relatively simple scenario of watching television requires the integrated deployment of several cognitive mechanisms. The main character Robby is either a human or an agent controlled by a cognitive architecture:

> *Robby has a previous experience in reading stories and watching movies involving duels between knights, warriors and gangsters. He is currently watching a TV program which is a dramatization of Homer's Iliad. On the screen he sees the fight between Hector and Achilles, while the voice-over narration comments on the story. Robby fears that the story will end in the death of one of the characters. Suddenly, the program is interrupted by a commercial. Frustrated, Robby tries to envision a way in which the story will end peacefully.*

We will argue that in this simple scenario Robbie deploys at least four different cognitive mechanisms:

- **Perception:** the processing of a real-time, data rich input from a set of sensors.
- **Narrative following:** the processing an input in the form of a high level narrative, roughly along the lines of a literary story.
- **Premonition:** the ability to judge that certain events are likely to occur in the immediate future.
- **Confabulatory continuation:** the ability to generate a fictional continuation of the ongoing series of events.

Of course, human behavior includes many other cognitive mechanisms: for instance, attention management and body control at the low level and problem solving, question answering and planning at the high level.

The objective of the paper is to illustrate a possible way in which these four cognitive mechanisms can be integrated in a common architecture. We start by a general, architecture independent description of the requirements of these four cognitive mechanisms. Then, we briefly introduce the Xapagy cognitive architecture and its Xapi language, and discuss the way in which these mechanisms can be implemented and integrated in that system.

Four cognitive mechanisms

Perception

Perception is the cognitive mechanism which integrates the real time input of the sensors into the cognitive model of the agent. In this paper we assume that the sensor and associated processing units had already converted the low level input (pixels and audio waveforms) into a *perception stream* of higher level, symbolic inputs (objects, movement, spatial relationships, utterances).

One of the most important attributes of perception is that it happens in **real time:** the stream follows the temporal succession of events happening in the real world. While we can allow for a delay before the events reach consciousness as shown in the Libet experiments (Libet et al. 1983), the perception would not include significant delays, nor present events out of order.

Another attribute of perception is its **specificity:** the perception stream is anchored in concrete, physical reality. The perception stream does not refer to abstractions, nor use high level verbs. We do not perceive a "dogness", but see a concrete dog. We do not hear "hammering", we only hear individual hammer strokes. We don't see "fighting": we only see two warriors performing specific movements. The specificity does not necessarily mean correctness. In some cases (optical illusions, phantom limb phenomena, hallucinations) it is possible that the low level processing makes mistakes in the segmentation of reality, or in assigning attributes to objects.

Finally, compared to the narrative, the perception stream is **dense**, contains a rich stream of details and **unfilterable**. The latter attribute requires some explanation. Robbie might choose to look at the television or to close his eyes. However, once he looks at a scene, it can not choose which attributes and relations it perceives. For instance, it can not look at Achilles and Hector without noticing which one is on the left side (spatial relationship), it can not look at a red object without seeing it is red, or observe a human without observing his hands, legs, head and clothes.

Following a narrative

Humans cannot speak (write) or listen (read) fast enough to convey the full richness of the perception stream. Human communication happens at the level of a narrative, which we will associate with a list of statements through which salient aspects of the story are extracted and presented in a higher level, summarized and abridged form.

In contrast to the synchronous nature of perception, a narrative can be **asynchronous:** there is no obvious time bind between stories and the physical reality. A narrative might refer to events in the past, predict events in the future, it can describe events faster or slower than the time span they might take in reality. Also a narrative might reverse the order of events.

A narrative is not bound to the truth as exists in the physical reality, and it can represent physical or even logical impossibilities. We can, for instance, say that Hector is simultaneously to the right and to the left of Achilles, a statement which can not appear in the perception stream.

Compared to the richness of perception, a narrative stream is **sparse**; it can omit details which are necessarily part of a direct perception. For instance, we can narrate the fight of Hector and Achilles without specifying which one is on the right or the left side. One way in which sparseness can be achieved is by simple omission of details: we do not need to mention the color of Hectors armor. Another way is to use expressions which summarize longer ranges of perception: we say that we hear hammering, instead of describing every beat, or we see Achilles and Hector fighting, instead of enumerating every blow.

Finally, the narrative can refer to abstractions which are not directly observable in the perception. We can talk about the patriotism of Hector or the love he feels for his wife.

Premonition

We defined premonition as the ability to judge that certain events are likely to occur in the immediate future. The existence of this cognitive mechanism in humans is attested by introspection but also by the fact that humans show external signs that they are affected by premonitions. A fan of horror movies, for instance, would visibly shift in his seat in the expectation of a scary scene, even if he cannot visualize the scary scene which will follow.

While premonitions are commonly associated with the subconscious, the fact the horror-movie fan acquires its expectations through experience, and it can project it to the line of the movie, shows that premonitions are learned and cognition-dependent (which does not exclude their physical manifestations).

Premonitions are **unspecific:** they don't offer details or even a consistent narrative of the expected events. Nevertheless, a human can easily spin a concrete story based on his premonitions (or even multiple alternative ones) - these narratives are usually based on first hand or indirect *experiences*. This act of spinning a a specific story can happen either at an external prompting or as an internal mechanism of the agent. We will argue, however, that this story spinning is confabulative continuation, a cognitive mechanism distinct from premonition. From the fact that humans can very quickly spin different stories about their premonitions, we might hypothesize that a premonition is a probability distribution over groups of different, but related outcomes.

Humans are able to consciously notice when a certain premonition was fulfilled, and they can also notice if a high-impact event occurs for which they had no premonition. On the other hand, they rarely notice that a premonition was not fulfilled. We conjecture that humans notice that a premonition was not fulfilled only if they explicitly reason about it, which involves other cognitive mechanisms beyond premonition.

We can hypothesize that the premonition mechanism has a beneficial effect on human behavior, by controlling the allocation of physical and mental resources and guiding high level behavior, for instance in avoiding surprises.

Confabulative continuation

Confabulative continuation (CC) is the ability of an agent to explicitly continue an ongoing narrative without necessarily verbalizing it. The story generated by the CC is not a deterministic one. The agent has a considerable **freedom to steer** the confabulated story from the "most likely continuation" towards directions desired by the agent. In our example, the agent can choose to generate a continuation which has a positive ending, one which is sentimental, humorous or gory. In fact, the agent can even divagate from the story which it purports to continue by introducing new locations, scenes and characters, even to the point of completely ignoring the entities of the original story. The agent must have means to control how far it can divagates from the original story.

The freedom to steer the story is limited by the requirement of at least a **partial internal consistency**. Even in a fully confabulated story, most events must be plausible successors of the existing ones. It is in fact very difficult for a human to generate a story entirely consisting of non sequiturs. Even in the most experimental literary stories, with the notable exception of dadaism, the relatively few logical impossibilities are separated by long stretches of internal consistency.

Finally, we must observe that the confabulated stories are **sparse**: they resemble a narrative stream rather than a sensed stream. For instance, it is feasible to continue a story without specifying any detail: "Hector and Achilles solved their differences, and became friends from than on". Somewhat more detailed continuations are of course, possible.

An introduction to the Xapagy architecture

Xapagy is a cognitive architecture developed with the goal of mimicking the human reasoning about stories. The interaction of the Xapagy agent with the outside world happens through Xapi, a language with a simplified syntax, which approximates closely the internal representational structures of Xapagy. Xapi uses an English vocabulary, and it should be readable for an English language reader with minimal familiarity of the internal structures of Xapagy. For instance the snippet:

> *Hector is a Trojan warrior. Achilles is a Greek warrior. Achilles has a shield. Achilles's shield is big. Achilles hits Hector. Hector strikes Achilles.*

can be translated into Xapi as follows:

```
A "Hector" / exists.
"Hector" / is-a / trojan warrior.
An "Achilles" / exists.
"Achilles" / is-a / greek warrior.
"Achilles" / owns / a shield.
The shield --of-- "Achilles" / is-a / big.
"Achilles" / hits / "Hector".
"Hector" / hits / "Achilles".
```

Xapi sentences can be in subject-verb-object, subject-verb or subject-verb-adjective form. A single more complex sentence exists, in the form of subject-communication verb-scene-quote, where the quote is an arbitrary sentence which is evaluated in a different scene. Subjects and objects are represented as *instances* which can acquire various attributes in form of *concepts*. Xapi sentences are mapped to internal Xapagy structures called *verb instances* (VIs). VIs can be locally connected with succession, summarization, context, coincidence and question-answer *links*.

One of the unexpected features of Xapagy instances is that an entity in colloquial speech is often represented with more than one instance in Xapagy. These instances are often connected with *identity relations* but participate independently in verb instances, shadows and headless shadows. We refer the reader to the technical report (Bölöni 2013) for a "cookbook" of translating English paragraphs of medium complexity into Xapi.

The newly created VIs of a story are first entered into the *focus* of the Xapagy agent, where they stay a time dependent of their salience, type and circumstances. For instance, VIs representing a relation will stay as long as the relation holds. On the other hand, VIs representing actions are pushed out by their successors. During their stay in the focus, VIs acquire salience in the *autobiographical memory*. After they leave the focus, VIs and instances will not change, and they can not be brought back into the focus. As we will see, newly encountered events or story snippets are *shadowed* by elements of the memory, while *headless shadows* project into the future, providing prediction or in the past, suggesting elements which might be missing in the current narrative.

Implementation and integration

Background knowledge

Reasoning in Xapagy is based on its autobiographical memory. There is a minimal core of hardwired semantic information, referring to existence / non-existence, grouping, exclusive and non-exclusive relations, temporal succession and identity. Everything else, even topology and basic spatial relations are acquired from autobiographical experience, and they are not extracted into rules.

The first challenge in creating a Xapagy agent with knowledge for a specific situation is the creation of a *synthetic autobiography* (SYAB). While ideally we would have a SYAB capturing the experience of an adult human, in practice current experimentation using Xapagy involves custom-built SYABs focusing on stories relevant to the domain of interest. For our example, we need two types of stories in the autobiography:

Perception SYAB: physical observations which have the nature of perception (*i.e.* dense and specific). They involve general observations about animate and inanimate objects, spatial arrangements, as well as observations about human beings: body parts, clothing, tool use, sounds, harm and dying. As hand-crafting long sequences of perception SYAB is impracticable due to the density of the stream, our ongoing work on the perception SYAB involves generating it from a virtual environment. The ideal future solution would be the acquisition of real world observations from a sensor attached to, for instance, a robot.

Story SYAB: consists of stories which we anticipate to be relevant to the story we investigate. These stories present a level of detail similar to the one in literary works. Thus, it is feasible to translate the English text manually to Xapi (automated English to Xapi translation is future work). We translated to Xapi a number of stories involving duels, such as Lancelot and Maleagant, David and Goliath, Macbeth and Macduff. We also added a description of the story of Patrocles and Hector, which is a preliminary scene in the Iliad sequence we consider, and normally had been heard by the agent before the Hector and Achilles fight.

Perception, narrative following and their integration

The Xapagy architecture is built around a *narrative bottleneck*: the perception as well as spoken or read narratives are represented through a common internal structure of a stream of VIs. The two streams arrive simultaneously and are merged together in the narrative bottleneck. Naturally, however, the perception stream and the narrative stream will have a different content. Figure 1 shows a snippet of the two streams before their merge.

The left side of the figure shows the perception stream, which contains a dense series of concrete, low level observations. In this case, the comma terminator after the statements show that they arrive at intervals of 0.1 seconds, in contrast to the period terminator used in the right which indicates intervals of 1 second. This means that we have about 10 perception VIs for every narrative VI. All the perception instances and VIs refer to a single scene labeled #Perception, which in our case is the battlefield shown on the television. This scene contains two physical animate objects labeled #A and #H. Note that the names of Achilles and Hector are not directly observable, thus they are not attributes of the instances[1]. Xapagy uses the same autobiographical memorization mechanism for all VIs irregardless of their provenance: in its simplest form, a VI accumulates memory salience throughout its stay in the focus. Due to the energy limits of the focus, the high arrival rate, the "pushout effect" through which incoming VIs steal energy from their predecessors, few of the perception VIs will reach sufficient salience to be stored in the autobiographical memory. The implication is that we are rarely remembering our ongoing perception. Nevertheless, a perception can be remembered if it has a very high impact (explosions, unexpected events) or if it is connected to an external or internal narrative.

A snippet of the narrative stream is shown in Figure 1-right. Narrative VIs, converted from spoken words, arrive at the speed at which speakers deliver them multiplied by the number of Xapi statements into which a specific English expression is translated. In general, complex grammatical expressions require a larger number of Xapi sentences (see (Bölöni 2013) for details).

The majority of the narrative stream takes place in the scene #Narrative and involves the Greek warrior Achilles and the Trojan warrior Hector. However, the narrative does not necessarily take place in a single scene. Various scenes, connected to each other with `view`, `fictional-future` or `succession` relations can be used to represent aspects of planning, daydreaming, flashbacks and so on. For instance, in our case, the narration specifies that "the shield of Achilles was made by the god Hephaestus". To express this sentence, the Xapi text creates a new scene #Past, reverse linked with a succession relationship to the #Narrative scene. In this scene, we have the god forging a shield. The shield made by the god is a different instance than the one currently held by Achilles, and they can have different properties - new and shiny for the one in scene #Past, scratched and dusty in the scene #Narrative. The fact that these two shield are considered identical in colloquial language[2], is marked by the identity relation between the two instances.

The narrative stream also differs from the perception stream by the use of concepts and verbs which are not directly perceivable. The attribute of `trojan` is not directly observable by perception. The verb `forging` (a shield) is a summarization verb, which allows us to summarize in a single VI what in the perception would be thousands of individual movements such as strikes of hammer.

The way it is shown in Figure 1 the perception stream and narrative stream are strictly separated. Indeed, it is possible for humans to listen to a narration which is completely unrelated to their current perception. In this case, however, in order to understand the narrated television program the agent needs to perform processing to *integrate* the streams.

The first step is to recognize that the scene of the narration is a *view* of the observed scene. In Xapagy, two scenes are in the view relation if they share the same core of events. Frequently, a view is an interpretation of a scene: for instance, when reading a book, the meaning extracted by "reading between the lines" can be represented in a separate scene, connected with a view relationship to the text of the book. Similarly, agents might use view scenes to model the "theory of the mind" of other agents they are interacting with. In our case, the agent must recognize that the scene #Narrative is a view of the scene #Perception. As view scenes interpret the same story as their target, the entities of the story are often replicated as instances both in the view and the viewed scene. In Xapagy, these are represented by separate instances, connected through identity relations. Thus, what we have discussed until now can be expressed in Xapi as follows:

```
#Narrative / is-view-of / #Perception.
"Hector"--in-- #Narrative / is-identical/
   object #H--in--#Perception.
"Achilles"--in--#Narrative /is-identical/
   object #A--in--#Perception.
The shield--of--"Achilles"--in--#Narrative/
   is-identical / #S--in--#Perception.
```

Once the relations between the scenes and instances had been set up, the agent will establish links between the incoming VIs. The succession links between VIs and the context links to the relations currently existing in the scenes are automatically set up, but they do not extend beyond the individual streams.

The integration of the streams is using two types of links which can cross scenes. *Coincidence* links connect VIs which refer to the same event, and they are interpreted as describing different aspects of the same single event. For instance, an event which is described by the perception stream as the raising of a hand, in the narrative stream it is interpreted as a greeting.

[1]Hashtag labels are a syntactic sugar used in Xapi to simplify instance references. They are not taken into account in reasoning.

[2]In the philosophy of personal identity, such a relation is called *somatic identity*.

```
$NewSceneCurrent #Perception
An animate object #H / exists,
An animate object #A / exists,
#H / is-left-of / #A,
#A / has / a hand #ALH,
#A / has / a hand #ARH,
#A / has / a leg #ALL,
#A / has / a leg #ARL,
#ARH / holds / a long shiny object.
#ALH / holds / an object #S,
#S / is-a / round,
#S / is-a / dusty,
#H / has / a hand #HLH,
#H / has / a hand #HRH,
```

```
$NewSceneCurrent #Narrative
A "Hector" trojan warrior / exists.
An "Achilles" green warrior / exists.




"Achilles" / holds / a famous shield.
$NewSceneCurrent #Past, shield->shield,
    god "Hephaestus"
Scene #Past / is-past-of / scene #Narrative
"Hephaestus" / forges / a shield.
The shield / is-identical /
    shield--of--"Achilles"--in--#Narrative.
```

Figure 1: Illustrative snippets of the perception stream (left) and the narrative stream (right). To save space, these examples understate the difference in density between the perception and narrative stream.

In contrast, *summarization* links connect a group of VIs to a summary VI which can represent them as a group. Summarizations are automatically created even in the perception stream for patterns such as repetition and alternation. More interestingly, VIs from the narrative can summarize a series of perceived VIs: for instance, the statement

```
"Hector" + "Achilles" /
   in-summary are-fighting.
```

can represent a long series of observed movements.

The shadowing mechanism

Shadowing is the main reasoning technique of the Xapagy architecture. Each instance and VI in the focus has an attached *shadow* consisting of a weighted set of instances, and respectively VIs from the autobiographical memory.

The maintenance of the shadows is done by a number of dynamic processes called *diffusion activities* (DAs)[3]. The diffusion activities resemble the spreading activation models deployed in many cognitive architectures and neural network models. For instance, a feed-forward neural network can be perceived to perform two steps of forward spread of its input activations. Once these two steps had been made, the output of the network is well defined and unchanged. The DAs, in contrast, are recursive, evolve continuously and are subject of complex temporal dynamics. Even if an equilibrium would be theoretically reachable in the long run, the periodic arrival of new VIs disturbs the equilibrium of the system. The closest relatives of DAs might be the continuous time recurrent neural network models (Beer 1997).

Some of the DAs add or strengthen shadows based on direct or indirect attribute matching, others decay shadows, while again others (such as the scene sharpening and the story consistency sharpening DAs) rearrange the weights between the shadows. The DAs for instance and VI shadows interact with each other: a shadow of a subject-verb type VI will make the subject instances shadow as well, even if they are not sharing any attributes.

[3]The Xapagy system also employs a number of single shot processes called *spike activities* (SAs).

The dynamic nature of the DAs make shadows constantly evolve in time. Shadows take time to develop, and different shadows might appear at different times in the process. In general, shadows based on immediate attribute match will appear first, but later they might be weakened by inconsistencies between the current and the shadow story. Less immediate matches appear later. The implication is that the shadowing process depends on the pace of the arrival of the VIs. If the VIs arrive very fast, only a very superficial matching is done (which will impact the premonition and confabulative continuation mechanisms). In contrast, if there is a very long delay between VIs, the shadowing mechanism exhausts the possibilities of bringing in new shadows from the autobiographical memory, and the sharpening-type DAs will gradually create a “winner”: i.e. the shadows will be dominated by a single story line from the past.

For the practical purposes of our running scenario, the ongoing fight between Hector and Achilles will be shadowed by duel fights from the memory of the agent. The alignment might not be perfect. For instance, the fight between Achilles and Hector will be shadowed by the fight between Hector and Patrocles. The questions, however, is how the instance shadows are set up? Is Hector shadowed by his own previous instance, or by Patrocles? In practice, both warriors will be present in the shadow. Initially the shadow of his own previous instance will be stronger (due to the larger number of shared attributes and the identity relation). Later, however, the story consistency DAs will force a stronger mapping between Hector and Patrocles (for being on the loosing side of the fight).

Premonition and its integration with perception and narrative

Let us consider the moment when we have added to the focus several VIs describing the fight between Hector and Achilles. The VIs are connected by succession links. These will be shadowed by VIs from the memory. On their turn, these VIs have succession links, some of them extending further into the future than the current story.

A DA in the Xapagy agent (a) finds the links from the shadow VIs, (b) interprets them in the terms of the current story through a process called *reverse shadowing*, (c) clusters them in weighted groups based on a similar interpretations and (d) calculates scores for the individual groups based on the weights and links between the shadows which generated them (their *supports*).

These weighted groups are very similar to shadows, except that the VI in the current story, the *head* of the shadow, is an interpretation which does not exist as an instantiated VI. Thus the Xapagy name for these structures is *continuation-type headless shadows (HLSs)*[4]. The continuation HLSs implement the premonition mechanism by modeling plausible future events. Multiple continuation HLSs can exist simultaneously. For instance, the agent might entertain both the possibility of Hector killing Achilles, Achilles killing Hector, or a friendly reconciliation.

When a new VI is inserted in the focus, it is first matched against the existing set of continuation HLSs. If it matches one of them, the HLS is automatically converted into a shadow of the new VI. This corresponds to a *validated premonition*, and it is a well defined event. HLSs, just like the shadows which give rise to them, evolve in time. If the story moves in a different direction, the HLS will decrease in strength, and ultimately will disappear, a low key event not normally recorded in memory. We find, therefore, that the premonition mechanism is ultimately implemented by a mechanism strongly tied to and integrated with the perception and narrative following mechanism. The continuation HLSs implementing the premonition can be taken into consideration by the resource allocation and planning components of the agent. Similarly, the asymmetry between the validation and lack of validation of a premonition matches the way in which humans remember more often the validation than the failure of their premonitions.

Confabulative continuation and its integration with other components

Confabulative continuation is implemented in Xapagy through a mechanism which instantiates internally one of the predictions of the continuation HLSs. This will create a new VI inserted in the focus. The shadowing mechanism will update the new shadows by reinforcing those shadows which are consistent with the instantiated continuation, and weakening those which are not. This will, in its turn, create a new collection of continuation HLSs, from which the agent will choose another one for instantiation, and so on. The resulting VIs form a new, confabulated story.

The most plausible continuation of the current story can be obtained by instantiating at each step the continuation HLS with the strongest support. The definition of plausibility, in this case, means that the continuation is in line with the experience of the agent as reflected in its autobiography. For instance, if the agent had seen mostly duels with a friendly outcome, it will also confabulate a friendly termination of the Hector - Achilles fight.

[4]Similar mechanisms are used for inferring missing actions, hidden relations, summarizations and answers to questions

The agent can steer the confabulated story by instantiating a HLS which is weaker, but more desirable due to other aspects - for instance, it might have a positive emotional valence. Performing such a selection for just one or several HLSs is sufficient to steer the story in a new direction. The updated shadows will generate HLSs compatible with the new direction of the story, and the agent can go back to instantiate the HLS with the most support. This way, the agent can achieve confabulative continuations which match the human behavior of partial internal consistency.

Related work

It is impossible to cover in the space available all the relevant work done by the cognitive architecture (CA) community with regards to integration of cognitive mechanisms: all CAs had to consider the problem of integration, in the context of the other design decisions of the given CA. Thus, we will only try here to position the assumptions made by CA in the general field of CAs.

One way to classify CAs is to arrange them on a continuum ranging from *tight-core CAs* which try to cover the range of human cognition using a small number (possibly only one) knowledge representation (KR) structure and reasoning algorithm, and *wide-core CAs* which integrate a large number of eclectic KR structures and reasoning types. The differences, to be sure, are not necessarily one of principle: they can be an impact of software engineering, human resource management and external constraints. Focusing on a tight core and investigating how wide range of human cognition it can cover might be the best use of a small research team[5]. In contrast, sometimes it is convenient to bring in well tested algorithms, even if it requires extending the core and adding adaptation layers.

Soar (Lehman et al. 1998) and ACT-R (Anderson and Lebiere 1998; Anderson et al. 2004), the cognitive architectures with the longest history of active development, had started out as tight-core systems, but had been gradually augmented with a large set of KR structures and reasoning architectures. In contrast, other cognitive architectures, such as Polyscheme (Cassimatis et al. 2004) had been designed from the beginning to integrate an eclectic mix of representation types and cognitive models. Some newer architectures are also tight-core: for instance Sigma (Rosenbloom 2011) uses factor graphs as a general purpose reasoning model. Similarly, Xapagy exhibits a common internal representation we call the *narrative bottleneck:* virtually all the internal representation takes a single form, that of a *story*, on which there is a single reasoning technique (shadows / headless shadows) and a single form of memory (autobiographical memory).

Naturally, the challenge of integrating cognitive mechanisms will take different forms in tight-core versus wide-core systems. For instance, in (Kurup et al. 2012) ACT-R is integrated with robotic perception using the Felzenswalb algorithm for detecting objects in images (in this case pedestrians). While the object detection algorithm is outside the ACT-R architecture, the actual processing of perception,

[5]Xapagy falls in this category

through the Perceptual (Visual and Aural) Modules, happens inside the architecture. The cognitive architecture is used to eliminate false positives, fill in the gaps and improve performance of perception by directing the attention of the object detecting algorithm to parts of the visual field where pedestrians are expected to appear. The key element in each of these functions is the prediction of the locations the system expects a pedestrian to be.

It is interesting to compare the approaches taken by ACT-R and Xapagy in the reasoning about perceptions. The ACT-R model reasons about the perceived objects by looking up a relevant chunk in the *declarative memory*. This chunk can be either retrieved by exact matching (if available) or by partial matching and choosing the best match. Alternatively, ACT-R can *blend* multiple chunks together to form a better match. In contrast, Xapagy looks up matching entities in the *autobiographic memory*. Almost always, this process will result in multiple matches, which form the weighted group called the shadow. The equivalent of the blending process does exist in Xapagy, for instance in the clustering process forming the headless shadows, however, the results of the blends are not normally stored. Both systems model the internal temporal dynamics of memory lookup (Lebiere 1999).

In the following we will try to position the Xapagy architecture by investigating its relationship to some influential trends in cognitive system design, and its relationship to concrete systems with respect to these trends.

The strong-story hypothesis states that reasoning about stories is a critical part of intelligences (Winston 2011). As Xapagy aspires to mimic the cognitive activities humans use in thinking about stories, it naturally subscribes to this view.

The role of worlds: Many cognitive systems deploy multiple, individually consistent, usually closed models which can represent an interpretation of the present state of the world, a moment in the past, a possible future or an alternate version of reality. These models are often called *worlds* or *contexts*, although many alternative names exist. For instance, Soar dynamically creates structures called *substates* whenever it encounters an *impasse* in reasoning, and needs new knowledge added to the reasoning pool. In Cyc (Lenat et al. 1990) *subtheories* are used to represent alternate versions of reality, for instance, the description of the state of the world in a certain moment in the past (for instance, we can have a microtheory in which Nelson Mandela is still a prisoner). The Polyscheme architecture (Cassimatis et al. 2004) integrates different representations and reasoning algorithms by allowing them to operate over simulated worlds.

The role of the autobiographical memory: Many cognitive systems implement an episodic/ autobiographical memory – see (Nuxoll and Laird 2007) for Soar, and (Stracuzzi et al. 2009) for ICARUS. However, the importance of the autobiographical memory for Xapagy is more critical, as the reasoning method of the system uses the authobiography the source of shadows and headless shadows.

Common serial mechanism. Xapagy makes the assumption that acting, witnessing, story following, recall and confabulation are implemented by a common serial mechanism. A number of other cognitive architectures make the same assumption, for instance, ACT-R. Combined with the other design decisions of Xapagy, however, this triggers several unexpected implications. The first is the *undifferentiated representation of direct and indirect experiences*. The stories exiting from the story bottleneck are recorded together in the autobiographical memory, with no fundamental distinguishing feature. The second implication is the *unremarkable self*. The Xapagy agent maintains an internal representation of its cognition (the real-time self), in the form of an instance labeled `"Me"`. However, this instance is not fundamentally different from the instances representing other entities. As the entity of the self can not be retrieved from memory, only recreated, an agent remembering its own stories will have simultaneously several representations of itself, only one of them marked as its real time self. Thus, *every recall of a story creates a new story*.

Handling questions. As a system reasoning about stories, Xapagy needs to deal with the problem of questions - both in terms of questions appearing in the dialogs of the stories, as well as possible questions about the story which might be answered by the agent. Recent work in statistical NLP had successfully demonstrated answering questions by parsing large databases (Ferrucci et al. 2010).

The Xapagy system's approach is more closely related to the structural question answering models, where the answers are provided from the autobiographical knowledge of the agent, rather than from the parsing of large databases. In particular, the Xapagy approach is more closely related to that of the AQUA project of Ashwin Ram (Ram 1991; 1994). The main difference is that in AQUA the ongoing assumption is that there is a schema based knowledge model which is enhanced through question driven learning: AQUA seeks truthful answers to questions externally posed, and raises internal questions for seeking knowledge. There is no such model in Xapagy: questions and answers become part of the autobiographical memory, but there are no higher level knowledge structures. In Xapagy questions are simply a specific type of sentence and they can represent a number of scenarios: questions for which an answer is sought, rhetorical questions, questions whose goal is to find out if the interlocutor knows something, questions asked to influence the discourse and so on.

Relationship to logic-based approaches. Reasoning about stories had been successfully demonstrated using logic-based approaches such as situation calculus, event calculus and action languages. For instance, Eric Mueller had applied event calculus to the understanding of script based stories of terrorist incidents (Mueller 2004).

Another approach for representing stories using logic is the *episodic logic* proposed by Lenhart K. Schubert. Episodic logic allows the translation of the English language stories into a rich logical model. In (Schubert and Hwang 2000) a portion of the story of LRRH is translated into EL. This covers not only basic narrative facts, such as "who did what to whom", but very sophisticated sentences, for instance:

The wolf would have very much liked to eat [LRRH], *but he dared not do so on account of some woodcutters nearby.*

The challenge of the use of episodic logic is that it requires a significant amount of knowledge engineering of the background rules. The authors state as a general principle that the knowledge must be specified at the most general level possible. An example of this would be rules such as "if a predator encounters a non-predatory creature not larger than himself and it is enraged or hungry, it will try to attack, subdue and eat that creature". The same problem, of course, appears in Xapagy as well, but in a different way: instead of background rules, the system requires relevant autobiographical knowledge. The history of the Xapagy agent must contain experiences with predators, which will shadow the current story.

Learning from reading. Xapagy acquires most of its knowledge by reading narratives. One of the recent approaches which acquire knowledge by reading is the work led by Kenneth Forbus in the DARPA sponsored "learning by reading" project at Northwestern University (Forbus et al. 2007). For instance, the system is able to learn about the history of the Middle East by reading historical narratives and newspaper reports and it can answer specific questions about the problem domain. The system reads texts in simplified English and relies on the Direct Memory Access Parsing (DMAP) model (Martin and Riesbeck 1986) which tightly integrates the processing of the natural language with the knowledge representation and reasoning. Background knowledge is provided by the ResearchCyc database.

Conclusions

In this paper, through a small scenario of watching television, we argued that a cognitive architecture needs to integrate at least four different cognitive mechanisms: perception, narrative, premonition and confabulatory continuation (and of course, more complex scenarios will include many others). We described how a tight-core cognitive system (Xapagy) can implement these mechanisms in an integrated way.

References

Anderson, J., and Lebiere, C. 1998. *The atomic components of thought*. Lawrence Erlbaum.

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

Beer, R. D. 1997. The dynamics of adaptive behavior: A research program. *Robotics and Autonomous Systems* 20(2):257–289.

Bölöni, L. 2013. A cookbook of translating English to Xapi. *arXiv:1304.0715*.

Cassimatis, N.; Trafton, J.; Bugajska, M.; and Schultz, A. 2004. Integrating cognition, perception and action through mental simulation in robots. *Robotics and Autonomous Systems* 49(1-2):13–23.

Ferrucci, D.; Brown, E.; Chu-Carroll, J.; Fan, J.; Gondek, D.; Kalyanpur, A.; Lally, A.; Murdock, J.; Nyberg, E.; Prager, J.; et al. 2010. Building Watson: An overview of the DeepQA project. *AI Magazine* 31(3):59–79.

Forbus, K.; Riesbeck, C.; Birnbaum, L.; Livingston, K.; Sharma, A.; and Ureel, L. 2007. Integrating natural language, knowledge representation and reasoning, and analogical processing to learn by reading. In *Proc. of AAAI*, 1542–1547.

Kurup, U.; Lebiere, C.; Stentz, A.; and Hebert, M. 2012. Connecting a cognitive architecture to robotic perception. In *Proc. of SPIE*, volume 8387, 83870X–1.

Lebiere, C. 1999. The dynamics of cognition: An act-r model of cognitive arithmetic. *Kognitionswissenschaft* 8(1):5–19.

Lehman, J.; Laird, J.; Rosenbloom, P.; et al. 1998. *A gentle introduction to Soar, an architecture for human cognition*, volume 4. MIT Press. 211–253.

Lenat, D.; Guha, R.; Pittman, K.; Pratt, D.; and Shepherd, M. 1990. Cyc: toward programs with common sense. *Communications of the ACM* 33(8):30–49.

Libet, B.; Gleason, C. A.; Wright, E. W.; and Pearl, D. K. 1983. Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential) the unconscious initiation of a freely voluntary act. *Brain* 106(3):623–642.

Martin, C., and Riesbeck, C. 1986. Uniform parsing and inferencing for learning. In *Proc. of AAAI*, volume 86, 257–61.

Mueller, E. 2004. Understanding script-based stories using commonsense reasoning. *Cognitive Systems Research* 5(4):307–340.

Newell, A. 1994. *Unified theories of cognition*. Harvard Univ Pr.

Nuxoll, A., and Laird, J. 2007. Extending cognitive architecture with episodic memory. In *Proc. of AAAI*, 1560–1565.

Ram, A. 1991. A theory of questions and question asking. *Journal of the Learning Sciences* 1(3-4):273–318.

Ram, A. 1994. AQUA: Questions that drive the explanation process. In Schank, R. C.; Kass, A.; and Riesbeck, C. K., eds., *Inside case-based explanation*. Psychology Press. 207–261.

Rosenbloom, P. S. 2011. Rethinking cognitive architecture via graphical models. *Cognitive Systems Research* 12(2):198–209.

Schubert, L., and Hwang, C. 2000. Episodic Logic meets Little Red Riding Hood: A comprehensive, natural representation for language understanding. *Natural Language Processing and Knowledge Representation: Language for Knowledge and Knowledge for Language* 111–174.

Stracuzzi, D.; Li, N.; Cleveland, G.; and Langley, P. 2009. Representing and reasoning over time in a symbolic cognitive architecture. In *Proc. of the Thirty-First Annual Meeting of the Cognitive Science Society*.

Winston, P. H. 2011. The strong story hypothesis and the directed perception hypothesis. In Langley, P., ed., *Technical Report FS-11-01, Papers from the AAAI Fall Symposium*, 345–352. Menlo Park, CA: AAAI Press.