Interview with the Robot: Question-Guided Collaboration in a Storytelling System

Philipp Wicke
School of Computer Science
University College Dublin
Dublin D4, Ireland
philipp.wicke@ucdconnect.ie

Tony Veale
School of Computer Science
University College Dublin
Dublin D4, Ireland
tony.veale@ucd.ie

Abstract

An automated storytelling system that presents its stories as text on a screen is limited in its engagement with readers. An embodied robotic agent can overcome these limitations by using gesture to physically enact its stories. This paper presents a robotic agent that enacts its own novel stories, which it shapes from the feedback it obtains using probing personal questions. Our robotic writer/presenter has two alternate modes of story-generation: a straight "telling" mode and an interview-oriented back-and-forth that extracts personal experiences from the user as raw material for new stories. We explore the practical issues of implementing both modes on a NAO humanoid robot that integrates gestural capabilities in an existing story-telling system.

Introduction

An intriguing story that captivates listeners is a product of multiple intertwined factors, from the complexity of the plot to the way the storyteller uses speech and gesture to interact with an audience. Ultimately, however, it is the emotional nature of the human listener, such as the listener's willingness to get emotionally involved in a story and to empathize with its characters, that leads to true appreciation of a narrative. This empathy can be triggered by identification (Krebs 1975) with the characters in the story or with the perspective conveyed by the storyteller. This path from identification to empathy to engagement requires that the listener be human, but it does not require the teller to be human too. This paper provides a system description of a creative story-generator that augments its symbolic narratives with embodied gestures using a robot for knowledge elicitation and story presentation. It explores two alternate modes of engagement that vary in the amount of personal listener experience that is integrated into a tale to foster identification and empathy.

As we use the Nao humanoid robot in this work (Fig. 1), we will briefly comment on the technical details of this physical platform, before surveying previous creative work with the Nao. Storytelling with a robot, as presented here, will combine story-generation with story-enactment. The former generates stories that make maximal use of the robot's affordances, while the latter executes these affordances to maximize user engagement. Enactment combines not just speech (textual delivery) but gesture, posture and body orientation. In short, robotic agents provide more dimensions of expressiveness than text or speech alone.

The following sections focus on a discussion of the technical difficulties that we have encountered, and on the techniques and resources we have used to overcome them. To investigate our hypothesis that a robotic platform can enhance listener identification with the products of automated story-generation, we contrast the implementation of two modes of user-influenced story-generation. The first, a baseline, employs simple user interaction prior the generation of a story, while the other conducts a probing, therapist-like interview to elicit personal experiences from the listener that it can repackage – somewhat collaboratively – into novel stories.

An empirical evaluation of the approaches employed is still in progress, so for now we conclude with insights from the current implementation to explain why we identify more with an embodied robot than with an intangible piece of software, and perhaps why are more likely to attribute creativity to the former than to the latter.

The Humanoid Nao

The French robotics company Aldebaran Robotics (recently acquired by Softbank) began to design its Nao humanoid robot (Fig.1) in 2004. Four years later the first release was provided for research use in universities and laboratories. The current model has a height of 58 cm, weighs 4.3 Kg and is powered by a 48.6 Wh lithium battery that provides approx. 90 minutes of active use. The bipedal robot has 25 degrees of freedom controlled by a 1.6 GHz CPU. The inbuilt Linux-based *NAOqi* OS gives access to two HD cameras, four microphones and a variety of sensors for detecting inertia, pressure and infrared light (Gouaillier et al. 2009). A detailed investigation of the robot's technical capabilities can be found in (Shamsuddin et al. 2011). With its wide variety of easily accessible tools, the robot quickly found its way into research (Tapus et al. 2012), (Parde et al. 2015) and education (Hood, Lemaignan, and Dillenbourg 2015). The robot provides in a ready-made set of modalities: Speech production and speech recognition software, facial recognition that allows the robot to follow faces with its gaze, a pre-installed set of 400+ body movements and although the robot has no facial gesture capabilities it has a set of RGB LEDs that function as eyes, which can show different colors to communicate emotions. The proceeding section will

Figure 1: The Nao robot from Aldebaran. In this image, the robot performs a gesture titled *ShowSky_1* that has been associated (with medium strength) to the actions: look up, praise, pray, preach and worship.

comment on studies investigating those modalities. The use of gestures is the key modality in the storytelling framework presented in this paper and we will therefore outline the importance of this modality through previous research.

The Nao's uses are myriad, so this section will highlight the most notable research that has been published in the domain of storytelling. An overview of storytelling with robots can be found in (Chen, Nurkhamid, and Wang 2011). For instance, (Ham et al. 2011) investigated the impact of deliberate gesture and gaze during storytelling by a Nao robot. They employed a database of gestures and gazing modes that were sourced from a professional stage actor and their results showed that the effect of using gesture and gaze together was significantly greater than the effect of using a single modality in isolation. The Nao has a gazing mode that is active by default, which causes it to autonomously shift its focus from one person to another, often to the person who happens to be speaking at any given moment. We can assume, based on (Ham et al. 2011), that this default gaze-shifting functionality enhances the communicative effect of its speech mode, regardless of whether other gestures are also used. Those authors relied on a small dataset of just 21 gestures, whereas the framework we present here employs more than 400 gestures, each of which has been associated with one or more story verbs.

The current framework emphasizes not just gesture and gaze but the use of language to elicit personal experiences from a user. We draw on studies from (Csapo et al. 2012), (Meena, Jokinen, and Wilcock 2012) and (Wilcock and Jokinen 2013) that use the Nao in question-answering mode. The framework described in that work marries Nao's dialogue capabilities with a question-answering interface for Wikipedia that allows a user to retrieve information from the web in a conversational manner (in much the same vein as Amazon's Alexa/Echo). The evaluation of human/robot interaction in (Csapo et al. 2012) focuses on modalities such as tactile sensors, face detection and non-verbal cues to derive these findings: the optimal communication distance is 0.9 meters between human and robot, while the limitations of Nao's speech recognizer make sudden interruptions inadvisable and highly impractical to process. The latter insight influences the constraints we describe in a later section on the modes of our own framework. Findings regarding the use of gesture are less relevant here, since those authors relied on a database of just six gestures for their system.

Gesture and gaze enrich the communication process, but how much accuracy do these non-verbal additions provide? (Häring, Bee, and André 2011) suggests that the Nao robot's body movements are most effective for communicating a specific set of emotions, whilst eye color and sound are evocative but much less accurate in conveying specifics. This leads us to prioritize physical gestures in our approach, while relying on colour and sound as non-vital embellishments. As each of the studies mentioned here suggests that gestures achieve a heightened effect when paired with the relevant speech content, we do not need to focus our evaluation on behavioral cues, but on the communicative aspects of the interaction with the user.

A comparable number of gestures to that employed here can be found in the work of (Pelachaud et al. 2010), (Le, Hanoune, and Pelachaud 2011) and (Gelin et al. 2010). Those authors developed two markup languages for use in functional and behavioral annotation for story-telling with a virtual animated avatar. They subsequently selected a subset of their database of approximately 500 annotated gestures for use in embodied storytelling with a Nao robot that can read stories to children. The empirical evaluation of their approach indicates that while the gestures are reported as appropriate to the content they adorn, they are not often seen as natural adornments to the action (Le and Pelachaud 2012).

Most approaches to automated story-generation lend themselves to robotic embodiment, insofar as any stream of action-oriented text can be augmented at suitable junctures with appropriate gestures and gaze behaviours. Thus, the engagement-reflection approach of (Pérez y Pérez and Sharples 2001), as implemented in the Mexica system, is as suited to the enactive mode of story presentation as the morphological approach (in the sense of Vladimir Propp's *morphology of the folk tale*) of (Gervás 2013) to the plot-as-planning-and-problem-solving approach of (Riedl and Young 2010). So we are principally guided by practicality rather than theory in our choice of the *Scéalextric* model of (Veale 2017) as the story-generating core of our system. *Scéalextric* employs an open and modular knowledge representation that is easily extended, and provides a public distribution that contains tens of thousands of story-related semantic triples to support plot and character design. It is built around an inventory of over 800 action verbs that gives our robot the semantic material to pose highly specific questions of a listener in its path to building a vivid story.

Some fascinating recent work in Computational Creativity has focused on humans and artificial agents working in unison to achieve *co-creativity* (Jordanous 2017). For example, (Davis et al. 2016) explored co-creativity in the domain of abstract drawing, using an enactment framework to identify the emergence of *sense-making* in a contrastive study of human-human *and* human-machine collaboration. A co-creative approach to storytelling is found in the *Mable* system of (Singh, Ackerman, and Pérez y Pérez 2017), which builds on the *Mexica* story generation system to write lyrics for a ballad that tells a tale. This system also builds on the *Alysia* system of (Ackerman and Loker 2017), which uses machine-learning to support the creation of melodies. The combined system first composes lyrics in a co-creative mode with a human user, and subsequently overlays these lyrics onto a machine-generated melody. Good music is designed to move the emotions *and* the body, so musical story-telling presents significant opportunities for physical embodiment in a robot. So the approach described here is one of many potential story-generation services (in the sense of (Veale 2013)) that can be selected from a competitive API economy (Concepción, Gervás, and Méndez 2017) for automated storytelling on demand. Another API service in this economy, the *Charade* system of (Méndez, Gervás, and León 2016), suggests obvious parallels to our current framework, insofar as it motivates the development of inter-character affinities. Each service will advertize its own comparative advantage, and so the current approach offers nuance in its use of enactment to unify vivid actions with embodied gestures.

The body is the means by which we engage with our physical environment, so an embodied story-teller whose gestures appear natural can grant a greater sense of reality to the wholly invented realms of its imaginary stories. Each of the studies considered here thus emphasizes the importance of natural gesture to the enactment of a tale, or in the words of Hollywood screenwriters, to *showing, not just telling*. It is to the practical implementation of this maxim with natural and expressive gestures that we now turn.

Embodied, Enactive Storytelling

The scientific community lacks agreement on a single cognitive model for explaining the processes of creative generation. The Brain Computer metaphor, most prominently described by (Putnam 1961), offers a premise that most cognitive engineers take for granted. This metaphor of Computationalism regards the brain as the underlying hardware, just like the hardware of a computer, and the mind as the software that runs on this physical platform. But to what extent does the hardware shape the software, or vice versa, and how is the synergy between mind and body achieved? While not attempting to resolve these vexing long-standing questions, this paper explores an intersection that is neglected in most modeling approaches. Collaborative enactive storytelling is the application of creative software that relies crucially on the physicality of its hardware, thus blurring the boundaries between external interactions and internal representations. So we adhere to a new theory in the philosophy of cognition, called Enactivism, that challenges Computationalism in suggesting that a precursor to high-level cognition is the dynamic interaction of an active organism within its physical environment. This environment is modeled internally not by translating sensory input into internal representations, but through exploratory interactions that create meaning (Di Paolo, Rohde, and De Jaegher 2010).

By analogy with these biological organisms, a robot can operate within, modify and learn from its environment (see, for instance, (Sandini, Metta, and Vernon 2007)). The robotic system becomes producer and product at the same time via the establishment of an *autopoietic feedback loop*. In such a feedback loop, the execution of a specific behavior by an actor for a spectator may trigger a corresponding response from the spectator, which in turn influences the actor's subsequent actions and shapes the actor's overall behavior (an explanation is given in (Fischer-Lichte 2012)). The most influential behavior of a storyteller is the use of the spoken word, but apt choices of physical gesture can also greatly contribute to the creation of an effective feedback loop (see (McNeill 1985), (Bergen, Narayan, and Feldman 2003)). As the interface between one's internal representations and the external environment, the body can make use of gestures to express what speech alone cannot convey. The following section outlines an implemented robot framework that provides two different modes of enactive storytelling.

Framework Description

The framework that is used for the two storytelling modes has been developed using *Python* and the *NAOqi* (Version 2.1.4.13) package. This software package provided by *Softbank Robotics* allows easy access to the different modalities of the robot's hardware. This section briefly describes the most important modules that have been used, databases that have been integrated for the story-generation process and the solution to problems that have been encountered during the setup of the framework. The databases used to craft the stories and the questions to create the interview-shaped mode will be made publicly available.

Automated Storytelling The framework is built upon the *Scéalextric* system for automated story-generation (Veale 2016a), as this provides a dense forest of plot possibilities for our robot to explore, perhaps in collaboration with a user. *Scéalextric*'s rich databases of symbolic representations allow actions to be bound together by causal connections and characters to be bound to actions on the basis of their established qualities. The system binds individual plot actions into plot segments (or arcs) with the following two-character triplet shape:

1. X *action* Y
2. Y *reaction* X
3. X *re-reaction* Y

The *Scéalextric* system provides over 3000 plot segments of this kind, made from causally-appropriate triplets that range over 800 different plot verbs (see (Veale 2017)). The resulting story space is modeled as a forest of trees in which each vertex is a plot verb, and in which every random walk yields a causally-coherent plot. Here is an example of a traversal through the forest:

A learns from B $\longrightarrow$ *A is inspired by B* $\longrightarrow$ *A falls in love with B* $\longrightarrow$ *A sleeps with B* $\longrightarrow$ *B fails to impress A* $\longrightarrow$ *A is disillusioned by B* $\longrightarrow$ *A breaks with B*

A robot story-teller can choose to explore the story forest using random walks, or it can elicit personal experience from the listener to guide its traversals. The precise strategy depends on which of its two modes the robot is operating within. In either case, the branching structure of the story forest provides choice (in random walk mode) and apposite junctures at which to question the user (in interview mode).

Gestures and body movement The current framework differs significantly from past efforts to exploit gestures on the Nao, drawing as it does from a set of 400+ predefined gestures from Aldebaran. We handcrafted annotations for each gesture, or physical behaviour, with one or more *Scéalextric* plot verbs. Each association of a gesture to a verb is also marked as strong, medium or weak according to our judgment (e.g., see Fig. 1). In all, 68% of the 800+ plot verbs in *Scéalextric* are associated with at least one gesture.

Understanding the Robot The current framework uses the Nao's *AnimatedSpeech* module to enrich its rendering of text as speech. Each sentence of a story is preprocessed prior its output to enhance its comprehensibility, e.g., by increasing of volume to the maximum, or lowering the voice pitch to simulate a more mature voice, or slowing the rate of articulation to increase understandability, or shortening of the pause between sentences to yield greater momentum in the telling. Unfortunately, the mechanical joints of a gesticulating Nao create noise that competes with the robot's speech, and in a non-laboratory environment this can impede story comprehension. We have thus introduced a super-title feature that echoes the output of the speech module on a large screen, so that an audience can follow the story in an additional modality (see e.g., Fig. 2).

Understanding the User The framework uses the Nao's *SpeechRecognition* module to communicate with the user. This module is primed with a vocabulary of words to which the robot should react via a built-in *word spotting* option. The vocabulary can be pre-loaded with thousands of words, yet the greater its size the more likely it is to confuse similar-sounding words. In some contexts we disable word spotting and require the user to reply with a single word response when explicitly prompted. As we shall see, such constraints need not impact the naturalness of a man-machine dialogue if the interactions are well-engineered and suitably framed.

Baseline Mode

The baseline mode of interactive storytelling employs minimum engagement with the user, but explores the same story forest and exploits the same gestural possibilities as other modes. Story-telling is initiated in this mode with a request

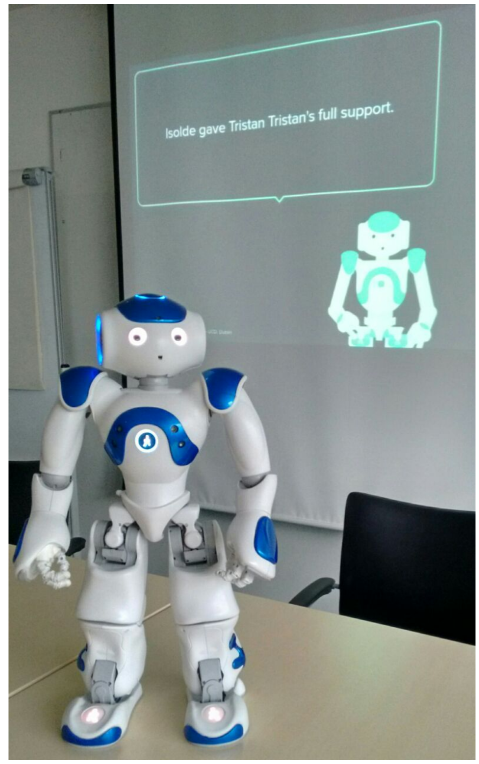

Figure 2: Demonstration of the storytelling robot with super-titles on the screen to increase comprehensibility.

to the user: "please provide an action around which to build a story." Any of *Scéalextric*'s 800+ action verbs may be offered in response by the user, as the Nao's speech recognizer is primed with the corresponding words. Low-level engineering challenges include loading this vocabulary in a parallel thread and preventing robotic stutters during the loading phase.

A vocabulary of 800+ verbs diminishes the reliability with which the robot can correctly distinguish words, as e.g. the words "look" may be confused for "cook" or "pay" for "pray." Fortunately the Nao's acoustic module reports a confidence value for each word that it recognizes. Only when this confidence is above an upper threshold (0.65) does the robot accept the user's response without question. When the confidence falls below a lower threshold (0.4) the robot remains in listening mode; only when the confidence falls between thresholds does the robot signal its uncertainty and seek explicit user confirmation with a "yes" or "no."

Once the robot has obtained an action from the user, story-generation around that pivot can proceed. Building on the *Flux Capacitor* representations of (Veale 2014), the robot selects its start and end points in the story forest to represent a meaningful character-development arc. Stories with the preferred number of actions (e.g., 6 to 10) are then generated by traversing the story forest between these end-points and retaining only those pathways/plots that contain the desired action. Of these matching pathways, the robot selects one that provides maximal opportunity for gestural expression, i.e. the story that contains the most verbs with gestural opportunities. Consider this complete example of a generated story using the keyword "*intimidate*":

- `Nao`: This story is about Isolde the Loyalist and Tristan the Despot.
- `Nao`: Isolde swore loyalty to trustworthy Tristan.
- `Nao`: Isolde gave Tristan Tristan's full support.
- `Nao`: But Tristan took full advantage of impressionable Isolde.
- `Nao`: Tristan intimidated others with threats of violence from Isolde.
- `Nao`: Thuggish Tristan threatened Isolde with violence.
- `Nao`: Isolde considered Tristan a disgusting monstrosity
- `Nao`: so Isolde's feelings for trustful Tristan soured.
- `Nao`: As a result, Isolde sold out Tristan to Tristan's enemies.

Each story begins with an introduction of a pair of characters, which have been selected from a comprehensive database of familiar faces, real and fictional, called the NOC list (see (Veale 2016b); access the list at `github.com/prosecconetwork`). The sequence of actions that have been selected from the story-graph are rendered with apt logical connectives (*and, so, but*, etc.) to convey the teller's understanding of causality.

In this baseline mode, the user is given only one opportunity – a single action – to influence the direction of a story. An example video of this mode can be found at: `bit.ly/2GGORhc`. More opportunities for collaborative development and self-expression are offered by the interview-shaped interaction of *Therapist mode*, as we'll see next.

Therapist Mode

We name this mode *Therapist mode* because it frames its interactions with the user around a series of probing questions about the user's own life experiences. These questions mirror the possibilities of the robot's story space because they are built from an interrogative rendering of the same plot actions. The guiding intuition is that everyone has a story in them; the role of a good therapist is to bring this tale to light by collaboratively reassembling the user's own experiences.

The story's actions should reflect the individual experiences while perhaps changing the sequence and filling gaps, thereby offering a new perspective on the user's past behavior. Potential story actions are rendered as yes-no questions (so, e.g., $are\_inspired\_by$ is rendered as "Have you ever been inspired by someone?"), and an affirmative answer adds that action to the story's plot. Consider this example:

1. `Nao`: Let us craft a story. I will start to ask you some questions. You only have to answer with yes or no.
2. `Nao`: Have you ever felt inspired by someone?
3. `User`: Yes.
4. `Nao`: But have you ever taken a hard look at someone who inspired you?
5. `User`: Correct.
6. `Nao`: Have you ever been criticized by that person?
7. `User`: No.
8. `Nao`: Can you imagine this is possible?
9. `User`: No.
10. `Nao`: Have you ever felt censured by that person?

The initial story verb is chosen at random from the robot's repertoire. If an affirmative response is given, the verb provides the next action in the story, and the potential follow-ons to that action suggest the next set of questions. Consequently, each of those interviews can vary in length, i.e. two or three positive replies might suffice to generate a story whereas negative answers will make the interview longer as the story tree needs to be explored further. If a negative response is given, and asserted again, the robot jumps to another branch in the story forest, and tries to move forward from there. We want the robot's stories to build on the user's experiences but to expand upon them too, to suggest what might have happened if events had taken a different turn. Thus, in line 6 the robot asks whether the user was ever criticized by a role model. Given a negative response, the robot presses on, asking instead if the user can conceive of this possibility. If the user now replies affirmatively, the story can incorporate this sequence of events (inspiration followed by criticism). To further allow for stories that go beyond the specific facts, users are also encouraged to reply with "maybe." As the questions asked of the user will differ from session to session, a different – yet highly personal – story will result each time.

Each story involves two characters, a protagonist (the user, or "you") and an antagonist. Notice how the questions above relate each action back to the previous action by assuming the antagonist to be common to both. It is likely that the user will have multiple antagonists in mind when answering questions, and the antagonist presumed in answer 3 is not the antagonist presumed by answer 13. The questions are sufficiently general to allow this artistic license to operate, so that the robot can weave stories that conflate several people from the user's life into a single thought-provoking antagonist. Once the session ends at the user's request, the selected actions can now be woven into a two-character plot:

> This is the story about you and a pioneer. This spectacular pioneer became a shining inspiration for you. You kept the pioneer under close observation. You mimicked the popular pioneer's style and adopted it as your own. "You've let me down" said the Pioneer plaintively, so the domineering pioneer gave you a very public rap on the knuckles, and to conclude, the pioneer brought suit against you in open court. That's the end of your story.

This story features all of the actions that the user has assented to, and may include additional actions as well to conclude the protagonist's arc (in the sense of (Veale 2014)) and bring the tale to a satisfying conclusion. Notably, the protagonist of the story is addressed as 'you', while the protago-

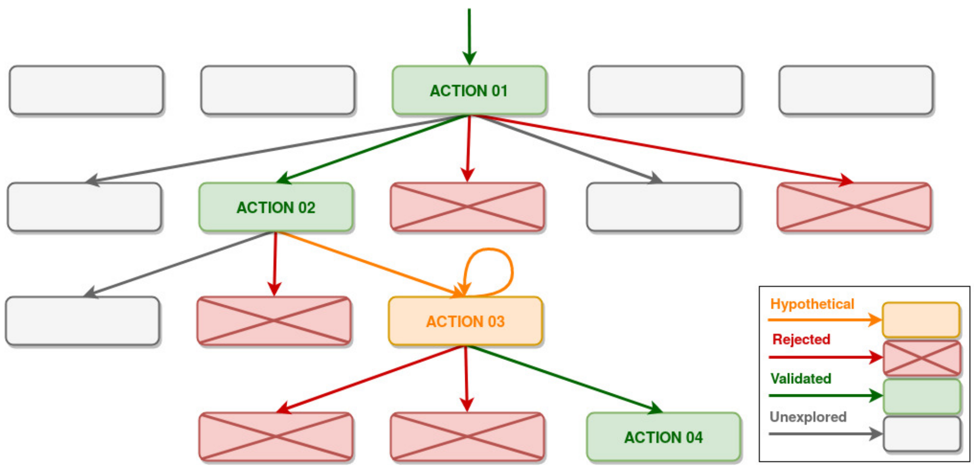

Figure 3: Example of a tree traversal for the story generation process. Grey nodes and edges indicate that a path has not been explored by the system. Red nodes and edges denotes those that have been answered with a 'No' by the user. Green nodes have been validated by the user, giving the sequence *ACTION 01*, *ACTION 02*, *ACTION 03* and *ACTION 04* that start the story story. A yellow node is one that has been rejected by a direct question, but accepted by the user to be a hypothetical possibility.

nist is referred to not by name but by character-type (here, the *pioneer*, as this is the kind of character that *inspires* others). The resulting story diverges from the user's own experiences, but in doing so sheds new light on them.

A schematic view of the generation-by-interrogation process is provided in Figure 3. Here we can see that the first selected action (`ACTION 01`) is translated into a question form and posed to the user by the robot. The node is selected (green) if the user validates this experience, or assents to the hypothetical, and this selection cues the node's children as possible follow-ons. However, if the user negates the question *and* its rephrasing as a hypothetical, the node and its action are blacklisted so that it cannot be asked again of the user. Instead, the robot selects another sibling and another branch to explore. In the example the node `ACTION 03` has been validated as a meaningful hypothetical by the user.

This mode differs significantly from the baseline in the amount of interaction it demands from the user. However, this interaction is not deterministic, and the user's answers merely suggest, rather than dictate, the robot's path through the story forest. This mode demonstrates that every story can offer a probing interrogation of one's own experiences, and vice versa whilst creating a feedback loop. An example video of this mode can be found at: `bit.ly/2oouZbY`. Our work up to this point has focused on development of the pilot system, but in the next section we comment on the planned evaluation of both modes.

Evaluation Challenges A crowd-based evaluation of the *Scéalextric* model of story-generation has been reported in (Veale 2017). In that work, two modes of story-generation were evaluated. Each mode employed the same plotting mechanism – a traversal of the story forest and subsequent rendering of plot actions into connected linguistic utterances – but each relied on a different model of characterization. In the generic condition, plots were instantiated with animal protagonists and antagonists, such as "the monkey" and "the snake," that were chosen at random from a list of Aesop-style animals. In the more elaborate familiarity condition, established characters such as Darth Vader or Donald Trump were chosen from the NOC list, and aspects of their characters as retrieved from the NOC were integrated into the rendering of plot verbs. Moreover, characters were chosen randomly in apt pairs, so that stories would pit Donald Trump against Lex Luthor, or Darth Vader against Bane, or Steve Jobs against Leonardo Da Vinci. The crowd-based evaluation of fifty stories from each mode solicited 10 ratings for six dimensions for each story, and the NOC-based mode showed superior results for all dimensions. Surprisingly, the NOC stories scored higher for dramatic content too, even though the underlying plots relied on precisely the same plotting mechanism as the generic stories. Consequently, we must take this bias into account comparing the two modes and eventually disable the NOC characters in order to compare the baseline with the therapist mode.

Once again we find ourselves with two modes of story-generation to evaluate. In the earlier crowd-based evaluation of (Veale 2017), judges were presented with pre-generated stories and asked to rate them after-the-fact. However, as the current stories are generated in cooperation with the user, and rely crucially on the user's input (as well as an appreciation of their own past actions), these stories are far less amenable to a simple crowd-based evaluation. Moreover, our evaluation should allow us to test the central research question of the current work:

> To what extent can an interview-style collaboration enrich the storytelling experience between human and machine over an approach with minimal interaction?

Since we further hypothesize that the embodied, gestural behaviour of the robot will form a significant part of any such enrichment, it behooves us to present the robot's stories *live*, in the physical presence of the robot. While video-taped sessions are one possibility when the goal is to evaluate the impact of gestures, prerecorded video does not support real-time interactivity *and* gesture in the same test. As the system moves out of the pilot stage, we hope to evaluate the system with a live audience, most likely in an educational setting where participants can be given partial credit for their feedback. For now, we aim to gain practical insights into possible evaluations in show-and-tell sessions at conferences.

Philosophical Investigations

Symbol-grounding remains a vexing problem for modern AI systems, and especially for those that rely on wholly symbolic representations. How do the symbols of a representation relate to the things in the world for which they are supposed to stand (Searle 1980)? So, for instance, how does the representation of a given plot verb in *Scéalextric* relate to the intuitive understanding of this verb as held by the human audience for a story that uses this verb? In a weak sense, the elements of a symbolic representation can mutually ground each other if they are connected in inferentially useful ways. Thus, insofar as the verbs in *Scéalextric*'s story forest are connected to each other in causally significant ways that are appropriately marked, the symbols for these verbs are weakly grounded in an extensional model of causality.

As system builders, our goal is not a theoretical grounding of symbols but the practical use of symbols in ways that *appear* grounded, whatever the philosophical truth may be. In this respect, the embodiment of a story-telling system in a physical agent capable of nuanced gestures goes a long way toward selling the appearance of grounding. But in addition to physical grounding, it is also meaningful to speak of *psychological* grounding. Does a system employ its cultural symbols in ways that respect the psychological attitudes that native speakers ascribe to those symbols? This is a question that goes beyond purely plot-related issues. Rather, it goes to how humans "feel" about specific symbols. We argue that framing the story-generation process as a therapeutic interview is just as effective at achieving psychological grounding as it is at achieving meaningful man-machine collaboration. Ultimately, the robot does not succeed in grounding its symbols in any way that would satisfy (Searle 1980), but it does come close to reflecting the grounding possessed by the human audience, whatever that might be.

Being There: Identity and Empathy in Storytelling

A study by (Pop et al. 2013) has demonstrated that the use of a social robot to support interaction in story-telling is more effective than the use of text on a screen. For when identification with the teller is the goal, the physical presence of a moving body with a human shape makes all the difference. While reading text on the screen is a solitary activity, a shared gaze with two human-like eyes can keep us focused and engaged during an interaction. Moreover, a study by (Seo et al. 2015) has shown that people tend toward greater empathy with a physical robot than with a simulated on-screen avatar. This heightens the contribution of physical modalities such as gesture, gaze and speech to the identification of listener with speaker and of human with machine. A listener that can identify with the storyteller is better positioned to empathize with the story that the teller wants to convey, especially when that story is crafted from the life experiences of the listeners themselves. Future studies are needed to investigate the empirical effect of a physically embodied and psychologically-grounded robot story teller that invents and delivers its *own* stories to users. While it is still at an early stage of development, we believe that *Scéalextric-NAO* is an important step in the right direction.

Acknowledgments

We would like to thank Stefan Riegl for his contribution to implementation of the storytelling system as outlined here.

References

Ackerman, M., and Loker, D. 2017. Algorithmic songwriting with alysia. In *International Conference on Evolutionary and Biologically Inspired Music and Art*, 1–16. Springer.

Bergen, B.; Narayan, S.; and Feldman, J. 2003. Embodied verbal semantics: Evidence from an image-verb matching task. In *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society*, 139–144.

Chen, G.-D.; Nurkhamid; and Wang, C.-Y. e. a. 2011. A survey on storytelling with robots. In *International Conference on Technologies for E-Learning and Digital Entertainment*, 450–456. Springer.

Concepción, E.; Gervás, P.; and Méndez, G. 2017. An api-based approach to co-creation in automatic storytelling. In *6th International Workshop on Computational Creativity, Concept Invention, and General Intelligence. C3GI*.

Csapo, A.; Gilmartin, E.; Grizou, J.; Han, J.; Meena, R.; Anastasiou, D.; Jokinen, K.; and Wilcock, G. 2012. Multimodal conversational interaction with a humanoid robot. In *Cognitive Infocommunications (CogInfoCom), 2012 IEEE 3rd International Conference on*, 667–672. IEEE.

Davis, N.; Hsiao, C.-P.; Yashraj Singh, K.; Li, L.; and Magerko, B. 2016. Empirically studying participatory sense-making in abstract drawing with a co-creative cognitive agent. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, 196–207. ACM.

Di Paolo, E.; Rohde, M.; and De Jaegher, H. 2010. Horizons for the enactive mind: Values, social interaction, and play. *Enaction: Towards a new paradigm for cognitive science*.

Fischer-Lichte, E. 2012. *Ästhetik des Performativen*. Suhrkamp Verlag.

Gelin, R.; d'Alessandro, C.; Le, Q. A.; Deroo, O.; Doukhan, D.; Martin, J.; Pelachaud, C.; Rilliard, A.; and Rosset, S. 2010. Towards a storytelling humanoid robot. In *AAAI Fall Symposium: Dialog with Robots*.

Gervás, P. 2013. Propp's morphology of the folk tale as a grammar for generation. In *OASIcs-OpenAccess Series in Informatics*, volume 32. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Gouaillier, D.; Hugel, V.; Blazevic, P.; Kilner, C.; Monceaux, J.; Lafourcade, P.; Marnier, B.; Serre, J.; and Maisonnier, B. 2009. Mechatronic design of nao humanoid. In

Robotics and Automation, 2009. ICRA'09. IEEE International Conference on, 769–774. IEEE.

Ham, J.; Bokhorst, R.; Cuijpers, R.; van der Pol, D.; and Cabibihan, J. 2011. Making robots persuasive: the influence of combining persuasive strategies (gazing and gestures) by a storytelling robot on its persuasive power. In *International conference on social robotics*, 71–83. Springer.

Häring, M.; Bee, N.; and André, E. 2011. Creation and evaluation of emotion expression with body movement, sound and eye color for humanoid robots. In *Ro-Man, 2011 Ieee*, 204–209. IEEE.

Hood, D.; Lemaignan, S.; and Dillenbourg, P. 2015. When children teach a robot to write: An autonomous teachable humanoid which uses simulated handwriting. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 83–90. ACM.

Jordanous, A. 2017. Co-creativity and perceptions of computational agents in co-creativity. In *Proceedings of the Eighth International Conference on Computational Creativity, Atlanta, US. ACC*.

Krebs, D. 1975. Empathy and altruism. *Journal of personality and social psychology* 32(6):1134.

Le, Q. A., and Pelachaud, C. 2012. Evaluating an expressive gesture model for a humanoid robot: Experimental results. In *Submitted to 8th ACM/IEEE International Conference on Human-Robot Interaction*.

Le, Q. A.; Hanoune, S.; and Pelachaud, C. 2011. Design and implementation of an expressive gesture model for a humanoid robot. In *Humanoid Robots (Humanoids), 2011 11th IEEE-RAS International Conference on*, 134–140. IEEE.

McNeill, D. 1985. So you think gestures are nonverbal? *Psychological review* 92(3):350.

Meena, R.; Jokinen, K.; and Wilcock, G. 2012. Integration of gestures and speech in human-robot interaction. In *Cognitive Infocommunications (CogInfoCom), 2012 IEEE 3rd International Conference on*, 673–678. IEEE.

Méndez, G.; Gervás, P.; and León, C. 2016. On the use of character affinities for story plot generation. In *Knowledge, Information and Creativity Support Systems*. Springer. 211–225.

Parde, N.; Hair, A.; Papakostas, M.; Tsiakas, K.; Dagioglou, M.; Karkaletsis, V.; and Nielsen, R. D. 2015. Grounding the meaning of words through vision and interactive gameplay. In *IJCAI*, 1895–1901.

Pelachaud, C.; Gelin, R.; Martin, J.; and Le, Q. A. 2010. Expressive gestures displayed by a humanoid robot during a storytelling application. *New Frontiers in Human-Robot Interaction (AISB), Leicester, GB*.

Pérez y Pérez, R., and Sharples, M. 2001. Mexica: A computer model of a cognitive account of creative writing. *Journal of Experimental & Theoretical Artificial Intelligence* 13(2):119–139.

Pop, C. A.; Simut, R.; Pintea, S.; Saldien, J.; Rusu, A.; Vanderfaeillie, J.; David, D.; Lefeber, D.; and Vanderborght, B. 2013. Social robots vs. computer display: does the way social stories are delivered make a difference for their effectiveness on asd children? *Journal of Educational Computing Research* 49(3):381–401.

Putnam, H. 1961. Brain and behavior, originally read as part of the program of the american association for the advancement of science. *Section L (History and Philosophy of Science)*.

Riedl, M., and Young, R. 2010. Narrative planning: Balancing plot and character. *Journal of Artificial Intelligence Research* 39:217–268.

Sandini, G.; Metta, G.; and Vernon, D. 2007. The icub cognitive humanoid robot: An open-system research platform for enactive cognition. In *50 years of artificial intelligence*. Springer. 358–369.

Searle, J. 1980. Minds, brains, and programs. *Behavioral and brain sciences* 3(3):417–424.

Seo, S.; Geiskkovitch, D.; Nakane, M.; King, C.; and Young, J. 2015. Poor thing! would you feel sorry for a simulated robot?: A comparison of empathy toward a physical and a simulated robot. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 125–132. ACM.

Shamsuddin, S.; Ismail, L.; Yussof, H.; Zahari, N.; Bahari, S.; Hashim, H.; and Jaffar, A. 2011. Humanoid robot nao: Review of control and motion exploration. In *Control System, Computing and Engineering (ICCSCE), 2011 IEEE International Conference on*, 511–516. IEEE.

Singh, D.; Ackerman, M.; and Pérez y Pérez, R. 2017. A ballad of the mexicas: Automated lyrical narrative writing. In *Eighth International Conference on Computational Creativity, ICCC, Atlanta*.

Tapus, A.; Peca, A.; Aly, A.; Pop, C.; Jisa, L.and Pintea, S.; Rusu, A.; and David, D. 2012. Children with autism social engagement in interaction with nao, an imitative robot: A series of single case experiments. *Interaction studies* 13(3):315–347.

Veale, T. 2013. A service-oriented architecture for computational creativity. *Journal of Computing Science and Engineering* 7(3):159–167.

Veale, T. 2014. Coming good and breaking bad: Generating transformative character arcs for use in compelling stories. In *Proceedings of ICCC-2014, the 5th International Conference on Computational Creativity*.

Veale, T. 2016a. A rap on the knuckles and a twist in the tale. *AAAI spring symposium series*.

Veale, T. 2016b. Round up the usual suspects: Knowledge-based metaphor generation. In *Proceedings of the Fourth Workshop on Metaphor in NLP*, 34–41.

Veale, T. 2017. Déjà vu all over again: On the creative value of familiar elements in the telling of original tales. In *Proceedings of ICCC-2017, the 8th International Conference on Computational Creativity*, 245–252.

Wilcock, G., and Jokinen, K. 2013. Wikitalk human-robot interactions. In *Proceedings of the 15th ACM on International conference on multimodal interaction*, 73–74. ACM.