Article

An ACT-R Based Humanoid Social Robot to Manage Storytelling Activities

Adriana Bono [1], Agnese Augello [2], Giovanni Pilato [2,*], Filippo Vella [2] and Salvatore Gaglio [1,2]

[1]   Department of Engineering, University of Palermo, 90128 Palermo, Italy;
     adriana.bono@community.unipa.it (A.B.); salvatore.gaglio@unipa.it (S.G.)
[2]   ICAR-CNR, Italian National Research Council, 90146 Palermo, Italy; agnese.augello@cnr.it (A.A.);
     filippo.vella@cnr.it (F.V.)
*   Correspondence: giovanni.pilato@cnr.it

Abstract: This paper describes an interactive storytelling system, accessible through the SoftBank robotic platforms NAO and Pepper. The main contribution consists of the interpretation of the story characters by humanoid robots, obtained through the definition of appropriate cognitive models, relying on the ACT-R cognitive architecture. The reasoning processes leading to the story evolution are based on the represented knowledge and the suggestions of the listener in critical points of the story. They are disclosed during the narration, to make clear the dynamics of the story and the feelings of the characters. We analyzed the impact of such externalization of the internal status of the characters to set the basis for future experimentation with primary school children.

Keywords: cognitive agents; interactive storytelling; social robots; ACT-R

1. Introduction

Storytelling is an art whose ultimate goal is to touch the listener and completely engage him in an emotional journey. The old and fascinating art of telling stories is continuously evolving, until today, to meet new technologies and to exploit new fruition ways that characterize the digital era. This task has stimulated a growing interest in Artificial Intelligence (AI) and Robotics research fields and several storyteller robots were proposed to support teachers in their activities [1–5].

Many studies demonstrated the effectiveness of using virtual companions in learning [6,7]. On the other hand, robots represent an attractive and innovative tool that by capturing the pupil's attention, contribute to the achievement of many goals, such as the story recall and its emotional interpretation [8–10].

Hsu et al. [6] investigated the possibility of exploiting a physical robot in a learning environment. In particular, an experiment was carried out on the different effects of learning between the use of a virtual companion and the use of a physical robot. The authors of [6] report that an effective engagement leads to lower dropout rates and improves student success. The result of the study is that the robotic companion can lead to some advantage, mainly due to physical interaction, but it turned out that this is only true if the robot is well framed in the context, in the environment in which the lesson is delivered and in the adequate quality of the material that is delivered to the student. An interesting conclusion is that if the robot is not well coordinated with the whole environment, the participants consider the robot as simply a machine.

In a survey conducted by Belpaeme et al. [11] it is also pointed out that while many authors have found benefits from using a social behavior of the robot, leading to a better learning outcome, the social behavior of the robot has to be accurately designed taking in account the context and the educational task to carry on. The study shows that social behavior certainly improves the learning rate among humans. Some studies suggest that these influences also extend to human-robot interaction.

The motivation behind increasing attention with the use of a robot is to be found in the human psychological aspects that affect the developmental stages of children. It was shown that children can easily anthropomorphize robots by attributing them characteristics that are generally associated with human beings [12]. The anthropomorphized view of a robot facilitates the interaction with children, especially when the robot shows human features. Moreover, the tendency of human beings to aggregate in social groups, thus fostering interpersonal relationships, is strongly visible in children. The possibility of interacting naturally with a robot triggers curiosity and attention of children about the activities carried out by the robot throughout their interaction, which can thus become a moment of the development of social-cognitive and linguistic skills. The social relationship established between the child and the robot is also characterized by a certain degree of trust that a child feels he can assign to his *player*.

On the other hand, in [13], how social behavior of robots may impact the learning process of children was further investigated. While the study showed that the employment of a robot can improve the learning process, the adoption of social behavior by the robot in that context did not lead to significantly better learning. The authors explained this strange inconsistency with the consideration that perhaps during the lesson, students were more attentive to the social behavior of the robot than to the content of the lesson itself. A further explanation that the authors give is that the social behavior of the robot places a heavier cognitive load on the child, inhibiting his ability to process the provided information. While this can be a disadvantage in a learning environment, in our opinion, it can fit well a storytelling scenario, where the social behavior of a robot playing the role of a character is the subject itself and a key point of the whole approach.

Moreover, robots can be used in processing information about bullying experiences among children in an age group between eight and twelve years, as reported by C.L. Bethel et al. [14]. In that study, interviews were collected with a group of 60 children who were asked a series of questions about their experiences with bullying. About half of the children were interviewed by an educational professional, while the other ones were interviewed by an NAO robot. From the study, it arose that for one item, participants were significantly more likely to open up with the robot interviewer than to the human interviewer.

A social storytelling robot was implemented in [15]; the robot uses a persuasion model, based on a categorization of humans Personality Traits (PTs), to convince the player to choose the best action to perform within the plot.

Another example that focuses attention on the use of a robot in storytelling is reported in [16,17], in which it was demonstrated how the use of a Softbank Robotics NAO with human social behaviors can help pre-school children to better memorize the details of a story.

In [18,19], we introduced NarRob, a storytelling humanoid robot provided with modules that process the semantic content of a story and, as a consequence, enrich the verbal narration, including proper gestures that allows increasing the underlying communicative message [20]. The non-verbal narration is essential to facilitate a mood induction and to improve the storytelling experience. The non-verbal narration is essential to facilitate a mood induction and to improve the storytelling experience [21,22].

In [19], we stressed the importance of introducing NarRob storytelling activities for the improvement of children's social skills and awareness. In this work, we focus on this goal since our opinion is that a robot, through the narration of a story, can help children in understanding how to manage different social situations. By listening to a story, a child can realize what consequences could derive from some choices in specific circumstances. The robot should explain what is happening in the story, who is involved and with which role, engaging the child in the most critical moments of choice of the main characters, triggering his attention and also asking for his opinion.

We implemented a dynamic interactive storytelling activity about bullying. The story can be seen as the result of the actions performed by different characters, each one of them with his role within the given narrative context. The narrative context is a dynamic environment where the characters must be

able to reason according to their beliefs and objectives. Furthermore, the system allows the player to influence the actions of the characters, who remodel their objectives according to the choices made by the player.

To obtain such a kind of storytelling activity, we implemented the characters of the story as autonomous cognitive agents whose behavior was modeled exploiting a cognitive architecture. Considering the implementation through a cognitive process allows somehow simulating the reasoning processes and the adaptation to new situations that a real character would face, with the result of higher credibility and dynamism in the development of the story.

We exploit the ACT-R [23] cognitive architecture, to implement three agents: the *narrator*, the *protagonist* (or "main character"), and the *antagonist*. These three agents share the same narrative context, that is the story in which the characters act. The characters analyze the narrative context, and according to their knowledge and their goals, they plan the actions to perform, contributing to the evolution of the plot.

The story characters are interpreted by humanoid robots using the NarRob framework. The role of the robots, at present, is only that one of digital mediators. They are used as tools for understanding social situations. We did not give the robots a predominant social behavior.

In particular, the SoftBank NAO and Pepper robots were used. The reasoning processes of the characters are disclosed during the narration, to make clear the dynamics of the story and the feelings of the characters. The narration is enriched with gestures, with a tone of voice modulated according to the events of the story being told, and with the change of the color of the NAO LEDs to simulate the emotions related to the narration.

The story can evolve according to the suggestions of the child, leading to distinct epilogues and raising different emotions in the characters. However, we conceive this external intervention in the story as something that should not directly change the plot in a predetermined manner. Instead, it should influence the beliefs and the goals of the characters driving them to plan the best actions to perform.

In what follows, after a brief overview of the state of the art and the ACT-R architecture, we describe in detail the implemented system focusing on the cognitive models of the characters and the enactment of the interactive storytelling process on NarRob. Finally, an evaluation section discusses the impact of the externalization of the character's internal status on the story listeners, and a concluding discussion is reported.

2. Interactive Storytelling

The evolution of storytelling has not only affected the technology of communication channels, but it has, above all, revolutionized the listener's role by removing him from the old figure of being a passive recipient and giving him an active role. Modern storytelling claims a new role for the listener, projecting him into the story and increasingly involving him. The listener is called to actively participate in the narration by helping the characters along their path. This increased involvement has the effect of strengthening the emotional link established between the player and the characters [24].

Several interactive storytelling systems were developed. Façade [25,26], for example, is an interactive drama, conceived by Michael Mateas and Andrew Stern, where the user plays the role of a friend of Trip and Grace, the characters of the story, a married couple who invite him to spend a funny evening in their apartment. During the development of the story, the problems of the couple emerge, and Trip and Grace take dramatic and heavy tones, forcing the player to play the role of an intermediary. Trip and Grace are implemented as *believable agents*, characterized by a large set of behaviors written in ABL (A Behavior Language) [25]. These behaviors are modeled following a hierarchical logic, and they are represented through a dynamic tree structure. The behaviors are instantiated in *beats* (the atomic units of the story) by a drama manager.

Another interactive drama is Versu [27]. In this game, the player has to improvise a specific role in a dramatic situation. In Versu, all the characters are represented by autonomous intelligent agents that choose the actions to be performed according to their desires; the drama manager takes part in the

story in deadlock situations. Another interesting interactive game is *FearNot!*, designed for educational purposes. FearNot! narrates a story of bullying where the player acts as an "invisible friend" and gives advice to the victim by entering text through the keyboard to solve his problems. The narration is based on an emerging narrative mechanism that allows the characters to independently select the actions, thus allowing the narrative to naturally emerge in a non-deterministic manner. The choices of the characters are driven by their emotions (emotional weights are assigned to their perceptions). The emotions influence both a reactive behavior and the planning process of the agents. The agent could reduce the importance of a goal or even delete it if it is "impossible" to reach from an emotional point of view. In the present work, we take inspiration from FearNot! to improve the storytelling system of NarRob, starting from the story plot available at [28].

The main difference with FearNot! is that we implemented the characters of the story by using a cognitive architecture, namely ACT-R. This choice was driven by the desire of mainly focusing on the cognitive processes developed and activated by the characters of the story. The characters were implemented as cognitive agents free to reason according to their own knowledge and to act to achieve their goals.

Even if the story takes inspiration from the plot presented in FearNot!, it was reshaped in terms of several possible episodes, characters, and contexts in which the characters are in. Some episodes were rewritten or added.

3. The ACT-R Cognitive Architecture

Cognitive architectures (CAs) are formal models of cognition; they can be exploited not only to better understand how the human mind operates, but they can also be used as a basis for implementing artificial intelligent agents that can, somehow, simulate some mechanisms of the human thought.

In this context, the perspectives and challenges of CAs in AI and general AI are summarized in [29]. Among the several CAs that were suggested in recent years, the ACT-R (Adaptive Control Things-Rational) was proposed as a hybrid cognitive architecture, characterized by symbolic and sub-symbolic components [23].

This architecture was successfully used to create models for learning and memory, problem-solving, language processing, decision making, perception and attention, and cognitive development. Figure 1 shows a general representation of the architecture, where it is possible to examine the set of modules and buffers that define its symbolic components.

These modules represent particular human cognitive and perceptual abilities aimed at processing multiple and significant information from the environment, as well as acting on it by implementing actions that allow the achievements of specific goals. Each module is associated with a specific *buffer* whose main function is to forward requests to its corresponding module and to retain the received information, which is made available for the evaluation process by the *Procedural System*.

The procedural system acts as a coordinator. It contains the set of known production rules that can be applied by the model for achieving its goals. Its main function regards the analysis of a buffer at a given time looking for specific patterns to identify a production rule that can be activated. The rules, written in Lisp language, have a particular (*if-then*) structure; the first part declares the conditions on the buffers that must be satisfied in order to apply the changes, while the requests are present in the second part of the rules themselves. The choice of the production that must be activated can be made flexible by applying a *Partial Matching*, one of the sub-symbolic features of this architecture. When this feature is enabled, the tests on the buffer values are slightly relaxed. If more than one production matches, it will activate the production with the highest utility value. The utility associated with a rule can be fixed, or it can dynamically change according to the number of its execution. The *Declarative module* has the task of managing the knowledge of the model through a set of *chunks*. A chunk represents the atomic knowledge of the model. Chunks are characterized by a list of key-value pairs, where the keys and values are based on a set of symbols, e.g., constants or references to other blocks. Depending on

the concept that must be represented in the agent's knowledge, different types of chunks, each one with a proper structure, can be defined. A general definition of a chunk is the following:

$$(chunk - type \;\; name\_chunk \;\; slot_1...slot_N)$$

where *name_chunk* uniquely identifies the type of defined chunk, *slot* represents a field that can contain a value constituting a part of the atomic information of the chunk. A chunk can usually contain more than one slot, thus forming a key-value pairs list. The collection of instances of the defined chunk-types represents the set of information held by the agent about a certain concept. Each chunk intrinsically has an activation value, indicating its access frequency. When several chunks satisfy a request, the one with the highest value is considered and, in the case of multiple chunks with the same activation value, one of them is randomly selected.
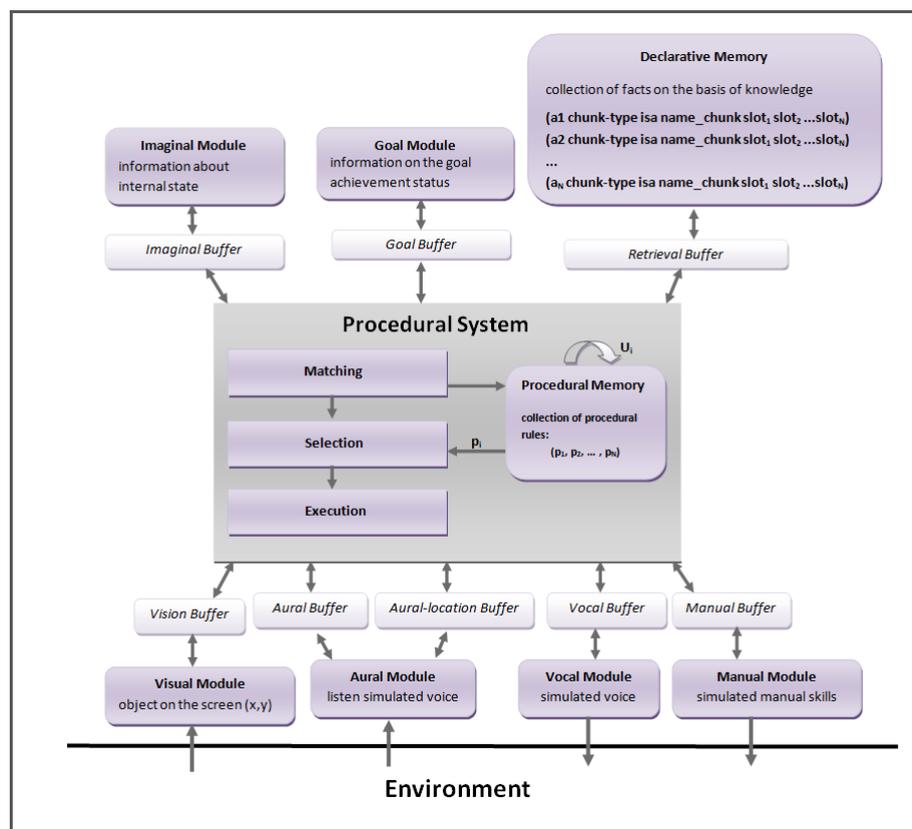

Figure 1. A representation of the ACT-R architecture.

The sensory abilities of the generic model in this architecture are implemented using two main modules: the *Visual Module* and the *Aural Module*, which continuously scan the environment to process specific types of information. In particular, the *Visual Module* implements the visual capability of the agent, allowing it to shift the attention towards a specific point in the perceived scene. This module includes the *visual-location buffer*, to keep track of the position of an object on the screen, and the *visual buffer* to identify its symbolic representation. The *Aural Module* provides the agent with the ability to perceive sounds. It works simultaneously on two different buffers: the *aural-location* buffer handles requests about the source of the sound, the *aural buffer* keeps track of what has been heard. Each audio event is managed by this module, which transforms each audio into chunks callable by from the model when it is necessary.

The ability to act on the environment is provided by two other modules. The *Speech Module* provides the model with the skill of communicating through words with other models using the

vocal buffer. This module works through three internal states: preparation, word-sound processing, and execution.

Finally, the *Motor Module* provides the model with motor skills allowing the use of external devices, such as a keyboard.

The current state of the model is identified by the *Goal module* and the *Imaginal Module*. The *Goal Module* provides the system with a *goal buffer*, that is typically used to maintain the current goal of a model. The *Imaginal Module* maintains the internal state of the model. The corresponding buffer, the *imaginal buffer* contains a chunk, whose slot values can be updated during its execution to carry out the correct assessments on the actual state of the model within the environment.

The coordinated work of these modules, with their respective buffers, allows for the definition of interactive ACT-R activities simulating human cognitive and perceptive abilities.

4. ACT-R Based Storytelling

The work illustrated in this paper exploits the ACT-R architecture to model a storytelling activity aimed at improving children's social skills. By modeling the characters using appropriate cognitive models, the narrative emerges from the choices of the characters rather than from a predefined script. In turn, the character's decisions derive from their knowledge and goals and the succession of events. Such an implementation gives greater importance to the characters' autonomy and reasoning. The storytelling activity is designed to help children to better understand the effects of their behavior within society [30–32].

From an implementation point of view, our system works through the coordination of two files: a Lisp file named *Models* containing the individual ACT-R models that implement the characters, and another file, named *Environment* that deals with managing the choices made by the players who influence the narrative path of the characters of the story.

The implementation strategies focus on the aspects related to the memory of the ACT-R model, the type of modules to be used, and the degree of generalization of the applicable rules. The sub-symbolic components provided by the architecture are used to provide greater flexibility to the models, as happens with the use of partial matching whose purpose is to relax the constraints on buffers to provide a greater variability. The storytelling is obtained through the dialogue between the characters, implemented as independent cognitive models.

The following sections describe in detail the system implementation and its fruition through the interaction with a humanoid robot.

4.1. Story Plotting and Evolution

In this preliminary phase of the work, the proposed solution is largely inspired by the FearNot! plot. The story told in FearNot! addresses the delicate issue of bullying among children. The main character of the story is John: a child recently moved to the school of Meadow View and who, due to his shyness, struggles to make new friends. Some of the boys led by Luke, strong in their popularity, decide to target John by opposing him in his activities and humiliating him as soon as possible.

However, although our story leaves a few points of FearNot! story unchanged, for example, the scene of chocolate theft or the names of the characters, the rest of the story has been rewritten. Our episodes were planned and described to provide the agents with all the information required to implement their internal cognitive processes. So we chose to manage some episodes, necessary for the representation of a complete narrative dramatic arc, to focus more on the implementation of the knowledge of the world to be provided to the agent, on his reasoning ability to evaluate the situations and events and, as a consequence, letting it deliberating the most appropriate actions to achieve its goals. In particular, we wanted to dwell on the reasoning and motivations that push a character to perform one action or another one, rather than on the type of action carried out by an agent.

The story evolves according to a typical dramatic arc, following the Freytag dramatic structure (Figure 2) [33], starting from an exposition phase, and continuing with a set of events, among

which there is an accident that leads to the point of maximum tension (the *climax*). Finally, the climax is resolved to reach an epilogue.

Specifically, the *initial event* is directly managed by the narrator, who introduces the context of the events and the main characters to the player. The goals pursued by the different characters (according to their cognitive models) inevitably change along the course of the story, according to the interactions with the other models. The different goals pursued by the protagonist and the antagonist lead to the presentation of a complication for the main character. The complication is triggered by the behavior of the antagonist who carries out one or more actions whose ultimate aim is to move the protagonist away from its goal.

The complication leads to the *climax*, for which it is necessary to find a solution to close the conflict. The subsequent actions and the intervention of the player, who is asked for an opinion in situations of stalemate, brings the narrative arc towards the end of the story, the *final event*.
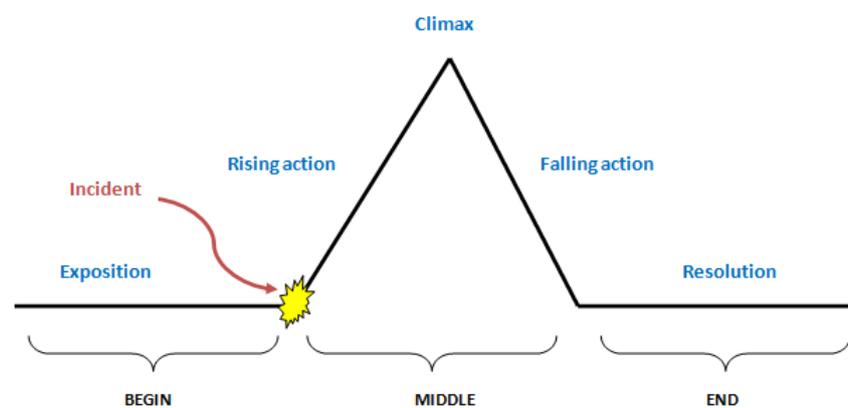

Figure 2. Freytag's Pyramid.

4.2. Characters Cognitive Models

The ACT-R architecture, as seen above, is characterized by different modules, each of which has the important function of implementing a specific human cognitive function.

In the proposed storytelling system, the story takes place in the form of a dialogue between the characters, who interact with each other within the same environment. For our purposes, therefore, we chose to use only some modules of architecture that mainly allow simulating the dialogue and the cognitive reasoning of the characters. The modules used for the implementation of the models of our system are:

- Declarative Module
- Procedural Module
- Speech Module
- Aural Module

The declarative module assumes an important role in the implementation of the actors who behave as cognitive agents ACT-R. The knowledge of these agents is put in the form of chunks allocated in the declarative memory. The knowledge of the agent is the collection of the information regarding the other agents and the information related to the semantic meaning of some words that allow the evolution of the dialogue between the characters. The Procedural Module instead has a collection of all the rules that the model uses to process inputs coming from the outside world and managing its knowledge. The use of these rules thus allows the agent to run the actions. The Speech and Aural modules are used together to simulate the dialogue between the characters of the story. The Speech module allows us to synthesize the vocal strings outwards through its *vocal* buffer, while the Aural

Module is responsible for the cognitive process of listening. This module processes the source of the heard sound-word and its semantic content through its two (*aural-location* and *aural*) buffers.

Each character in the story is represented as a cognitive model by ACT-R. Three different models were defined:

- M1: The victim's model;
- M2: The bully's model;
- M3: The moderator model.

In particular, the moderator leads the narration of the events and facilitates the resolution of deadlock situations through the help of the human player. This ensures the interactivity of the system and gives the listener the opportunity to establish an empathetic relationship with the characters of the story.

The interaction between the models within the environment represents one of the fundamental points of our work; it would not be possible to generate the dynamics of the story without it. This interaction between the characters is made possible by exploiting the *Speech* and *Aural* Modules, which make it possible to simulate a real dialogue. Each model uses these two modules to interact with the other models and to generate changes within the environment.

The models, running on the same narrative context environment, use these modules as sensors to filter new inputs, and generate changes in the environment. When the model synthesizes a vocal string through the *Speech Module*, the string is reproduced within the environment. Each running model uses its *aural-location* buffer to intercept the sound strings spread in the environment by the other agents. When this buffer identifies a string, it recognizes its source and sends a request to the associated *aural* buffer to create a chunk of *sound* type, whose function is to contain at its internal the perceived string. The definition of this particular type of chunk is typical of the ACT-R architecture. After identifying a vocal string through the appropriate modules, the cognitive processes of the model begin to focus on the analysis of the declarative memory to identify a possible match between what was listened to and what it is represented in its knowledge base.

As mentioned above, each model will try to pursue its goals during events that could be modified due to interactions with other models. These goals are managed by the *Goal* Module of ACT-R, while the *Imaginal* Module is used to keep track of the internal states necessary to achieve the goals among the various steps.

The declarative memory, made up of chunks, is used to express the *facts* owned by the model. The collection of facts represents, thus, the agent's knowledge of the world.

In this phase of the work, we chose to equip the model with some information that can be used during the interaction with the other models, such as basic information about the other characters in the story and a set of notions regarding the meaning of some terms to manage the problem of understanding specific sound words. For example, regarding the *M1* model, we defined the *info* chunk-type with atomic information such as details about the name, the role played within the story, and the age of models *M2* and *M3*.

```
(chunk-type info model role name age)
```

Starting from this definition of the chunk-type it is possible to instantiate in the declarative memory of the *M1* model three different chunks (called *p1, p2, p3*), containing information regarding each actor of the story:

```
(p1 ISA info model m1 role victim name "Luke" age 11)
(p2 ISA info model m2 role bully name "Paul" age 11 )
(p3 ISA info model m3 role assistant name "Pepper" )
```

Instead to improve the management of the dialogue and allow the model to understand the speech, we defined in our code the *semantic* chunk-type. It provides the model the skill to understand the *meaning* of the *word* exchanged with the other models. So the set of instances of this chunk-type

allows the model to think about what it has just heard and to make the correct evaluations based on its knowledge. It is formalized as follows:

```
(chunk-type semantic meaning word)
```

Among the instances of this chunk-type contained in the declarative memory, we can identify the underlying assets, which have the purpose of associating positive or negative feedback with certain words:

```
(p isa semantic meaning positive-feedback word "yes")
(q isa semantic meaning negative-feedback word "no")
```

In the underlying *listening-comprehension* production, it is possible to examine how the *semantic* chunk-type is used by the model in the process of understanding. This production is used when the model listens to a sound-word. At this specific moment it is possible to find the word just heard in the aural buffer, which is responsible for the listening activity. The model is thus ready to search its meaning by forwarding a request to the retrieval-buffer to access its declarative memory to find some match with what it knows.

```
(p listening-comprehension
=aural>
isa sound
content =word
?retrieval>
state free
==>
+retrieval>
isa semantic
word =word
!output! (M1 start reasoning..)
)
```

At this point, the procedural system evaluates which action must be executed considering the current state of the buffers.

If no match is found between the sound-word heard and one of the chunks present in the declarative memory, then the model ignores what it has heard, and it continues the activity it was carrying out. Instead, if the match is successful, the chunk extrapolated from the declarative memory will be placed in the retrieval buffer, and it can be evaluated by the Procedural System for the identification of an applicable rule. For example, the retrieval buffer forwards to its module the request for information regarding a sound-word listened to. The Declarative Module finds the match and loads the identified chunk into its buffer, and, according to the values identified by the chunk's slots, the Procedural System chooses to activate one rule instead of another one. In the specific case shown below, the listened word has a meaning relative to an answer, or feedback, expected by the model. The model evaluates the sequence of actions to be performed based on the value corresponding to the *meaning* slot.

In this particular example, if the value contained in this slot is *negative-feedback*, the model chooses a closing reaction and protection towards oneself, such as "ignoring the interlocutor", as defined in the following rule:

```
(p negative_feedback
=retrieval>
isa semantic
meaning negative-feedback
?vocal>
state free
==>
+vocal>
cmd speak
string "ignore"
!output!(Ignore model...)   )
```

In this rule, after evaluating the value contained in the chunk located in the *retrieval buffer*, a test is carried out on the vocal buffer. If this buffer is free, a specific vocal string can be synthesized.

Instead, if the content of the slot meaning indicates a *positive-meaning*, then the model can evaluate the execution of an action that has the purpose of strengthening the affiliation with his interlocutor; for example, it can propose to start a collective activity or to entertain a discussion on a given topic.

The choice of collective activity to be proposed or the type of topic to be discussed is completely free from further constraints. Since in our system the sub-symbolic activation values of the chunks contained in the declarative memory of the models are set to the same value, the chunk that is selected from the set of chunks that satisfy a request is chosen in a completely free manner. This degree of freedom provides a certain level of uncertainty on the behavior chosen by the model, starting from a set of feasible behaviors known, and this inevitably induces variability in history. So providing more knowledge to the model allows you to get a character capable of implementing actions that can affect the course of history. However, it could arise the problem that the continuous identification of new actions could change the structure of the story so to the Freytag's dramatic arc is not respected. It could happen that the models get stuck in the initial phase without ever reaching the central phase of the conflict. To avoid this problem the narrator, interpreted by the *M3* model, observes the interactions between the models and the progress of the events, intervening only to redirect the story to avoid that the structure can shift from the one described by Freytag providing new elements and new narrative contexts to the characters. The structure of the story is therefore kept stable by the narrator while the uncertainty is linked to the behaviors assumed by the models in certain contexts.

In ACT-R, the sub-symbolic components provide a powerful tool to make a model dynamic and flexible within the environment. Each rule can be associated with a utility value; this value can be either static, or it can dynamically vary along with the execution of the model through a learning process based on the learning rule of Rescorla-Wagner [34]. In this phase of the work, we exploited this learning rule only on those rules more frequently used by the model, such as *listening-comprehension*. We also exploited the partial matching feature, to allow the model to take into consideration those rules whose conditions partially satisfy the current state of the buffers. Since the rules refer to actions or behaviors that the model can perform in certain situations, this element provides the story with a certain degree of indeterminacy concerning the actions performed by the characters. The Figures 3 and 4 show two different states of the active buffers in the model of John, the main character, in an initial phase of the story, where there is no conflict with Luke. Here John is walking towards school and meets Luke. The model will evaluate the status of its buffers and the applicable productions, and finally, it will choose to execute one of the possible actions (Status A depicted in Figure 3).
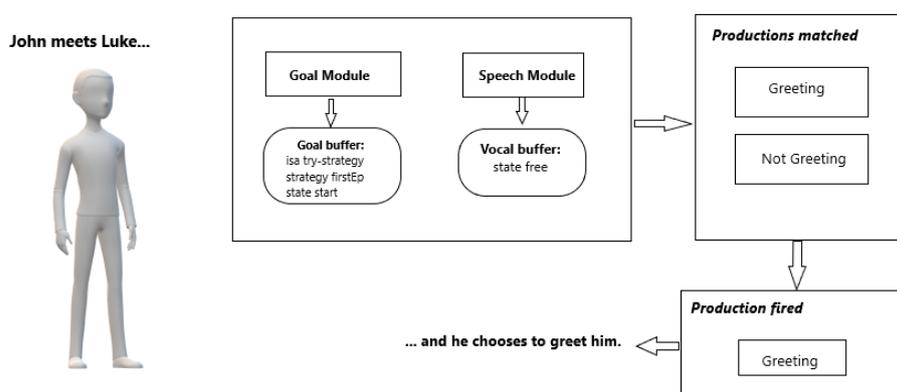

Figure 3. Cognitive model of M1: Status A.

The type of choice implemented by the model is important since it can speed up or delay the evolution of the conflict within the story. For example, if John decides to say goodbye to Luke, he will expect his greeting to be returned. If this happens, then he can choose to start a friendly discussion with Luke or continue with his initial goal of going to school (Status B depicted in Figure 4).
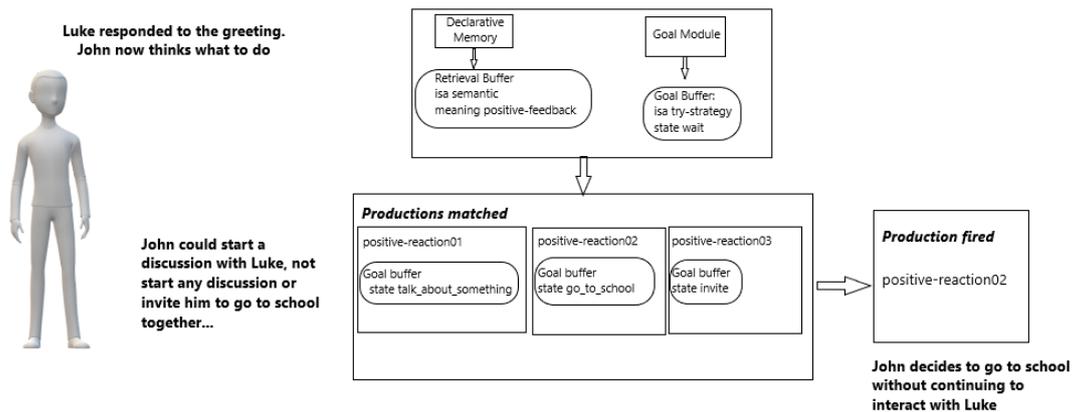
Figure 4. Cognitive model of M1: Status B.

Another important aspect of our system is the interaction between the characters and the player. This is implemented through the distinction of particular points in the plot, called *turning points*, used by the narrator to take control of the story. In these points of the plot, the main character is called upon to make a certain choice that could have important effects for him and the other characters involved in the story. In this case, the narrator will directly address the player inviting him to choose one of the two proposed actions to be performed by the protagonist.

In our system, we implemented several similar situations, including the following one: "*It's breakfast time and John, the bullying victim of the story, starts eating his chocolate. The bully thus decides to oppose John even in this situation and chooses to steal the chocolate bar from the victim solely to humiliate him in front of the other companions.*"

In this particular episode, the victim can do two things: run away with chocolate or leave it to Luke. Hence, the narrator proposes the two following alternatives to the player by showing a screen in which it will be possible to select a specific choice. Figure 5 depicts the selection window that is shown at a given point of the narration.
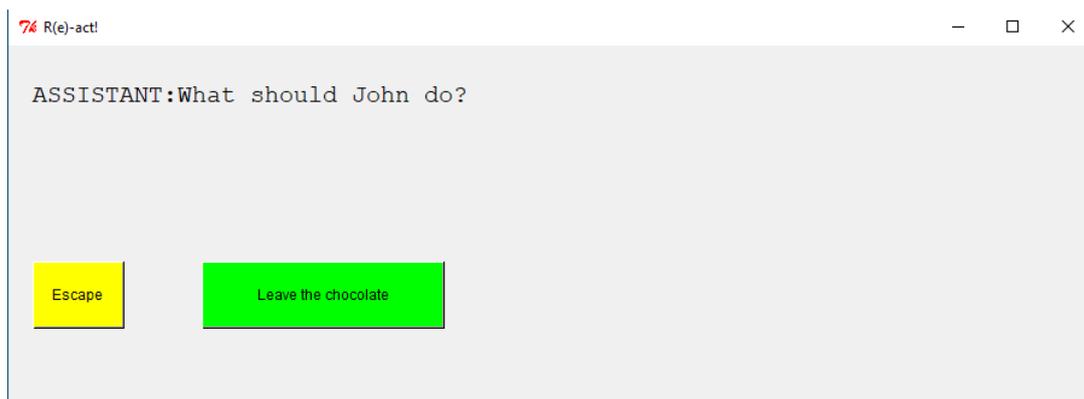

Figure 5. Selection Window.

When the player makes his choice, the chunk in the model's goal buffer is appropriately modified. For example, if the player's choice falls under "Escape" then the objective of the *M1* model is modified as follows:

```
actr.set_current_model(player1)
actr.define_chunks(['goal','isa', 'strategy','state','escape'])
actr.goal_focus('goal')
```

In this way a new goal is set in the respective buffer and the following *escape* rule is activated.

```
(p escape
=goal>
state escape
?vocal>
state   free
==>
+vocal>
cmd speak
string "escape"
=goal>
mode escape
!output! (I escape with chocolate)
)
```

If instead the choice falls under "Leave the chocolate" the chunk placed in the *Goal* buffer will be accordingly modified:

```
actr.set_current_model(player1)
actr.define_chunks(['goal','isa', 'strategy', 'state','leave'])
actr.goal_focus('goal')
```

and the Procedural System of the model will activate the following rule:

```
(p leave-chocolate
=goal>
state leave
?vocal>
state   free
==>
+vocal>
cmd speak
string "leave"
=goal>
mode leave
!output! (I leave chocolate bar...)
)
```

The player is thus called to take on an important role in the course of the story. The goal is to be able to create an empathetic link between the end-user himself and the main character of the story he is trying to help. In this way, the user will be assigned a certain degree of responsibility during the evolution process of the story, which will end with a positive outcome or failure.

5. The Interactive Storytelling on NarRob

The ACT-R story characters were embodied in humanoid SoftBank Pepper and NAO robots, using the NarRob framework [18,19]. Each robot interprets a character of the story. In particular, the NAO robot performs the narrator's role, while the two Pepper robots play the bully and the victim roles, respectively. Each robot exploits a different vocal setting, obtained by modulating the pitch and speed of the speech, to interpret a specific character. Each robot is also able to add gestures and emotional expressions by analyzing the story content, using specific modules of the NarRob framework, and exploiting a labeled dataset of gestures. In particular, the content of the story is analyzed, and the text is automatically annotated with tags representing the gestures and the emotions that the robot must perform. The text annotation with gestures tags is obtained by using the Stanford CoreNLP tool. The text is first segmented in sentences, then there is a phase of words lemmatization, and finally, a Dependency-parsing procedure analyzes the syntactic structure of each sentence and compares the elements of the dependency graph with the annotations of the available gestures. The gestures are included in a manually annotated dataset composed of animations created by using both the SoftBank Choregraphe suite recording the postures obtained by using the robot as a puppet and an acquisition device (a Kinect camera) [18].

The emotional annotation of the story is obtained by using the Synesketch tool [35]. The emotions tags are then translated in robotic expressions, by modifying the color of the LEDs, the speed of its

speech, and the head inclination. The SALVE chatbot architecture [36] allows the robot to manage the interaction. Every time there is a situation where a critical choice has to be taken by the main character, the robot, playing at that moment the role of the moderator, asks for the human player's opinion by opening a selection window in the tablet screen, as shown in Figure 5. After that the child makes the selection, and the story evolves according to the reasoning processes of the characters, as described in Section 4.2.

Interaction Scenario

Figure 6 shows the interaction between a child and the storyteller robots. The experiment has been carried out in our laboratory to test the implemented system. The figure shows a girl sitting in front of three robots. The narrator, impersonated by the smallest robot, the NAO, has been placed on a table at the center to be more visible. The first Pepper robot from the left plays the role of the bully, while the second plays the role of the main character, i.e., the victim. The tablet on the second Pepper shows a form that is used by the child to introduce some information required by the robot. The child interacts directly with the story through the tablet, influencing the behavior of the victim. In the first phase, the narrator dialogues with the child to acquire initial data to be used during the interaction (in particular, it asks for the child's name). Then, the storytelling performance starts with the introduction of the characters and the narrative context by the narrator. The story emerges and evolves according to the cognitive models of the characters and their subsequent interaction. In what we called *turning points*, representing the points of the story where the protagonist is in trouble, the child selects an option by touching the tablet. Figure 7 shows an example of selection through Pepper's tablet. The narrator, who calls the baby by his first name, invites the child to give advice. This behavior of the narrator is aimed at obtaining a certain affiliation between the child and the robot. After evaluating two alternatives, the player selects his option, thus influencing the evolution of the narration.
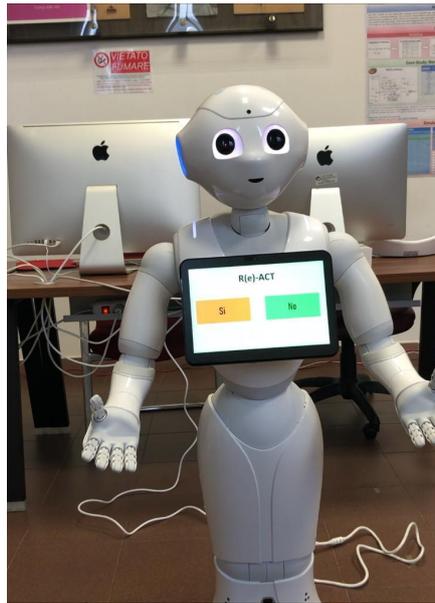

Figure 6. Human-Robot interaction.

Figure 7. Choices selection through the Pepper's tablet.

6. Evaluation

We carried out an evaluation of the system to understand whether the choices of modeling the characters of the story according to a cognitive architecture and the performances in the role of the characters by humanoid robots contributed to a better understanding of the story and a satisfactory involvement of the child.

We analyzed: (1) the impact of such modeling of the characters on the perception of their *credibility*, and (2) the effect of the implemented robotic narration on the storytelling *performance*.

We recorded two interactions of the same player with the system. In the first interaction (*video 1*), the characters (the victim and the bully), during the interaction with the other character, also explained the processes of their inner reasoning that led them to take decisions, and the reasons leading them to have doubts and ask for advises in case of significant conflict.

In the second interaction (*video 2*), the character's reasoning processes and the motivations behind the choices made during the narration were not explained. In this last case, therefore, the characters representing the victim and the bully only interacted with each other without giving any further explanation of their behavior.

The two videos were analyzed by forty users. Each user, after watching the two videos in a random order (to avoid an evaluation bias due to the sequence), filled in a survey. The survey is composed of two groups of questions, shown in Table 1 with their associated English translation.

The first group of five questions is aimed at assessing the credibility of the narrative and of the characters; the second group of three questions is aimed at assessing the overall interpretation of the robotic storytelling system. All the questions were evaluated on a 4-value Likert scale (*1 None, 2 Little, 3 Enough, 4 Much*). Finally, an open-ended question was included to assess any suggestions and criticisms as feedback to further refine the storytelling activity.

A dependent sample t-test was conducted to compare the credibility perceived by users after watching the movie 1 ($M = 17.66, SD = 2$) with that perceived after watching movie 2 ($M = 13.78, SD = 2.69$); $t(40) = 12.103, p = 0, d = 1.576$. These results suggest a large difference between the two conditions. Figure 8 shows the distribution of credibility values in response to the two different stimuli (movies). A dependent sample t-test was also conducted to compare the quality of interpretation perceived by users after watching movie 1 ($M = 10.88, SD = 1.44$) with movie 2 ($M = 9.73, SD = 1.52$); $t(40) = 5.6664, p = 0, d = 0.775$. These results suggest a medium difference

between the two conditions. Figure 9 shows the distribution of the interpretation values in response to the two different stimuli.

Table 1. Evaluation Survey.

	Per nulla	Poco	Abbastanza	Molto
It: Quanto ritieni naturale il comportamento dei personaggi? *En: How much genuine is the behavior of the characters?*				
It: Hai compreso le dinamiche della storia? *En: Do you understand the dynamics of the story?*				
It: Durante la storia hai compreso lo stato d'animo dei personaggi? *En: During the narrative, did you understand the characters' moods?*				
It: Quanto ritieni comprensibile il pensiero dei personaggi? *En: How understandable do you consider the thoughts of the characters?*				
It: Le emozioni dei personaggi sono comprensibili? *En: Are the characters' emotions understandable?*				
It: Come valuti la performance dei robot? *En: How do you judge the performance of the robots?*				
It: Pensi che dei robot siano adatti a narrare una storia? *En: Do you think that robots are appropriate for telling a story?*				
It: Pensi che la messa in scena della narrazione con dei robot sia attrattiva e coinvolgente? *En: Do you think the narrative robotic performance is attractive and engaging?*				

The results show in general how the narration through humanoid robots has been appreciated and that making explicit the internal state of the characters significantly has increased the perception of their credibility while it does not have any significant influence on the perception of their performance. Several comments, translated in what follows in English, inserted in the free answer question highlighted how the most appreciated aspect of the system is the explanation of the internal reasoning processes of the characters (the name Roberto in the comments refers to the bully character):

- *I appreciated that Roberto explains the reason for his behavior*
- *The episodes compared to the other version are faster, and they are not easy to understand. I prefer the version with the explanations of the characters*
- *The idea to let Roberto explain his actions is great, he seems even less bad*

Particularly interesting is the last comment, as the user shows empathy with the bully. Several more critical comments highlighted aspects that can be improved, in particular, those concerning the story dynamics and its duration, the number of scenes, and the interaction between the characters:

- *I liked the realistic interpretation of the robots and the help of the narrator, there's nothing I didn't appreciate, maybe I would improve the dialogue between the characters.*
- *I appreciated the narrator's explanation, I did not appreciate the immediate actions of the characters. I'd increase the length of the episodes.*
- *I appreciated the thematic and the characters in the school context, but I suggest to improve the story by inserting more dialogues and scenes.*

Some comments focused more on the emotional aspect of the characters:

- *The theme and the characters' performances are appropriate. I suggest to draw attention towards the suffering of the victim*
- *Interesting the idea of using robots, I'd improve the emotional aspects.*

Finally, the following comments are very interesting. After viewing the video 2, the user commented highlighting a lack of understanding:

- *I like the idea of robots playing the role of children in a school, although I can't easily understand neither the thoughts of the characters nor their emotions.*

  After watching the video 1 the same user commented as follows:

- *In this version, the story is more fluent, the characters are more realistic, and the story is more engaging. I suggest adding more dialogues.*
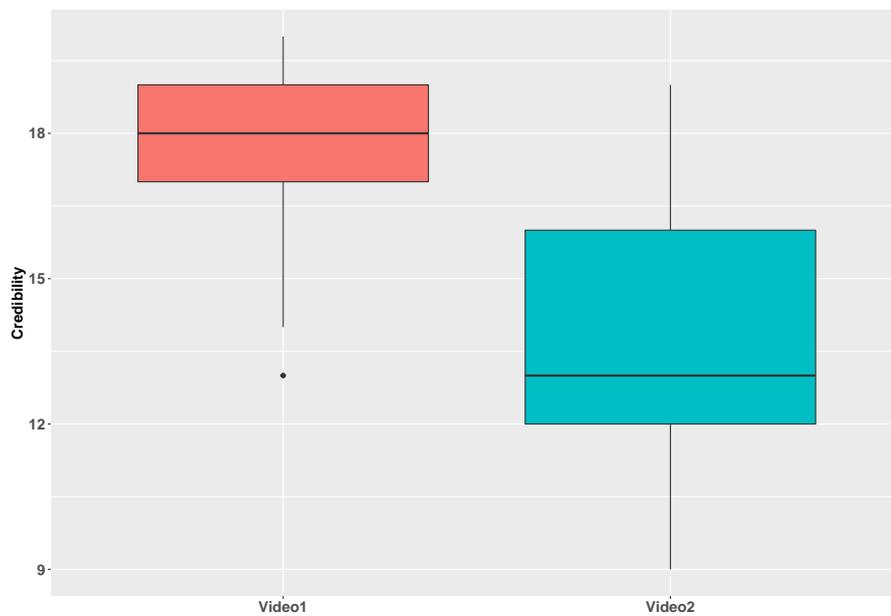

Figure 8. Distribution of credibility for the two kinds of interaction with the explanation of the internal reasoning of the characters (**left**) and without it (**right**).
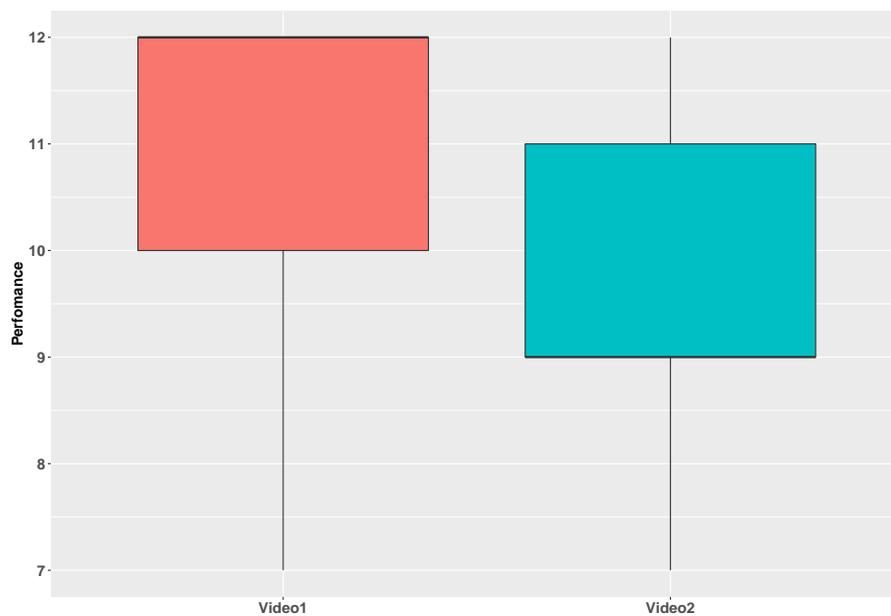

Figure 9. Distribution of performance for the two kinds of interaction with the explanation of the internal reasoning of the characters (**left**) and without it (**right**).

7. Conclusions

In this work, a story character modeling was performed by using the ACT-R cognitive architecture. The storytelling process is performed by NarRob, a robotic interactive storytelling system. Modeling the characters of the story through a cognitive architecture makes it possible to analyze the cognitive processes of the agents running in the environment, coinciding with a narrative context. By using the coordinated work of the modules that simulate the cognitive processes of the human brain, it was possible to make the characters entities able to understand, reason, listen and speak. In this preliminary phase of the work, each robotic actor is capable of understanding who is interacting with it, and it is also able to associate a semantic meaning to the exchanged sentences. The agent can deal with the situations that are presented to it, choosing from a set of available actions, coherently with the plot of the story.

As it was observed in the evaluation, the robots capture the attention and the curiosity of the players towards both the plot and the fate of the characters. The robots, during their storytelling, explained their internal reasoning processes, trying to stimulate the player's empathy.

The main limitations of the system emerged from its evaluation, are related to the little interaction between the characters of the story and a weak manifestation of their emotional state. For this reason, we will improve the system modeling the personal motivations and desires that lead the characters to make their decisions, giving importance to inner aspects of human beings such as the inclination to socialize, and their internal conflicts in balancing the compliance with the social norms, and the fulfilling of their desires. This choice would allow us to further explain the reasons behind the choice of some actions. We will also focus more on the social interactions among the characters of the story, considering the Social Practice model [36,37].
We will focus more on the agent's ability to freely choose among a large set of actions according to their behavioral knowledge, implemented through the flexible Social-Practices model.

By behavioral knowledge, we mean a set of information that a human being uses to act and interact with other human beings in a given social context. This step would also make it possible to generalize the writing of the rules, making them more flexible. As discussed, the system is conceived as serious storytelling; the aim of the narrative is not pure entertainment, but it has the primary purpose to make explicit some dynamics that may arise in social contexts and the effects of certain decisions better comprehensible.

In this phase, we focused on the evaluation more from an architectural and implementation point of view. Before testing the system with children in real serious game activity, it is important to understand if the design choices contribute to more credibility of the characters and if they bring a better understanding of the internal mechanisms. All these features must be taken into account to carry out a subsequent validation on children. The obtained results are encouraging. Therefore, the educational impact of the system will be evaluated through an experiment that we are developing and will be carried out in the upcoming months. The measurement of the effectiveness of this type of intervention will be achieved through the monitoring of the child/player's empathy toward the characters.

References

1. Mutlu, B.; Forlizzi, J.; Hodgins, J. A Storytelling Robot: Modeling and Evaluation of Human-like Gaze Behavior. In Proceedings of the 6th IEEE-RAS International Conference on Humanoid Robots, Genova, Italy, 4–6 December 2006.

2. Kory, J.; Breazeal, C. Storytelling with robots: Learning companions for preschool children's language development. In Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication, Edinburgh, UK, 25–29 August 2014.

3. Gelin, R.; d'Alessandro, C.; Le, Q.A.; Deroo, O.; Doukhan, D.; Martin, J.C.; Pelachaud, C.; Rilliard, A.; Rosset, S. Towards a Storytelling Humanoid Robot. In Proceedings of the AAAI Fall Symposium: Dialog with Robots, Arlington, VA, USA, 11–13 November 2010.

4. Striepe, H.; Lugrin, B. There Once Was a Robot Storyteller: Measuring the Effects of Emotion and Non-verbal Behaviour. In Proceedings of the International Conference on Social Robotics, Tsukuba, Japan, 22–24 November 2017.

5. Kennedy, J.; Baxter, P.; Senft, E.; Belpaeme, T. Social robot tutoring for child second language learning. In Proceedings of the 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Christchurch, New Zealand, 7–10 March 2016.

6. Hsu, S.H.; Chou, C.Y.; Chen, F.C.; Wang, Y.K.; Chan, T.W. An investigation of the differences between robot and virtual learning companions' influences on students' engagement. In Proceedings of the 2007 First IEEE International Workshop on Digital Game and Intelligent Toy Enhanced Learning (DIGITEL'07), Jhongli City, Taiwan, 26–28 March 2007.

7. Li, J. The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *Int. J. Hum.-Comput. Stud.* **2015**, *77*, 23–37. [CrossRef]

8. Leite, I.; McCoy, M.; Lohani, M.; Ullman, D.; Salomons, N.; Stokes, C.; Rivers, S.; Scassellati, B. Narratives with robots: The impact of interaction context and individual Differences on story recall and emotional Understanding. *Front. Rob. AI*, 2017. [CrossRef]

9. Chella, A.; Barone, R.; Pilato, G.; Sorbello, R. An Emotional Storyteller Robot. In Proceedings of the AAAI Spring Symposium: Emotion, Personality, and Social Behavior, Palo Alto, CA, USA, 26–28 March 2008.

10. Infantino, I.; Pilato, G.; Rizzo, R.; Vella, F. I feel blue: Robots and humans sharing color representation for emotional cognitive interaction. In *Biologically Inspired Cognitive Architectures 2012*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 161–166.

11. Belpaeme, T.; Kennedy, J.; Ramachandran, A.; Scassellati, B.; Tanaka, F. Social robots for education: A review. *Sci. Robot.* **2018**, *3*, eaat5954. [CrossRef]

12. Belpaeme, T.; Baxter, P.; De Greeff, J.; Kennedy, J.; Read, R.; Looije, R.; Neerincx, M.A.; Baroni, I.; Zelati, M.C. Child-robot interaction: Perspectives and challenges. In Proceedings of the International Conference on Social Robotics, Bristol, UK, 27–29 October 2013; pp. 452–459.

13. Kennedy, J.; Baxter, P.; Belpaeme, T. The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning. In Proceedings of the 2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Portland, OR, USA, 2–5 March 2015; pp. 67–74.

14. Bethel, C.L.; Henkel, Z.; Stives, K.; May, D.C.; Eakin, D.K.; Pilkinton, M.; Jones, A.; Stubbs-Richardson, M. Using robots to interview children about bullying: Lessons learned from an exploratory study. In Proceedings of the 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), New York, NY, USA, 26–31 August 2016.

15. Paradeda, R.B.; Martinho, C.; Paiva, A. Persuasion based on personality traits: Using a social robot as storyteller. In Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6–9 March 2017; pp. 367–368.

16. Conti, D.; Di Nuovo, A.; Cirsa, C.; Di Nuovo, S. A comparison of kindergarten storytelling by human and humanoid robot with different social behavior. In Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6–9 March 2017.

17. Conti, D.; Cirasa, C.; Di nuovo, S.; Di Nuovo, A. Robot, tell me a tale!: A social robot as tool for teachers in kindergarten. *Interact. Stud.* **2019**, *20*, 1–16.

18. Augello, A.; Pilato, G. An Annotated Corpus of Stories and Gestures for a Robotic Storyteller. In Proceedings of the 2019 Third IEEE International Conference on Robotic Computing (IRC), Naples, Italy, 25–27 February 2019; pp. 630–635.

19. Augello, A.; Infantino, I.; Maniscalco, U.; Pilato, G.; Vella, F. Introducing NarRob, a Robotic Storyteller. In Proceedings of the International Conference on Games and Learning Alliance, Palermo, Italy, 5–7 December 2018.

20. Maricchiolo, F.; Gnisci, A.; Bonaiuto, M.; Ficca, G. Effects of different types of hand gestures in persuasive speech on receivers' evaluations. *Lang. Cognitive Processes* **2009**, *24*, 239–266. [CrossRef]

21. Ham, J.; Cuijpers, R.H.; Cabibihan, J.J. Combining robotic persuasive strategies: The persuasive power of a storytelling robot that uses gazing and gestures. *Int. J. Social Rob.* **2015**, *7*, 479–487.

22. Xu, J.; Broekens, J.; Hindriks, K.; Neerincx, M.A. Effects of a robotic storyteller's moody gestures on storytelling perception. In Proceedings of the 2015 International Conference on Affective Computing and Intelligent Interaction (ACII), Xi'an, China, 21–24 September 2015.

23. Anderson, J.R.; Matessa, M.; Lebiere, C. ACT-R: A theory of higher level cognition and its relation to visual attention. *Hum.–Comput. Interact.* **1997**, *12*, 439–462. [CrossRef]

24. Roberts, D.L.; Narayanan, H.; Isbell, C.L. Learning to Influence Emotional Responses for Interactive Storytelling. In Proceedings of the AAAI Spring Symposium: Intelligent Narrative Technologies II, Palo Alto, CA, USA, 23–25 March 2009.

25. Mateas, M.; Stern, A. Architecture, authorial idioms and early observations of the interactive drama Façade. Available online: http://ra.adm.cs.cmu.edu/anon/usr0/ftp/usr/anon/2002/CMU-CS-02-198.pdf (accessed on 3 April 2020).

26. Mateas, M.; Stern, A. Writing Façade: A case study in procedural authorship. Available online: https://electronicbookreview.com/essay/writing-facade-a-case-study-in-procedural-authorship/ (accessed on 3 April 2020).

27. Evans, R.; Short, E. Versu—a simulationist storytelling system. *IEEE Trans. Comput. Intell. AI Games* **2013**, *6*, 113–130. [CrossRef]

28. "FearNot!". Available online: https://sourceforge.net/projects/fearnot/ (accessed on 10 December 2019).

29. Lieto, A.; Bhatt, M.; Oltramari, A.; Vernon, D. The role of cognitive architectures in general artificial intelligence. *Cognitive Syst. Res.* **2018**, *48*. [CrossRef]

30. Peter, M. DRAMA: Drama: Narrative pedagogy and socially challenged children. *Br. J. Spec. Educ.* **2009**, *36*, 9–17. [CrossRef]

31. Ying, K.T.; Sah, S.B.M.; Abdullah, M.H.L. Personalised avatar on social stories and digital storytelling: Fostering positive behavioural skills for children with autism spectrum disorder. In Proceedings of the 2016 4th International Conference on User Science and Engineering (i-USEr), Melaka, Malaysia, 23–25 August 2016.

32. Rhodes, C. Do Social Stories help to decrease disruptive behaviour in children with autistic spectrum disorders? A review of the published literature. *J. Intellect. Disabil.* **2014**, *18*, 35–50. [CrossRef]

33. Madej, K.S. "Traditional Narrative Structure"–not traditional so why the norm? In Proceedings of the 5th International Conference on Narrative and Interactive Learning Environments, Edinburgh, Scotland, 6–8 August 2008.

34. Rescorla, R.A.; Wagner, A.R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II Curr. Res. Theory* **1972**, *2*, 64–99.

35. Krcadinac, U.; Pasquier, P.; Jovanovic, J.; Devedzic, V. Synesketch: An open source library for sentence-based emotion recognition. *IEEE Trans. Affect. Comput.* **2013**, *4*, 312–325. [CrossRef]

36. Augello, A.; Gentile, M.; Dignum, F. Social agents for learning in virtual environments. In Proceedings of the International Conference on Games and Learning Alliance, Utrecht, The Netherlands, 5–7 December 2016.

37. Reckwitz, A. Toward a theory of social practices: A development in culturalist theorizing. *Eur. J. Soc. Theory* **2002**, *5*, 243–263. [CrossRef]