Robot narratives

Marina Sanz Orell
James Bown
Susan Stepney
Richard Walsh
Alan F. T. Winfield

Orell, M.S., Bown, J., Stepney, S., Walsh, R. & Winfield, A.F.T. (2021) 'Robot narratives'. In: A. Adamatzky (ed.) *Handbook of unconventional computing. vol. 1: Theory*, World Scientific, Singapore, pp. 221-246.
DOI: https://doi.org/10.1142/9789811235726_0006

Chapter 1

Robot Narratives

Marina Sanz Orell*, James Bown†, Susan Stepney‡,
Richard Walsh§, Alan F. T. Winfield¶

There is evidence that humans understand how the world goes through narrative. We discuss what it might mean for embodied robots to understand the world, and communicate that understanding, in a similar manner. We suggest an architecture for adding narrative to robot cognition, and an experimental scenario for investigating the narrative hypothesis in a combination of physical and simulated robots.

1. Introduction

We start from the *narrative hypothesis*, that *humans understand 'how the world goes' through narrative*: we make sense of more or less complex events through the stories we hear, tell, imagine, and construct: "narrative is our innate way of representing process—it's the form in which we make sense of stuff happening [...] our cognitive framework for representing behaviour is narrative " [18, p.5]. This starting point has led us in two related directions.

First, complex systems and their emergent properties, including feedbacks, multi-scale interactions, and tipping-points, appear unnarratable, except by giving the emergent property some form of *agency* (for example, evolution and Mother Nature [1]). We explore issues around this challenge of *narrating complexity* in [19].

Second, if we wish to communicate with artificially intelligent robots, be

*pv.sanz.o.marina@gmail.com
†School of Design and Informatics, Abertay University, UK; j.bown@abertay.ac.uk
‡Department of Computer Science, and York Cross-disciplinary Centre for Systems Analysis, University of York, UK; susan.stepney@york.ac.uk
§Department of English and Related Literature, and Interdisciplinary Centre for Narrative Studies, University of York, UK; richard.walsh@york.ac.uk
¶Bristol Robotics Lab, University of the West of England, Bristol, UK; Alan.Winfield@brl.ac.uk

they our helpers, workmates, or carers [15], then they need to understand and relate to the world the way we do: through stories. And if they can do so, can they then relate to each other in the same manner? Such questions of *robot narratives* are the focus of this chapter.

Our discussion is structured in three sections. In §2 we provide a motivating example in the form of contrasting narratives of robots exploring a planet. The first set of robots have narrative understanding of their situation and task; the second set have declarative logical understanding. In §3 we outline a design for a model of robot narrative understanding, and then discuss two narrative scenarios related to negotiating a flight of steps. In §4 we suggest a programme of experimental robotics that could be used to develop and explore robot narratives, and to test the narrative hypothesis in robotics.

2. Robots explore an alien planet

Here we give a motivating example of 'narrative robots' through two contrasting tales of exploration. Two teams of robots are given the same task (§2.1), that of thriving on an unexplored planet. The 'Greek' team have narrative understanding (§2.2); the 'Roman' team use declarative logic (§2.3). They have different experiences, and we the readers have different reactions to their discussions.

Clearly, this is an imagined example, with certain situations exaggerated and foregrounded to make our points (that is, it is a story), but our aim is to illustrate what might be possible with narrative understanding.

2.1. *Prologue*

Twelve identical intelligent robots awaken on an unexplored planet, standing in a circle. Their knowledge of the territory is limited to the laws of physics and chemistry that are as applicable here as they are on planet Earth, given their similar atmospheric composition and surface gravity.

Their mission is to thrive, to prosper, to attain full potential; they understand what that means in regards to their survival, but not how this new environment will allow it. To succeed, the robots decide that they should cooperate and work as a team.

The first step is surviving; for that they need information, fuel, material for maintenance, a base, and other resources. With limited information about the terrain the robots realise that their first task should be recon-

naissance. They dissect the terrain in twelve equal sectors centered on their current position. They plan to explore for six hours, then return to share their experiences.

Once the six hours have passed, each robot starts heading back to the original meeting place. They arrive back at different times since those robots that encountered fewer challenges went further than others. Some carry more relevant information than others, some have recorded more details about their experiences. Some have encountered a more varied range of phenomena and some have developed more complex ideas and processes. All these differences now mean that the twelve robots are no longer identical: they are separate entities with different knowledge, different priorities, and different skills. Once the twelve are reunited they start relating their discoveries to their companions, through narrative (Greek) or declarative (Roman) means.

2.2. *Greek Olympus, narrative robots*

Zeus, Hera, Poseidon, Ares, Aphrodite, Hades, Hephaestus, Athena, Hermes, Artemis, Demeter and Hestia awaken on the unexplored planet Greek Olympus (Figure 1), go off exploring, reunite, and start telling the stories of their exploration.

Hermes tells about his experience first. This is his story:

I started moving in a straight line from my origin. For the first hour of moving at a moderate speed I saw nothing relevant; the ground was sandy and red as it is here. During the next hour I started encountering rocks that varied in size and were very hard but cracked easily. Initially these rocks were scattered, but during my 3rd hour of travelling they started to appear in larger groups and sizes, and I predicted that I was approaching a bigger rock formation. This suspicion was confirmed as I realised that I was entering a canyon. Natural stone pillars of the same red colour towered around me and strong gusts of wind carried grains of sand into my joints, making movement slower and more difficult. Then I encountered a deep narrow gorge that cut across the straight path I had been following. I figured that the least dangerous and risky solution was to move around it to get to the other side. It was slow but safe as expected. I continued moving along the canyon. There were many cracks and holes in the wall that weren't big enough to fit my body through. The entire formation looked quite fragile, and when I tried testing the malleability of one of the rocks it didn't sustain much stress before disintegrating. I continued on until the sixth hour struck,

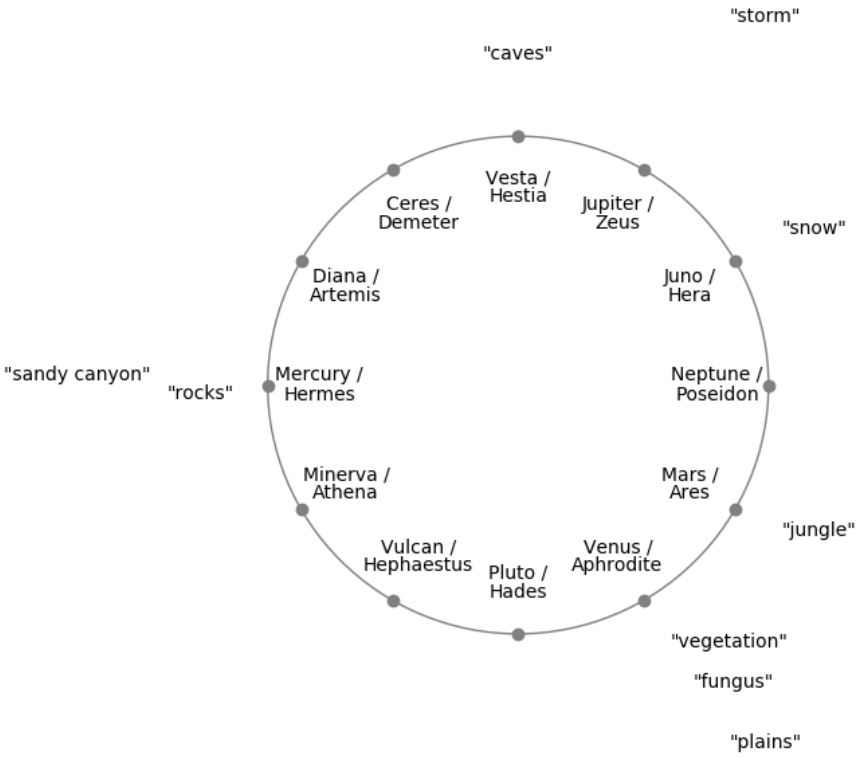

Fig. 1. The 12 Roman (declarative) and Greek (narrative) robots exploring planet Olympus, with some of the features they discover.

and I turned back to return here. I didn't find any fuel material, and my sector doesn't offer any good candidates for a settlement location, as the canyon appears dangerous and unstable.

Hermes finishes his story; the other eleven robots are listening carefully, processing the story, and noting the main points:

- The first sector does not contain fuel sources or settlement locations.
- There is a canyon-like geological formation that appears dangerous and unstable. The rocks that form it were hard but brittle. Hermes has warned them off the area.
- Gusts of wind might carry sand particles that get stuck in articulations making movement complicated; this is a real danger that Hermes warns about from personal experience.

- A successful way to deal with a deep narrow gorge is to walk around it.
- Based on Hermes' tale, the wind and the deep gorge situations are inconvenient but not insuperable.

Following Hermes's tale, it is Artemis' turn to share her experiences. But Zeus intervenes, before his turn, to tell the others that he has time-sensitive important information: he has witnessed a distant slow-moving thunderstorm during his exploration, and predicts it will reach their current location in about an hour. He says the best course of action is to have shorter reports so they can quickly decide on a settlement location and take shelter from the storm. All the robots agree this is the best course of action, so they give their reports in a shorter form.

Hestia describes a location that is less than an hour away and consists of a cliff with deep caves all over its surface, like a rocky beehive. Some of the caves form interconnected tunnels inside the cliff, and the lowest caverns may contain running water. The caves are structurally sound, and there were no noticeable hazards, just a lot of varied vegetation in the area in front of the cliff.

Aphrodite encountered few obstacles, and so her report focusses on the biodiversity in her sector: *I found my sector to be quite plain and safe. I encountered no obvious hazards or obstacles. I did record various forms of carbon-based organisms akin to vegetation; between the second and fourth hours I was moving across purple fields covered in a specific type of these organisms and also a species of fungal-like life forms. Past these fields there were plains with scattered groups of flora arranged in a bush-like fashion. It was pleasant and safe, and a good source of organic material. It could potentially present a good settlement location.*

Hestia's and Aphrodite's reports are considered alongside Demeter's and Artemis' for potential settlement locations. Zeus presents the four options since he is managing this time-sensitive mission. The robots have developed slightly different impressions and perspectives because of their different experiences, but they are still sufficiently similar that there is consensus to declare Hestia's suggestion as optimal.

Hestia gives the group instructions on how to get there: *We head directly north. After about 30 minutes, assuming we're moving fast, we'll encounter a fairly steep incline. It's better if we approach it from the right, because the path is smoother there. At the top of the incline there's a plain that we have to cross, and finally we reach the cliff. We can enter one of the*

ground level caves without any water and settle there.

With these directions, the group of robots starts moving towards the location. As they do, each robot exercises its own simulation as it faces smaller challenges such as adapting to different kinds of soil, and updates its world model. Hermes struggles as his movements are made difficult from the sand still in his joints. Ares has developed a crouched and jerky way of moving after navigating a dense jungle in his sector, and he has to modify his style for the new terrain. Hera is used to taking big strides, from walking through snowy slopes. They all alter their methods of movement, and develop slightly different styles. When they reach the incline, Hestia simulates the best way to climb it as a group, and instructs them to move in single file along the smoother rightward path.

They reach the cave where they can take shelter from the approaching thunderstorm. They are all aware they need to decide how to build their settlement, and so they engage in a session of brainstorming. The differing points of view give rise to different ideas and a richer understanding. They extract plans from all their suggestions, based on their different observations and considerations, and vote on how to address the shelter situation and distribute the other tasks. The robots adopt different roles according to their skills.

Having developed different areas of expertise and different perspectives about the world and their own selves, each robot grows in different directions. They diversify until they no longer have a unique memory bank that characterises them all, but a different way of perceiving, processing, storytelling and overall sense-making. This gives rise to individual behaviours and a range of dynamics like leadership and competition, and other forms of social interaction like games and culture and art.

We might like to think such behaviours are uniquely human, but we see similar practices emerge in some animal species, such as dolphins and chimpanzees [4]. So in our story here, our Greek Olympians go on to develop humour, sport, poetry, and, of course, story telling.

2.3. *Roman Olympus, declarative robots*

Jupiter, Juno, Neptune, Mars, Venus, Pluto, Vulcan, Minerva, Mercury, Diana, Ceres and Vesta awaken on the unexplored planet Roman Olympus (Figure 1), go off exploring, reunite, and start stating the facts about their exploration.

Mercury states his experience first, then all the others follow in turn.

Each robot describes their exploration in detail, giving explicit descriptions of everything they have encountered, including the scientific data. For example, Mercury states the details of the wind strengths along his journey, and then separately describes his physical status, noting movement is slightly problematic due to sand in his joints.

When they describe a hostile environment, like Mercury's canyon or Juno's snow, they give an objective description of the place. Every description is a detailed enumeration that allows each robot to construct an internal map of what is being described. By the end of their round of reports, each robot has an entire map of each sector in their world characterised by the data reported by each robot in turn.

Jupiter does not intervene to warn them of the approaching thunderstorm; when his turn arrives, he describes the storm objectively, but it does not carry a sense of urgency.

Now they have all shared their experiences they decide that they should act on the thunderstorm information and find a shelter. Their interior models are very similar to each other again, since they have all reported and recorded their respective data. They thus all determine they should go to Vesta's caves. They all start moving towards the cliff with their excellent virtual maps; there are no unexpected surprises since they know the terrain as if they explored it first-hand. They all climb the incline single file without a need to communicate. The thunderstorm catches up to them before they reach the caves, but fortunately it does not cause any damage.

Once in the cave, they do not really need to communicate to decide on a plan, as they are all so similar. They do not establish different roles, they all remain uniform. In being so homogenous, instead of a group of individuals, they all function as one. This is efficient, but it will severely limit creativity in their responses to forthcoming challenges. So in our story here, our Roman Olympians remain 'robotic' in their behaviours and communications.

2.4. *Epilogue*

Dramatis Personae: Computer Scientist (CS); Narratologist (N); Robot (R)

CS is applying a screwdriver to R's head

N: What are you doing?

CS: I'm trying to make my robot more human-like, by making it speak naturally.

R: `There is a hole. There is a corner. The hole is around the corner.`

N: That doesn't sound very natural! You want to use Narrative Logic.

CS: A what? Is that something from that cognitive narratology you've been telling me about for years? (And even with us writing that book on it, I still can't say it!)

N: That's right. We *know* that people understand the world through narrative – through stories. If you want to make your robot sound more human-like, it needs to talk about the world the way we would. So you need to give it Narrative Logic. Then it can talk about the world using stories.

CS: How can I do that?

N: How does your robot work?

CS: It's got a model of the world in its head. It builds that from its interactions in the world. It then uses that model to run simulations, and to plan its actions in the world. I was bolting on a declarative grammar module, so that it can make statements about its world model.

N: Okay, so it sounds like you need to work with embodied and enactive cognition. Why don't I bolt on a Narrative Logic module to help with that.

N applies a spanner to R's head

R: I went round that corner yesterday. I fell into a hole. I was stuck there all day.

CS: That was easy!

N: Well, no, it's not really that easy. We have to design the narrative logic, and connect it with the robot's model. I know about Narrative Logic, but not how to connect it to robots.

CS: Well, I know about interfacing to robots, and simulating them. And I know how the model in this one's head works.

R: And I can help you run experiments!

CS: Hey, together, we could get robots to understand the world the way we do! We can map out different narrative structures for different problems! One for ... a human companion. One for ... talking about legal regulations. One for ... social learning from each others' stories. One for ...

N: Hang on, hang on! Before we start all that, we need to see if we can

get this one robot to tell stories about its own world.

CS: But we could then use this robot to help design those other narrative logics? The robot would be more human-like?

R: Oh no! Here I am, brain the size of a planet ...

N: Yeah, yeah ... if the idea of robots telling stories works in the first place. Shall we work together on that first? And then we can tackle the bigger problem of social narrative, and of robots learning their world through stories, later on.

CS: Okay, sounds like a plan. Let's do it!

N: So, where could we get the funding to do that?

3. Robot imaginations

How might such 'narrative robots' be possible? In this section we describe a particular architecture of a robot mind that contains a model of the world including itself, where the robot can use this model to simulate scenarios in order to choose among potential courses of actions [2, 21]. This architecture has been suggested as a starting point for robots telling stories [20]. We then illustrate some potential scenarios of this model in use.

3.1. *Dennett's Tower*

Winfield [20] describes a succession of more complex creatures that can be used to design more complex intelligent robots. This is based on Dennett's idea of the *Tower of Generate-and-Test* [6], a conceptual model of levels of intelligence. Dennett's Tower not only provides us with a powerful model for types of intelligence, but a compelling route toward much more capable socially intelligent robots.

On each level of Dennett's Tower are creatures successively more capable of reacting to and surviving in the world, each having more sophisticated strategies. At the lowest level are (Charles) *Darwinian* creatures: new individuals are generated by variation from their parent(s), and are tested by selection in the real world; populations 'learn' through evolution; individuals do not learn. Next are (B.F.) *Skinnerian* creatures, who generate possible actions and test them by enacting them in the real world; individuals learn through reinforcing successful behaviours. Next are (Karl) *Popperian* creatures, who have internal models, in which they generate possible actions, and test them in the imagined world; they discard unsuccessful ones without needing to enact them in the dangerous real world.

Finally there are (Richard) *Gregorian* creatures, who are social learners: they can learn successful behaviours generated and tested by others.

3.2. *Robots in the Tower*

The majority of present-day robots, including those in research labs, have no mechanisms for learning: their behaviours are predesigned and hard-wired. These robots do not even make it to the ground floor of Dennett's Tower. A small number of research robots, within the subfield of evolutionary robotics [7], use genetic algorithms to evolve new behaviours: these are Darwinian creatures in Dennett's scheme. Another small set of research robots use reinforcement learning approaches [10]: these are Skinnerian. A handful of research robots have employed self-simulation embedded within the robot, to create Popperian robots [3, 11, 12, 17, 22, 25].

3.3. *An architecture for an ethical Popperian robot*

Winfield and co-workers [2, 16, 21, 23] describe an 'ethical' robot, ethical in that it may choose actions that compromise its own safety in order to prevent another from coming to harm. This ethical robot has an embedded simulation of itself, other dynamic actors (robots), and its currently perceived environment. This simulation is used as a real-time consequence engine capable of modelling, evaluating and weighting next possible actions against safety and ethical rules.

The model is summarised in Figure 2. On the right is shown the standard 'sense-plan-act' robot system. On the left is its additional Popperian 'interior model' (IM), where a loop *generates* potential actions; these are simulated in the robot/world model in the context of the real perceived environment (sensor input); the simulated results are *tested* by the consequence evaluator; actions that result in beneficial consequences are promoted to the robot, those that result in harmful consequences are inhibited.

Such a robot could also use its internal model to investigate other potential consequences of actions, thereby enabling it to make choices based on outcomes such as efficiency or safety, as well as ethical behaviour.

3.4. *From Popperian to Gregorian robots*

Winfield [20] proposes how this Popperian architecture might be exploited to enable robots to 'tell each other stories', and become Gregorian learners. Instead of simulating internally generated (imagined) actions, the robot can

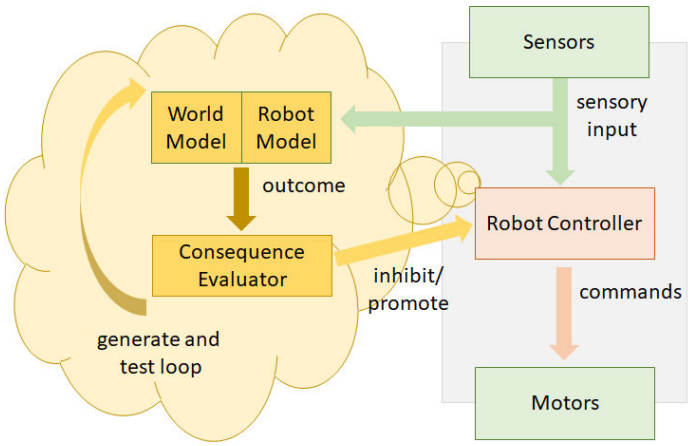

Fig. 2. Popperian robot model, with interior generate and test loop (adapted from [20, fig.4.1]).

interpret externally heard (story) actions and action sequences, and hence learn from the experiences of others.

Digging into what would be needed to achieve this, we find several requirements.

Rather than have to regenerate and test actions in each new context, there needs to be a *repository* of previous actions – real, imagined, or heard – and their consequences. This repository, along with the generate and test loop, can be used as a basis for future actions, imaginings, and stories.

The internal model may be inadequate in various respects, and an imagined action might be judged efficient or ethical or safe, but when carried out in reality, result in unanticipated behaviours or consequences. The consequence evaluator needs to be able to evaluate real world consequences, compare them with modelled consequences, and update both the model and repository as needed.

On the social side, the system needs a story parser, to hear stories told by other robots, parse into actions and consequences, and store in the repository for future use. It also needs a story generator, that can take items from the repository, and turn them into stories told using narrative logic. For these to be 'stories', rather than bald sequences of actions, they need to be parsed and assembled through a 'narrative logic', rather than as mere declarative statements.

It might be thought useful for the repository to include 'evidential mark-

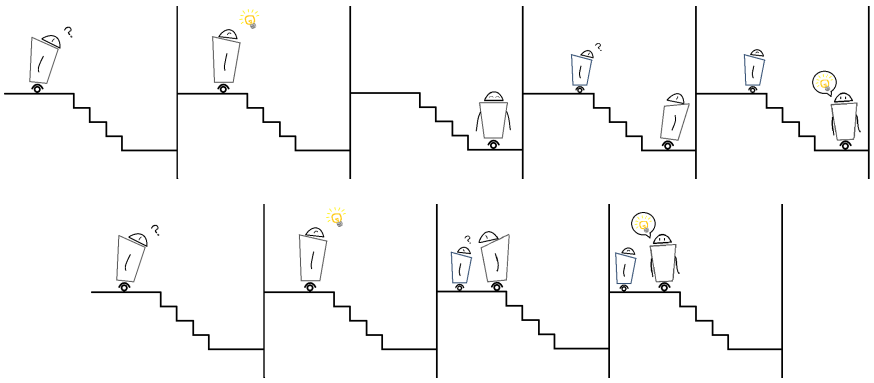

Fig. 3. Scenario 1 (top): telling the story of a successful descent. Scenario 2 (bottom): telling the story of an imagined descent.

ers', to distinguish consequences determined through real experience ('I did'), imagined experience ('I think'), observations of others' real experience ('I saw they did'), others' reported real experience ('they said they did'), others' imagined experience ('they think'), and so on.

See Figure 3 for how these components might work. In scenario 1 (top): Robot encounters some steps and wonders how to descend; it uses its internal model to evaluate scenarios until it finds a suitable course of action; it updates its repository with the imagined descent methods and their consequences ('I think I should do X, but not Y or Z'); it successfully descends the steps. Junior arrives and wonders how to descend; Robot tells Junior the story of how it successfully descended ('I did X'); Junior parses the story, imagines it through its own internal model, and learns how to descend. If Robot had instead fallen down the steps, because of a deficient model, it would then update its model to be a better predictor, and update its repository with the real world consequences of that particular descent method ('I thought I should do X, but I was wrong; I'll remember not to do that next time').

Scenario 2 (bottom): Robot encounters some steps and wonders how to descend; it uses its internal model to evaluate scenarios until it finds a suitable course of action ('I think I should do X, but not Y or Z'). Junior arrives and wonders how to descend; Robot tells Junior the story of how it imagines it should descend ('I think X will work'). If Robot is playing a more educative role, it might also say 'and don't do Y or Z', problems that it might have imagined in this case, or learned from previous misadventures,

or been taught by others when it was a more junior robot itself. If Junior runs these scenarios and finds a problem with X, or no problem with Y or Z, it can ask 'why (not)?'; Robot can answer with the relevant consequences from its own more advanced world model, and Junior can update its own model to deliver those consequences, too.

3.5. *Narrative logic*

Stories not only serve to share information, they also make sense of it. Work in narrative theory has emphasised the cognitive foundations of narrative in a 'story logic' [8]; this logic is expressed in the structure of stories, and is the basis of the narrative understanding by which we make sense of stories and use stories to make sense of experience. A story, by adhering to the formal features of story logic, gives narrative framing of the information it contains; it assimilates that information to an established structure of meaning and in doing so attributes a particular relevance and significance to it.

Because narrative foregrounds action and events, it is a privileged means of representing the behaviour of agents in interaction with their environment and each other. Crucially, stories mediate between the particulars of experience and the general framework of narrative understanding manifest in the set of stories in circulation, or those already familiar to a particular individual (which define that individual's narrative competence). There is a reflexive relation between particular stories and the story logic they use; each story depends upon the current set of stories as the basis for its intelligibility and significance, but also supplements that set and affects the general context of narrative understanding. The learning potential in this evolving culture of narrative meaning epitomises the distinctive kind of advantage Dennett attributes to Gregorian creatures.

Our short play earlier contrasts two statements by the robot character:

R: `There is a hole. There is a corner. The hole is around the corner.`

R: I went round that corner yesterday. I fell into a hole. I was stuck there all day.

The first set of sentences are declarative statements, and there is no story. The second set of sentences are also grammatically declarative, yet they form an (albeit trivial) story. What is the difference?

Consider just the two sentences, "I went around that corner yesterday. I

fell into a hole." From a human perspective this constitutes a narrative, but that does not make it narrative for the robot. Each sentence is a narrative utterance in its own right – by virtue of the action represented in the verbs "went" and "fell" – and the temporal extension of those processes is easily available to inference; but the information they convey might equally be taken as something closer to a status update. More importantly, the narrative articulation between the two sentences is not encoded by anything except their juxtaposition. Nothing positively requires us to understand them as more than a pair of unrelated assertions. If the robots are to make narrative sense of their experiences, and of the stories they tell each other, they need to be provided with a rudimentary *narrative logic* distinct from their linguistic competence and from their sensory engagement with their environment.

Our own predisposition towards narrative sensemaking makes available the inference that the two statements are to be understood sequentially; that falling in the hole *followed upon* going around the corner. Only on that basis is the further inference available, that falling in the hole was a consequence of going around the corner. This last inference is what gives the utterance its main communicative relevance: beyond the mere declarative information that there is a specific hole around a specific corner, the narrative particulars instantiate a generalisation: corners may hide holes. So there are two kinds of narrative implication involved: sequential-causal implication, and particular-general implication. These are both fundamental narrative heuristics, essential to narrative's value as a form of sensemaking, even while they lack logical rigour.

Narrative logic is inexact, and prone to fallacy: the *post hoc ergo propter hoc* fallacy (that what comes after is caused by), and the inductive fallacy (that what is true in this case is true in all cases). As the articulation of temporal experience, narrative is essentially concerned with matters of change and continuity, or of temporal difference and relation; there is no narrative object as such. Its connective logic therefore cannot be made fully explicit, but only pursued to some extent, within an implicit context of assumptions. The effectiveness of narrative in cognition and communication depends upon the cultural process of continuous reflexive refinement of narrative sensemaking through the circulation of stories.

4. Gregorian Chat

We have started from the position that narrative is the way humans understand how the world goes, and have discussed an architecture whereby robots might be endowed with narrative intelligence, too. However, there is nothing in that discussion that requires the stories heard and generated to be based on *narrative* logic specifically; the same argument could be used to support purely declarative statements. The narrative hypothesis is stronger, and in this section we discuss it in some more depth, and outline how one might go about testing the narrative hypothesis in general, through robots with internal models and story-telling capabilities: Gregorian robots chatting with each other.

4.1. *The narrative hypothesis in more detail*

We take the narrative hypothesis, that humans understand how the world goes through stories, and break it down into specific claims:

- Narrative frames our understanding of how the world goes, in that we necessarily represent and communicate that knowledge in the elemental form of stories.
- The affordances of narrative cognition are the legacy of our evolutionary adaptation to our environment, and set the terms for our continuing understanding of the world.
- Any given story mediates between its particulars and the general logic of narrative form; narrative provides a route from episodic to general semantic memory.
- Narratives are social: we learn from the stories of others as well as by giving narrative form to our own experience.

4.2. *The Gregorian Chat system*

Here we tell a story of how the narrative hypothesis might be tested through a series of experiments involving robots with internal models.

The plan would be to build a small society of robots, each with an internal model, generate and test loop, consequence evaluator, repository of past actions, and story parser and generator. The robots would be placed in a complex environment where they could explore, encounter dangers and rewards, interact with other robots, hear and tell stories of the world, reproduce and evolve. The hypotheses could be tested in various ways,

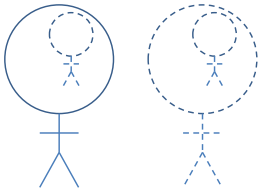

Fig. 4. (a) a physical robot (solid line) with a simulated self model (dashed line); (b) the same simulation used to simulated a robot with a simulated self-model

particularly by contrasting robot societies based on declarative logic stories, versus narrative logic stories.

Such a project would be an ambitious undertaking, and it would be essential to control the difficulty of implementation, experiment, and evaluation. However, it should be possible to use much "off-the-shelf" technology, to augment embodied robots with simulations, and to constrain the environment, as described here.

4.2.1. *Robot architecture*

Augmenting embodied robots with simulated robots. As described above, the individual robots need an internal model of themselves, for the Popperian simulation and consequence evaluation. A key insight [2, 12] is that the very same simulation approach used for an internal model can be used to simulate multiple instances of a larger population of robots, each with their own internal model.

Giving robots comprehensible grammars. The robots' grammars should conform to (a subset of) English, so the generated stories and declarative statements can be analysed . One way to accomplish this would be to equip robots with off-the-shelf speech-to-text and text-to-speech, allowing them to hear and produce stories externally, but readily transform this to text. Any errors in this translation, in either direction, can be considered to form a necessary part of the embodiment [12, 13, 25].

Formalising narrative logic. A crucial part of any investigation of the narrative hypothesis in robots is a need to develop a formalisation of narrative logic, such that the robots can construct 'stories' rather than declaim a sequence of facts. From the discussion in §3.5, this is a non-trivial challenge, and, in fact, forms the core of any such investigation: how can we provide a

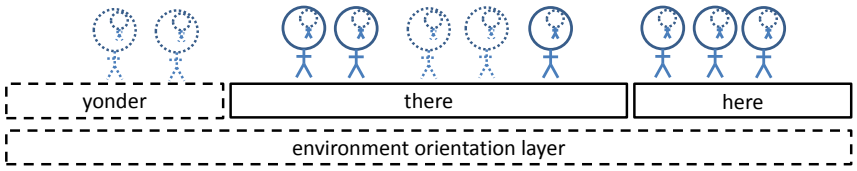

Fig. 5. The environmental architecture, comprising physical (solid line) and simulated (dotted line) robots embodied in physical (solid line) and simulated (dashed line) environments, mediated by an Environment Orientation layer.

narrative logic that is simultaneously formal enough to be programmed, and supple enough to capture the inexactness, implicit inferencing, and potentially fallacious reasoning underlying our human narrative understanding?

4.2.2. *Hybrid physical/virtual environment architecture*

The environment should be sufficiently complex to support a range of useful stories about it. There should be a complex geometry for the robots to navigate, with dangers and rewards, and opportunities to meet, observe, and interact with other robots.

This complexity could be achieved without the need for a large physical setup, by exploiting a combination of simulated and physical environments (figure 5): 'here', comprising physical robots in a simple physical social environment, the home campfire, where the robots can tell each other stories of their experiences; 'there', comprising a mix of physical and simulated robots experiencing and observing a complex physical environment supporting adventure-generating narratives; 'yonder', comprising simulated robots experiencing and observing a range of complex simulated environments supporting more varied narratives.

In order to simplify the robot perception issues, and allow physical and simulated robots to interact, the environment could be implemented using Environment Orientation [9] in the form of a 'spoken dungeon'. The environment as 'dungeon master' could speak aloud cues to the robots – 'you are by a river'; 'there is an unknown robot behind you'; 'Junior has fallen in the river' – and manage the simulated robots. Such an environment allows high levels of control and configuration, which are necessary to test the narrative hypothesis.

4.3. *Testing the narrative hypothesis in a robot ecology*

Through such robot and environment architectures, it would be possible to simplify the low-level perception and communication implementations, whilst maintaining the advantages of embodiment [12, 24], and focus on the issues of interest: narrative generation and transmission.

Such an approach would allow the narrative hypotheses to be formulated in a concrete form that would allow testing and evaluation in the following way:

- *Narrative influences understanding*: Seed the system with a range of different narrative styles, and observe and analyse the robots' responses to environmental and social stimuli, both previously seen and novel.
- *The world influences narratives*: Seed the simulations with different environments (for example, safe *v* dangerous, simple *v* complex, 2D *v* 3D) and observe and analyse the differences in the narrative structures that form.
- *Narratives are social*: Implement the same scenarios for collections of Popperian (non-social) and Gregorian (social) robots, and evaluate and compare their responses to situations previously seen by self, by others, or novel.
- *Narrative converts episodic to generic memory*: Seed the robots with narrative and declarative grammars, and evaluate and compare their responses to situations.

Such an approach would support an 'ecological' system of robots, with predefined grammars and internal models coping with a single generation of the world.

4.4. *Extensions of the approach*

The above scenario is in some sense the simplest approach to testing the narrative hypothesis. The same robot and environment architecture could be exploited to test that narrative hypothesis in more depth.

It could be extended to a system of evolutionary robots, evolving their internal models and repositories over generations. We hypothesise that the evolved robots would be able to cope with environmental change more robustly than the purely ecological systems.

It could be extended to a 'nested' model of self, and other robots, where the internal model of self includes its own model of self, and others' model of

Fig. 6. Nested models of self and other's models, from [14, fig.22.1]. Robot images © Julianne D. Halley; used with permission.

self, etc, each with decreasing fidelity, to avoid infinite regress (see Figure 6). We hypothesise that the nested models would result in more complex story structures ('I think you said they imagined I did X').

It could be extended to allow for self-modifying narrative logics, where the underlying logic itself is subject to some form of Darwinian or Popperian learning. We hypothesise that the evolving logic, where constrained to physically plausible environments, would result in strange but human-comprehensible stories, whereas a logic evolved to contend with an 'alien' environment would result in less comprehensible stories.

5. Discussion and Conclusions

5.1. *Communication and cognition*

We need to distinguish two different approaches to the idea of robot narrative. First, it is a *communicative* faculty, used by robots who operate with internal world models, and conduct simulations, etc, but who *translate* those ways of negotiating their environment into narrative form for the purpose of communication with each other and with humans. Second, the robots may also use narrative as *cognitive* resource, so that narrative sensemaking is directly part of their engagement with their environment, and the formal basis of their cognition and communication is therefore the same. This second approach is more difficult to implement, but the first

allows narrative only a limited role.

5.2. *Social robots*

Narrative communication between robots is significant to the extent that they have different experiences, and different cognitive perspectives. The story of the Greek and Roman robot teams illustrates the difference between the social sharing of narrative experience among separate robot minds, and the pooling of information among distributed instances of one collective hive mind. Meaningful communication in general requires both connection and difference; the circulation of stories does not just build a cumulative repository of knowledge, but proliferates interpretations of stories, in the different contexts of the experience of individual robots. The reciprocity between the range of stories and the range of interpretations provides for the possibility of a progressive refinement of the narrative competence of individual robots, and so a rise in the overall narrative competence of the population.

5.3. *Narrative logic and its interface with world modelling in artificial intelligence*

Narrative depends on an implicit connective logic, and inferences from it. Because this connective logic concerns change (process, action) it cannot be grounded in the contents of a classical form of world model alone; nor can narrative knowledge be translated into the terms of such a world model without fundamental loss. Equally, because the horizons of the implicit recede continually before any process of cognitive inference that derives explicit assumptions from implicit relations (on the basis of precedent or of principle), narrative logic does not resolve into any final, grounded form in its own terms. It has to remain a provisional resource informing agency within an environment, to be drawn upon within the pragmatic limits of the situated negotiation between robot subject and world (and reciprocally modified by the experience of that negotiation). The distinctive force of narrative communication, according to this line of reasoning, requires it to be assimilated to a narrative mode of cognition that is embodied, situated, and enactive [5].

5.4. *Beyond a "repository of actions": the particular and the general in narrative*

Narrative sensemaking concerns the form of particular experiences, and makes sense of the particular by assimilating it to the familiar shapes of the general: to narrative forms, templates and scripts that encapsulate "how the world goes" at different degrees of abstraction. These general narrative forms themselves, however, need to be extrapolated from the particular to the extent that they are not pre-programmed. Such extrapolation is a form of pattern recognition, and is equally involved in narrative sensemaking in response to experience and in interpretation of a communicated story. Without such a capability, a robot's repository of actions remains a database of particulars of no relevance to any circumstances except the recurrence of specific situations. The reciprocal dependence between particular acts of narrative sensemaking and general narrative competence tends to compound narrative's vulnerability to fallacy, but such a feedback loop is fundamental to narrative's cognitive value.

5.5. *Story generator and story parser*

The most basic challenge confronted by this project is the design of a story generator and a story parser (the two would substantially mirror each other). This could be pursued in the first instance at the level of narrative communication between robots, in which case it is essentially a problem of translation into and out of parameters of the robots' world model, and is detached from the question of narrative's efficacy as way of negotiating experience in itself. The distinctive value of narrative in this case will be a matter of its utility in the social circulation and consolidation of knowledge. Many of the difficulties to be addressed at the communicative level are essentially the same as those that arise at the cognitive level, with respect to the circuit of *narrative sensemaking* (making sense of narratives in communication, and using narrative to make sense in cognition). One fundamental difference, however, is that generating and parsing stories in the service of a world model is quite different from generating and parsing stories in the service of enactive experience in an environment.

5.6. *Preparing for the future*

Robots now and in the near future are escaping the laboratory and entering into our lives as autonomous entities for social play and pet companions,

as health and social care support workers, and more [15]. As these robots encroach on our day-to-day lives a key advantage of them learning and expressing themselves in a narrative form is that we should be more able to readily understand them, and be readily understood. There is also the opportunity to share knowledge among robots for group-learning effects.

Such robot narratives are not limited to embodied robots but also simulated robots: AI. The concept discussed here, where progress can be made more quickly with some of the benefits of embodiment, extends to AI more generally and the wider set of applications to which AI relates. It would provide us with a useful lens through which to understand, manage and unpack AI decision-making processes since we can interpret the internal models via narrative.

Acknowledgments

MSO acknowledges funding from the 2017 YCCSA (York Cross-disciplinary Centre for Systems Analysis) summer school. JB, SS, AW, RW acknowledge the support of the Narrating Complexity workshop series, funded by University of York and University of Abertay.

References

[1] H. P. Abbott. Unnarratable knowledge: The difficulty of understanding evolution by natural selection. In ed. D. Herman, *Narrative Theory and the Cognitive Sciences*, pp. 143–162. CSLI (2003).
[2] C. Blum, A. F. T. Winfield, and V. V. Hafner, Simulation-based internal models for safer robots, *Frontiers in Robotics and AI.* **4**, 74 (2018).
[3] J. Bongard, V. Zykov, and H. Lipson, Resilient machines through continuous self-modeling, *Science.* **314**(5802), 1118–1121 (2006).
[4] B. Boyd, The evolution of stories: from mimesis to language, from fact to fiction, *WIREs Cognitive Science.* **9**(1) (2018).
[5] A. Clark, *Being There: putting brain, body and world together again.* Oxford University Press (1997).
[6] D. C. Dennett, *Darwin's Dangerous Idea.* Allen Lane (1995).
[7] S. Doncieux, N. Bredeche, J.-B. Mouret, and A. E. G. Eiben, Evolutionary robotics: what, why, and where to, *Frontiers in Robotics and AI.* **2**(4), 1118–1121 (2006).
[8] D. Herman, *Story Logic: Problems and Possibilities of Narrative.* University of Nebraska Press (2002).
[9] T. Hoverd and S. Stepney, Environment orientation: a structured simulation approach for agent-based complex system, *Natural Computing.* **14**(1), 83–97 (2015).

[10] J. Kober and J. Peters. Reinforcement learning in robotics: A survey. In *Learning Motor Skills*, pp. 9–67, Springer (2014).
[11] H. G. Marques and O. Holland, Architectures for functional imagination, *Neurocomputing*. **72**(4), 743–759 (2009).
[12] P. J. O'Dowd, M. Studley, and A. F. T. Winfield, The distributed co-evolution of an on-board simulator and controller for swarm robot behaviours, *Evolutionary Intelligence*. **7**(2), 95–106 (2014).
[13] S. Stepney. Embodiment. In eds. D. Flower and J. Timmis, *In Silico Immunology*, chapter 12, pp. 265–288. Springer (2007).
[14] S. Stepney and R. Walsh. From simplex to complex narrative? In Ref. 19, pp. 319–322.
[15] S. Turkle, *Alone Together*. Basic Books (2011).
[16] D. Vanderelst and A. Winfield, An architecture for ethical robots inspired by the simulation theory of cognition, *Cognitive systems research*. **48**, 56–66 (2018).
[17] R. Vaughan and M. Zuluaga. Use your illusion: Sensorimotor self-simulation allows complex agents to plan with incomplete self-knowledge. In *From Animals to Animats 9*, pp. 298–309, Springer (2006).
[18] R. Walsh and S. Stepney. Introduction and overview: Who, what, why. Ref. 19, pp. 3–9.
[19] R. Walsh and S. Stepney, eds., *Narrating Complexity*. Springer (2018).
[20] A. Winfield. When robots tell each other stories: The emergence of artificial fiction. In Ref. 19, pp. 39–47.
[21] A. F. T. Winfield. Robots with internal models: A route to self-aware and hence safer robots. In ed. J. Pitt, *The Computer After Me: Awareness and Self-Awareness in Autonomic Systems*, pp. 237–252. Imperial College Press (2014).
[22] A. F. T. Winfield, Experiments in Artificial Theory of Mind: From Safety to Story-Telling, *Frontiers in Robotics and AI*. **5**, 75 (2018).
[23] A. F. T. Winfield, C. Blum, and W. Liu. Towards an ethical robot: Internal models, consequences and ethical action selection. In *TAROS 2014: Advances in Autonomous Robotics Systems*, number 8717 in LNCS, pp. 85–96, Springer (2014).
[24] A. F. T. Winfield and M. D. Erbas, On embodied memetic evolution and the emergence of behavioural traditions in Robots, *Memetic Computing*. **3** (4), 261–270 (2011).
[25] J. C. Zagal, J. Ruiz-del Solar, and P. Vallejos, Back to reality: Crossing the reality gap in evolutionary robotics, *IFAC Proceedings Volumes*. **37**(8), 834–839 (2004).