An Investigation into the Utility of Episodic Memory for Cognitive Architectures

Ladislau Bölöni
Dept. of Electrical Engineering and Computer Science
University of Central Florida
Orlando, FL 32816–2450
lboloni@eecs.ucf.edu

Abstract

In most cognitive architectures, episodic memory is either not implemented, or plays a secondary role. In contrast, in the Xapagy architecture episodic memory is the primary means of acquiring and using knowledge. Shadowing, the main reasoning method of the system, relies on unprocessed historical recordings of concrete events to determine the agent's behavior. This paper outlines the use of episodic memory in Xapagy, and investigates whether episodic memory might play a wider role in cognitive architectures at large.

Introduction

Episodic memory is the stepchild of cognitive systems. It is an accepted fact that humans retain autobiographical information - controversies only surround the fact whether this involves a separate memory system or not (Tulving 2002). Implementing episodic memory in a cognitive architecture is not particularly difficult. The problem appears to be more whether it buys us anything.

Episodic memory is a relatively recent addition to SOAR (Nuxoll and Laird 2007) and to ICARUS (Stracuzzi et al. 2009). On the other hand, many leading architectures, such as ACT-R (Anderson et al. 2004) choose to de-emphasize episodic memory.

In this paper we propose to take a second look at the usefulness of episodic memory in the context of the Xapagy cognitive architecture. In contrast to other cognitive architectures, in Xapagy episodic memory is the *primary* means of acquiring and using knowledge. The importance of episodic memory in Xapagy is partially due to its focus on *narrative reasoning*, that is, the mimicking of some of the mental processes which humans perform with respect to stories:

- Witnessing a series of ongoing events (a story), keeping track of the participants, their identity, properties and activities.
- Following a fixed story narrated in a language (for instance, by reading a text or listening to another agent's narration).
- Predicting future events in the story, expressing surprise when unexpected events occur.
- Inferring events which (for some reason) were not witnessed; understanding narrations where some events have been not explicitly said ("reading between the lines").
- Recalling a story, summarizing or elaborating on the remembered story, chaining together remembrances.
- Daydreaming, confabulating new stories.
- Self-narrate the story by verbalizing the recalled or confabulated story, for the narrating agent's own use.
- Narrate the story to an audience, adapt the narration based on feedback from an audience, elaborate on aspects of the story or selectively narrate.
- Act as an audience for a narration, express surprise or puzzlement, request clarification or elaboration for parts of the story and ask questions.
- Perform collaborative story-telling, develop a story by alternating narrations from multiple agents.

The remainder of this paper is organized as follows. First, we provide an informal introduction to the Xapagy architecture and describe the mechanism of episodic memorization and forgetting. Next we describe the headless shadow mechanism, through which Xapagy uses the episodic memory to perform various aspects of narrative reasoning. We will describe in somewhat more detail an example involving self-shadowing and story drift.

An informal introduction to Xapagy

External look: the pidgin language

The Xapagy architecture describes the operation of an autonomous agent which can directly witness events happening in the world, and it can communicate with humans and other agents through the *Xapi pidgin language*. Pidgin languages (Sebba 1997) are natural languages with a simplified syntactic structure which appear when two groups of people need to communicate without the time necessary to properly learn each other's languages. Pidgin languages are not the native language of any group of people, and uniquely among human languages, they have a limited expressiveness. Despite their limitations, pidgin languages represent a useful stopgap measure for communication between human communities, and their creative use can in fact express a wide of human concerns.

Xapi shares some important features with human pidgin languages. It has an uncomplicated causal structure: the only supported compound statement is the quotation statement. It uses separate words to indicate degrees of properties. It does not support quantifiers. It has a fixed word order and no morphophonemic variation (more exactly, accepts a range of morphophonemic variants as synonyms from the human speaker, but it does not provide them when generating it from the computer side[1]). In addition, it uses sentence part separators ("/") and sentence boundary markers ("//") as a way of circumventing the necessity of complex text segmentation (which is beyond the objectives of the Xapagy system).

Xapi is intended to be read and written by humans when communicating with Xapagy agents. It is also used for communication between Xapagy agents, but it is definitely *not* a formal agent communication language.

A line of Xapi text represents a single sentence, with the sentence parts separated by "/" and terminated with a period "." or question mark "?". Sentences can be of a subject-verb-object form:

```
The boy / hits / the dog.
```

subject-verb form:

```
The boy / cries.
```

subject-verb-adjective form:

```
"Hector" / is-a / warrior.
```

or verb instance-verb-adverb form:

```
"Achilles" / strikes / "Hector".
The strikes / action-is / hard.
```

One or more parts can be substituted with a `wh` component, transforming the sentence into a question:

```
Wh / eats / "Red-Riding-Hood"?
```

Xapi supports a single form of compound sentence, the *quotation sentence*:

```
"Red-Riding-Hood" /says/ conversation //
  the eyes / is-a / big.
```

In some cases, the semantics of other compound or complex sentences can be approximated by sentences which refer to shared instances or verb instances. We make, however, no claim that the expressive power of Xapi matches that of a natural language.

Subjects and objects are *instances* which are either currently in the *focus*, or are newly created by the sentence. A new instance can be created by prefixing a word with the indefinite article "a/an":

```
"Billy" / hits / a dog.
```

In this example we assume Billy has been referred to before, but the dog has been just introduced in the story. Subsequent references to the already introduced instance of the dog are prefixed with the definite article "the" (which can be omitted for proper nouns).

[1]In the current version Xapi 3.2.

```
The dog / changes / angry.
The dog / bites / "Billy".
```

In pidgin, we refer to instances through one or more of their *attributes*. When we mention the attribute `[dog]`, the reference will be made to the strongest instance in the scene which has the given attribute. In some cases, such as quotation sentences, the resolution process is performed in a different scene. It is the responsibility of the speaker to choose attributes which make a reference resolve correctly.

The verb word in a Xapi sentence actually maps to a mixture (overlay) of *verb concepts* in the internal representation of the Xapagy agent. The composition of this verb overlay determines the relationship between the sentences. For actions such as "hits" or "bites", the relationship between the sentences is one of a *weak temporal succession*. One can imagine these sentences connected by "then": Billy hits the dog, then the dog is angry, then the dog bites Billy. The relationship is stronger between sentences which share instances.

We can create verb overlays which convey essentially the same action but create a stronger succession to the preceding verbs by adding the word "thus". This can be used to represent a cause-effect relationship:

```
"Billy" / hits hard / a dog.
The dog / thus changes / angry.
```

Just like the "dog" in this story is an instance which has as attributes *concepts* such as `[dog]` and `[angry]`, the action "hit" is a verb instance whose verb part has as attributes the *verb concepts* `[hit]` and `[hard]`.

Finally, there are some sentences which do not represent actions in time, and thus they are not connected by succession relationships. Examples are verbs which set attributes to instances:

```
"Hector" / is-a / warrior.
```

or establish relationships between instances:

```
"Hector" / loves / "Andromache".
```

From words to concepts and verbs

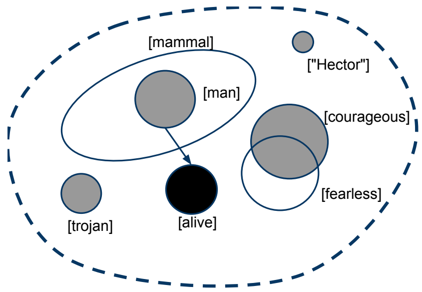
We have seen that the Xapi pidgin uses a simplified syntax, but otherwise regular English words. The *dictionary* of the agent maps nouns and adjectives to *overlays of concepts* while verbs and adverbs are mapped to *overlays of verb concepts*. We will discuss concept overlays, as the verb overlays are very similar.

An overlay is the simultaneous activation of several concepts with specific levels of energy. For instance the dictionary of a Xapagy agent might associate the word "warrior" with the following overlay:

```
[courageous=0.4, violent=0.3, strong=0.3]
```

The attributes of an instance are represented by an overlay which can be gradually extended through the side effects of the sentences. Thus, when reading the Xapi sentences:

```
"Hector" / is-a / man.
"Hector" / is-a / warrior.
```



The concept overlay of the Hector instance

Figure 1: A visualization of concepts, overlap, impact and concept overlays in the form of patches. Concepts directly added are represented with gray patches, concepts added through impact are black patches, and concepts which are implicitly present in the overlay due to their overlap with explicitly added concepts are represented as transparent contours.

the instance identified with the attribute Hector will acquire the attributes described in the overlay: man, courageous and so on.

Concepts are *internal* structures of the Xapagy agent. To distinguish them from *words*, which are external entities, we will always show them in brackets, such as `[warrior]`.

One way to develop an intuition for concepts and overlays is to visualize them as patches of a certain area in a fictional two dimensional space. Some concepts have a large area (e.g. `[animal]`), while others are smaller (e.g. `[dog]`). Proper nouns such as `["Hector"]` have a very small area. Overlays can be visualized as collections of such patches (see Figure 1).

Concepts can *overlap* on a pair-by-pair basis. For instance, there is a full overlap between man and human, meaning all men are human: *overlap*(`[man]`,`[human]`) = *area*(`[man]`). Thus, if we inquire whether Hector is human, we shall obtain a value of 1.0. There is, on the other hand, only a partial overlap between courageous and fearless: *overlap*(`[fearless]`, `[courageous]`) = 0.5·*area*(`[courageous]`). Thus, is we ask whether Hector is fearless, the answer will be $0.4 \times 0.5 = 0.2$.

Words denoting proper nouns, such as "Hector", marked in pidgin by quotation marks, are treated slightly differently: when the agent first encounters a proper noun, it will create a new concept with a very small area, and an entry in the domain dictionary associating the proper noun with an overlay containing exclusively the new concept. Other than this, proper nouns are just like any other attributes. Having the same proper noun as an attribute does not immediately imply any form of identity.

The dictionary which maps from a word to an overlay, the areas and overlap of the concepts are part of the *domain knowledge* of the agent. Different agents might have different domain knowledge - thus the meaning of the word might differ between agents.

Instances

The definition of an instance in Xapagy is somewhat different from the way this term is used in other intelligent systems. Instead of representing an entity of the real world, it represents an *entity of the story, over a time span limited by the additivity of the attributes*. For a particular instance, its attributes, represented in a form of an overlay of concepts, are additive: once an instance acquired an attribute, the attribute remains attached to the instance forever.

The advantage of this definition is that once we have identified an instance, there is no need for further qualification in order to identify its attributes (nor its actions).

What might be counter-intuitive for the reader, however, is that things we colloquially call a single entity are represented in Xapagy by several instances. Let us, for instance, consider Hector. In the Iliad, he is a central figure of the story: he appears first as a warrior, participates in several fights, while later he is killed by Achilles by a spear to his throat, and the action revolves around the return of his body to his father, Priam. Hector also appears in the Hollywood movie "Troy", but here he is killed with a sword to the chest. In the science fiction novel "Ilium" by Dan Simmons, Hector is quantum-recreated by aliens to replay the events in Iliad on the planet Mars. In the novel, Hector chooses to ally itself with Achilles and the Greeks against the aliens from an alternate reality who are playing the roles of the gods.

In the Xapagy system, these are all different instances, which share the attribute `["Hector"]` (but then, that is also shared by Hector Berlioz, the French composer). These instances, of course, are connected through various relations of identity (somatic, psychological, analogical and so on). Such identity relations are represented in Xapagy by *relations* among multiple instances, *not* by sharing the same instance.

The Xapagy system, however, moves a step beyond this. Not only Hector from the Iliad and Hector from the Ilium are represented by different instances, but Hector the live warrior and Hector the corpse in the Iliad are also two different instances, as the change from a living warrior to a dead one can not be represented as an addition of attributes.

```
1  "Achilles" / strikes / "Hector".
2  "Hector" / thus changes / dead.
3  "Achilles" / kicks / "Hector".
```

The sentence in line 2 will create a new instance, with a new set of attributes. In addition, it will connect these instances with a somatic identity relation:

```
i201 ["Hector"]/is-somatically-identical/
   i101 ["Hector"].
```

Whenever, at a later time, the Xapagy agent recalls Hector (for instance, in a conversation) it first needs to establish what instance is under consideration. Once this instance has been unequivocally established to be, for instance, the live Hector `i101`, all the attributes of the instance are also unambiguously established: we can say that he is strong, courageous etc., attributes which would not make sense applied to the dead Hector instance `i201`.

Table 1: The verb instance types and their mandatory parts

Type	Verb	SubI	ObjI	SubVI	ObjCO	ObjVO
`S-V-O`	x	x	x	-	-	-
`S-V`	x	x	-	-	-	-
`S-ADJ`	x	-	-	-	x	-
`S-ADV`	x	-	-	-	-	x
`QUOTE`	x	x	x	-	-	x

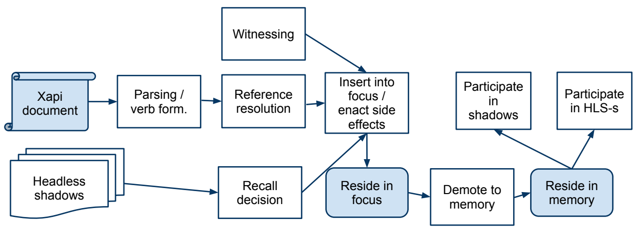


Figure 2: The life cycle of a verb instance.

Verb instances

A verb instance (VI) is composed of a random identifier and a number of mandatory relationships to its *parts*. Parts can be instances, VIs, COs or VOs. Table 1 describes the verb instance types and their mandatory parts. We can refer to the parts of a VI through a functional notation: for instance *SubI*(V) is the subject instance of VI V. The only part shared by all VIs is the *verbs* VO. The composition of this VO determines the *type* of the verb instance. For instance, the presence of the verb `[is-a]` implies a type of `S-ADJ`.

Figure 2 describes the life cycle of a VI. Most VIs are instantiated from Xapi statements. Those parts which refer to existing instances or VI are determined through the *reference resolution* process. The source of the verb instances is either external (from observation of ongoing events or reading/listening to a narration) or internal (from recall or confabulation).

The focus

The focus in the Xapagy system holds instances and verb instances after their creation for a limited time interval during which they are *changeable*. After an instance or verb instance leaves the focus, it can never return - and thus, it remains unchanged.

Instances in the focus can acquire new attributes, participate as a subject or object in verb instances and become part of relations. Verb instances in the focus can become part of succession or summarization relations, and they can be referred to by new verb instances.

A visual thinking oriented reader might think about the focus in the following way: the focus is a dynamically evolving graph. New nodes (instances and VIs) are added through various events. The same events might also create new edges among the nodes of the focus. When a node leaves the focus, it retains its attributes and edges, but it can not acquire new ones any more. So the focus can be seen as the actively growing boundary of a graph which represents the complete experience of the Xapagy agent. The graph will be only *locally connected*: it will *not* have long links, as only nodes which have been in the focus together can have links.

The instances and VIs participate in the focus with a dynamically evolving weight; the maintenance of these weights is a relatively complex part of the Xapagy system. In absence of any events, the weights are gradually decreasing. Instances are refreshed when they participate in new VIs (events). Action events are "pushed out" from the instance by their successors. In addition to these, the weights are affected by a number of other dynamic factors.

Shadows

Instances and VIs leaving the focus will be *demoted to the memory* of the Xapagy agent with a certain level of *salience*. They will never enter the focus again. On the other hand, each instance and VI in the focus has a *shadow*, a weighted collection of instances and, respectively, VIs from the memory.

The shadows are maintained through a combination of techniques whose goal is to make the shadows consistent between each other and match the ongoing story with the collections of stories in the shadows. To ensure the matching of the stories, the shadows sometimes need to forego the individual matching level between the instances. Let us consider that a Xapagy agent which had read about the duel between Hector and Patrocles:

```
"Hector" / cuts / "Patrocles".
The greek / thus changes / dead.
```

Several days later, he resumes reading the Iliad and reads:

```
"Achilles" / strikes / "Hector".
```

We already know that the two instances of Hector will not be the same: the days passed before resuming the story will be more than enough to remove the instance from the focus. Yet, the two instances will have a lot of common attributes: Hector, Trojan, warrior.

Yet the overall shape of the current fight will lead to a different shadowing: the strongest element in the shadow of Hector will be the previous instance of Patrocles, while the shadow of Achilles will contain the previous instance of Hector. The Achilles-strikes-Hector verb instance will be shadowed by the Hector-cuts-Patrocles verb instance.

Shadows are always matched to instances and verb instances in the focus. The verb instances in shadows, however, bring with themselves verb instances to which they are connected through succession, summarization and context relations. These verb instances can be clustered into weighted sets which are very similar to shadows, but they are *not* connected to current focus components. These sets are called *headless shadows* and they represent outlines of events which the agent either expect to happen in the future or assume that they had already happened but have not been witnessed (or they are missing from the narration). If an event matching the headless shadow happens, the two are combined to become a regular focus-component / shadow pair.

Shadowing is the fundamental reasoning mechanism of the Xapagy architecture. All the higher level narrative reasoning methods rely on the maintenance of the shadows. For instance, the Xapagy agent can predict the death of Hector through the shadow, and can express surprise if this does not happen. While in this example the shadow is created after the events are inserted into the focus from an external source (for instance, by reading), the opposite is also possible. In the case of recall, narration, or confabulation, the agent creates instances and verb instances in the focus based on pre-existing shadows.

Diffusion activities, spike activities and elements of prosody

The state of a Xapagy agent is modified by two kind of *activities*: *spike activities* (SA) and *diffusion activities* (DA).

SAs are instantaneous operations on overlays and weighted sets. Examples of activities modeled by SAs include inserting an instance in the focus, inserting a VI in the focus, and enacting the side effects of a VI. SAs are not parallel: the Xapagy agent executes a single SA at a time.

DAs represent gradual changes in the structure of the Xapagy agent; the output depends on the amount of time the diffusion was running. Multiple DAs run in parallel, reciprocally influencing each other. As a practical matter the Xapagy system implements DAs through sequential discrete simulation, with a temporal resolution an order of magnitude finer grained than the arrival rate of VIs.

One of the implications of the use of DAs is that Xapagy is a dynamic system, where the temporal distribution of actions, events or speech makes a difference. Xapi sentences terminated with "." imply an inter-sentence pause of 1 second, while sentences terminated with "," and ";" imply pauses of 0.1 and 0.5 seconds respectively. The Xapi sentences "-", "–", "—" and "—–" imply standalone pauses of 1, 10, 100 and 1000 seconds. During pauses, no SAs happen, but DAs do continue as normal. Using these notations, we can approximate some aspects of the prosody of natural languages. Actions in quick succession are less likely to be memorized. During long pauses, the successor / predecessor links are weaker, while for very long pauses the agent might fill in the lack of verb instances in the focus through daydreaming.

Episodic knowledge

Informally, the episodic knowledge of a Xapagy agent is the totality of the stories ever witnessed, read, heard, recalled or confabulated by the agent. Technically, the episodic memory is a repository of all VIs and instances which have been, at some time, part of the focus, and is implemented as two weighted sets (of instances and VIs).

The episodic memory is neither addressable nor directly searchable. The only way in which the content of the episodic memory influences future behavior of the Xapagy agent is through the shadow and headless shadow mechanisms.

We call the participation of the instances $S(I,t)$ and VIs $S(V,t)$ in the episodic memory their *salience* (at time t).

The salience is maintained by two DAs:

(D+) Memorization - the salience of the instances and the VIs increases while the instance or VI is in the focus

(D-) Forgetting - the salience of instances and VIs exponentially decays in time.

Memorization

The memorization DA increases the memory salience of an instance (or VI) during its stay in the focus. We will describe the equations for the case of VIs, the case of the instances is similar. Let us assume that the VI enters the focus at the creation time t_c and leaves it the demotion time t_d. In between these times, it's salience will be:

$$S(V,t_x) = \int_{t_c}^{t_x} m(t) w(VS_F, V) dt \qquad (1)$$

where $m(t)$ is the *marking rate* of the focus at time t. While the participation of the VI in the focus will gradually decrease, its salience will increase throughout its stay in the focus reaching its maximum salience at the moment when it is demoted to the memory:

$$S_{max}(V) = S(V,t_d) = \int_{t_c}^{t_d} m(t) w(VS_F, V) dt \qquad (2)$$

What this formula tells us is that the memorization level is proportional with the time spent in the focus. Action VI are memorized more than attribute assignment VIs[2]. Action verbs are memorized better when they are not a part of a quick succession of events (when the successors quickly push out the VIs from the focus). Summarization VIs are remembered more strongly than individual events: the Xapagy agent might remember the repeated hammering of a nail, but not the individual act of hammering.

In addition to this, the memorization also depends on the current marking rate of the focus. The marking rate is a slowly changing value, and is *focus-wide*. Verb instances inserted in the focus affect the marking rate not only for themselves, but also for other VIs before and after them. This way, it is possible that a marking action (such as a strong emotion) would affect the memorization of an unrelated story line, which, however, share the focus with the marking action.

The current version of the Xapagy system uses a heuristic approach for the setting of the marking rate, which, however, appears to be successful in mimicking a wide range of behaviors. The formula is based on an exponential smoothing of the contributions of each inserted VI according to the formula:

$m'(t) = \lambda_m m(t) + (1 - \lambda_m) m_c(V)$

Currently, the smoothing factor λ_m is set to 0.8. The continuation of the VI to the marking rate is dependent on the

[2]But the attribute itself is retained, because that one is dependent on the instance, which might have spent a lot more time in the focus. Thus the Xapagy agent might remember an attribute of the instance, but not when it acquired it - mimicking the limitations of human source memory.

source of the VI. To allow for an explicit setting of the marking rate, the CoreDKL contains a special verb `[marker]` which allows us to artificially increase the marking rate. The currently used values for the marking rate contribution $m_c(V)$ are summarized in the following table:

nature of V	$m_c(V)$
VIs with the `[marker]` meta-verb	1.0
witnessed events	0.5
verbalized recall	0.3
non-verbalized recall	0.1

Future versions of the Xapagy system will extend on this memorization model. Although it allows us to model explicit marker events, it does not model other aspects of human memory formation (such as the emotion caused by recalling certain experiences, which have not been very memorable when originally witnessed). There is a lot of existing work on human memory formation which can be modeled here - from the impact of levels of neurotransmitters in memorization (e.g. serotonin and dopamine levels), overall fatigue, as well as hard to measure aspects such as "interest".

Forgetting

After being demoted to the memory, the salience of a VI decreases along an exponential decay curve:

$$S(V, t_x) = \lambda^{t_x - t_d} S_{max}(V) \quad (3)$$

The salience of a VI will never increase after it leaves the focus. The recall of the VI does not increase its salience: it only creates a new, similar VI which might reinforce its recall.

For most situations, however, the main challenge of adequate remembering is not the decreasing salience of the VIs, but the initialization of the recall, which needs to create an appropriate focus and shadow.

The impact of the episodic memory on the current state of the agent

Instances and VIs which have been demoted to the episodic memory affect the current state of the agent by *shadowing* the focus. Each instance (or VI) in the focus is the *head* of a an associated instance set (or VI set) called the *body* of the shadow. The shadows are maintained such that their components reflect the previous experience of the agent with respect to the ongoing narration.

Shadows are dynamic and maintained by a number of activities:

(S+) The addition of an *unexpected* instance or VI creates a corresponding empty shadow.

(S+) The addition of an *expected* instance or VI creates a new shadow from the headless shadow which predicted it.

(D-) In the absence of other factors, all the shadows decay in time. The resources released in this DA are added to the resources of the environment.

(D+) Matching the head: instances from memory which match the shadow head will be strengthened in the shadow body. The resources for this DA come from the environment.

(D+) Matching the shadow body: instances from memory which match the shadow body will be strengthened in the shadow body. The resources for this DA come from the environment.

(D+/-) Consistency: the participation of the VI in a shadow and the participation of its parts in the shadows of the corresponding parts of the shadow head gradually moves towards a common average value.

(D+/-) Instance identity sharpening: if an in-memory instance participates in multiple shadows, the strong participations will be gradually reinforced, while the weak participations will be further weakened. The operation is resource neutral for a given memory instance.

(D+/-) Non-identity: if a shadow contains instances which are connected through the non-identity relation[3], the instance with the stronger participation is reinforced while the instance with the weaker participation is weakened. The operation is resource neutral for a given non-identity pair.

The shadow maintenance activities (and the closely related headless shadow maintenance activities controlling the narrative reasoning), are *self-regulating*, encompassing elements of negative feedback as well as resource limitation.

For instance, the head matching activity will not create an indefinitely large shadow even if the shadow head is very general, as the shadow instance relies on a limited set of resources, and once the shadow had grown beyond a certain size, its growth will slow.

Such interactions apply even between the activities. If a shadow is small, because there are few memory items matching its head, it can be extended by matching the shadow body, which brings in more remotely related items than the head match.

Reasoning with headless shadows

The narrative reasoning techniques of Xapagy are based on the mechanism of *headless shadows* (HLS), collections of related and aligned in-memory VIs which are not paired with any current in-focus VI. Like shadows, HLS-s are maintained on an ongoing basis through a collection of SAs and DAs. All the narrative reasoning models can be understood in terms of a single procedural pattern:

- **maintain** a collection of HLSs reflecting the current state of the narration
- **choose** an HLS for instantiation based on a specific criteria
- **instantiate** the HLS by creating a new VI
- **insert** the new VI to the focus and transform the HLS into its regular shadow

[3]The non-identity relation is explicitly created for distinct instances in the same story line. For example, Achilles is non-identical to the instance of Hector with which it is currently fighting. However, Achilles is *not* non-identical with Lancelot.

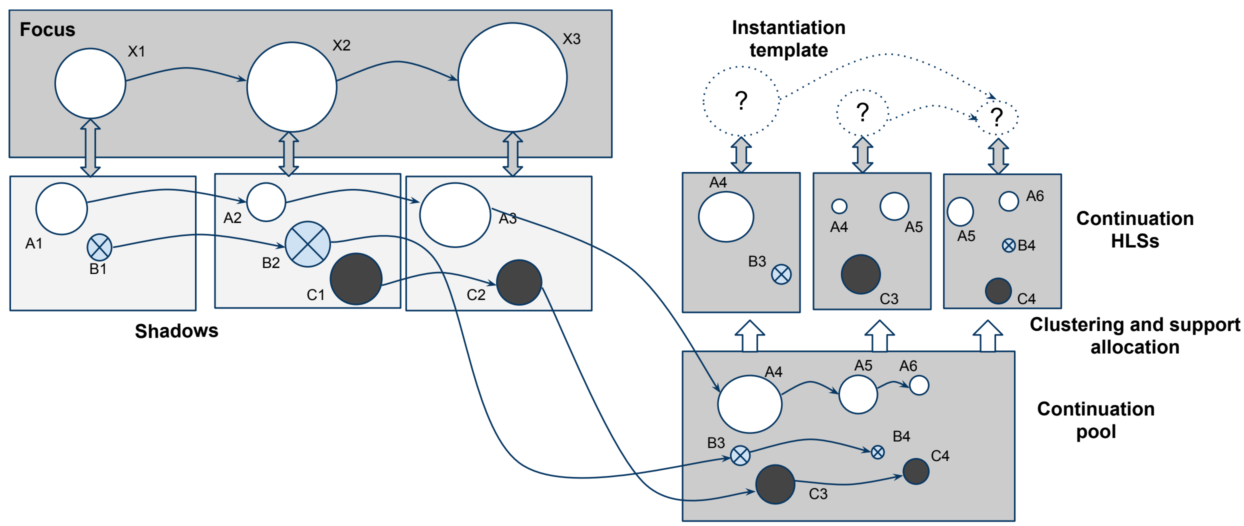


Figure 3: A simplified representation of the continuation HLS formation.

- **verbalize** the VI

The difference between the narrative reasoning methods stems from whether they skip one or more steps of the procedure, as well as the different criteria they might use to choose the HLS for instantiation.

Let us consider the example described in Figure 3. For illustrative purposes we will describe the HLS formation as a step-by-step process involving discrete steps. In practice, this is a continuous, ongoing activity performed by DAs.

In the first step, the agent maintains three VIs in the focus, with their respective shadows (the rightmost VI is the most recent one). The size of the disk denotes the participation of a given verb instance in the focus or shadow, respectively. The focus is shadowed by VIs from three different stories. The VIs linked by successor relationships from the shadows create a *continuation pool*.

In the second step, the elements of the continuation pool are clustered in continuation HLSs. The HLS contains a *template* of the possible VI it can instantiate. VIs from the continuation pool *lend support* to certain HLSs based on their match with the template.

Many of the DAs for the maintenance of shadows also operate for HLSs (for instance, story consistency). However, the low level details of the SAs and DAs performing the HLS maintenance, as well as the discussion of boundary cases such as what happens if the shadow predicts the apparition of a new instance, is beyond the scope of this paper.

An example: self shadowing and story drift

The concept of recall in Xapagy covers a wide range of situations, from exact recall of a previously seen sequence of actions to free range confabulation based on short fragments of recorded memory. We refer the reader to (Bölöni 2011) for discussion of these cases. In this section we discuss a specific situation, rarely encountered in other architectures, the problem of *self-shadowing* and *story drift*.

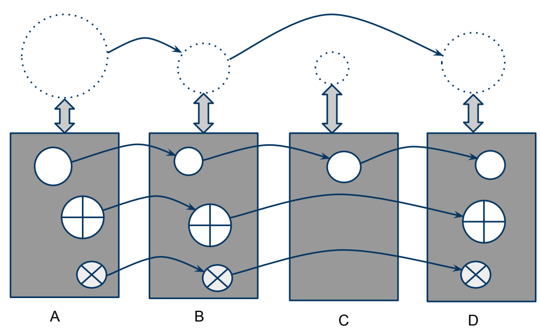


Figure 4: Self shadowing and drifting. The frequent ommission of the VI C from the retelling (for instance by constraints) drifts the story line, such that the retelling will proceed on the story line A $\rightarrow$ B $\rightarrow$ D even when no constraints are present.

The current elements in the focus of an agent are shadowed by elements from the episodic memory. The implications of these shadows create HLSs which, in the case when the agent performs a recall, determine the recalled elements. If the HLSs are supported only by shadows from a single, *dominant* story, we have a *pure recall*, where the recall exactly matches the dominant story. In almost any other situation, the recall represents a competition between stories supporting the HLSs. How accurately a recall matches the previous story depends on how well the recall initialization separates the dominant story from the competing stories. Especially when we are trying to recall one instance of many similar stories with no particular distinguishing factors (such as the events of one particular lunch), pure recall might not be possible. The psychology literature on autobiographical memory frequently calls this type of memories *repisodes*. The phenomena has been identified and well documented since the 1980s (Hudson and Nelson 1986). This is a well known challenge for the human memory, as the similarity between the events makes it more difficult to recall one particular event in an individual way.

The case we consider deals with a variant of this case, where the agent had witnessed a story only once (and the story can be highly specific), but recalled it many times.

Directly witnessed, heard or read VIs can equally be part of these shadows. This is a necessary and useful part of the system - for example, allows a Xapagy agent to make correct predictions when witnessing real world situations about which previously it had only "book knowledge". As a result, the re-narration and recall of a previously witnessed story will represented through new VIs, thus being themselves subject of successive recall. These chains of VIs will closely match the originally witnessed story, providing a competition for the original data during recall time.

Figure 4 illustrates the situation. The agent tries to recall the dominant story, which, we assume, was directly witnessed. The previous recalls are marked with $\oplus$ and $\otimes$ symbols. As these recalls are necessarily very close to the original story, they will inevitably support the specific HLSs. We call this phenomena *self-shadowing*.

This can have both positive and negative consequences. On the positive side, accurate recalls reinforce the HLSs of the dominant story, and make it less likely that foreign stories can compete with the recall. This is especially important if the agent recalls events for which the salience had naturally decayed.

With this mechanism, the Xapagy system automatically exhibits learning through repetition (it is a future task to investigate whether such observations about human learning such as the Eberhardt learning curve or Paul Pimsleur's spaced repetition technique can be automatically mimicked by the Xapagy agent).

Let us now consider, however, the case when the recalls have not been fully accurate. If the recalls differ from the dominant story line in a consistent way (for instance, by regularly skipping some events, the situation in Figure 4) there will be a strong likelihood that a specific recall will follow not the original story line but the "usual way of recalling it". This phenomena, which we will call story drift, mimics suppressed memories and self-deception in humans. A closely related situation is when the shadowing stories are not coming from internal recalls, but from external retelling of the same story in modified form. This might mimic human behavior where, for instance, excessive external praise might modify the person's own recollection of certain stories in the past.

A very famous example of self shadowing is the case of Nixon's counsel John Dean analyzed by Neisser (Neisser 1981). In several cases where Dean had recalled with high confidence details of meetings with Nixon, it turns out that what he had recalled was heavily modified by his fantasies and modified recalls of the event (both omission of events and insertion of events have been noted). It was also found that at its testimony, the main source of remembrance is not the original events but the statement about it he made a couple of days before.

Conclusions

An investigation into why episodic memory collects such a small mindshare in cognitive architectures must be done carefully - it is difficult to make general statements without building straw men, attributing researchers points of view which they never held. We are thankful for the reviewers to point this out. Yet, the point that episodic memory plays a minor role in current systems is undeniable. Some architecture designers might consider that episodic memory does not hold interest for the specific target applications: for instance, it might be unacceptable for the "life experience" of the agent to influence its performance at a specific task. Other researchers might consider episodic memory as a future task which need to be undertaken once solid foundations are laid in other areas. Arguments exist for the fact that unprocessed episodic memory, in the form of simple recorded experiences is not useful in itself, and these experiences must be integrated into more abstract cognitive structures before being used.

All these arguments have value. What our claim in this paper is, however, that a useful system can be built with a much stronger reliance on episodic memory than it is customary in cognitive architectures. Xapagy uses episodic memory in its most raw format of unprocessed recording as the foundation of all behavior types, and has a very simple, minimally structured model of conceptual and dictionary knowledge (and *no* model of procedural knowledge). At least in the field of narrative reasoning Xapagy exhibits a number of complex behaviors mimicking human reasoning models as shown in our example in this paper – for more examples, please consult (Bölöni 2011).

References

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.

Bölöni, L. 2011. Xapagy: a cognitive architecture for narrative reasoning. *preprint* arXiv:1105.3486v1.

Hudson, J., and Nelson, K. 1986. Repeated encounters of a similar kind: Effects of familiarity on children's autobiographic memory. *Cognitive Development* 1(3):253–271.

Neisser, U. 1981. John Dean's memory: A case study. *Cognition* 9(1):1–22.

Nuxoll, A., and Laird, J. 2007. Extending cognitive architecture with episodic memory. *Proceedings of the 22nd national conference on Artificial intelligence - Volume 2* 1560–1565.

Sebba, M. 1997. *Contact languages: Pidgins and creoles*. Palgrave Macmillan.

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

Tulving, E. 2002. Episodic memory: From mind to brain. *Annual review of psychology* 53(1):1–25.