Chapter 4
When Robots Tell Each Other Stories: The Emergence of Artificial Fiction

Alan F. T. Winfield

Abstract This chapter outlines a proposal for an embodied computational model of storytelling, using robots. If it could be built, the model would open the possibility for experimental demonstration and investigation of how simple narrative might emerge from interactions with the world and then be shared, as stories, with others. The core proposition of this chapter is that in such a system we would have a practical synthetic model of robot-robot storytelling. That model might then be used to experimentally explore a range of interesting questions, for example on narrative-based social learning or the relationship between the narrative self and shared narrative.

1 Introduction

The model set out in this essay has a surprising origin. It emerges from work toward making robots that can be safe in unknown or unpredictable environments (Winfield 2014). That work takes the idea of robots with dynamic, continuously updating, *internal models* (of themselves and their environment) and links that with Dennett's conceptual framework: the *Tower of Generate and Test*, leading to a new control system for safer cognitive robots. We then extend this schema, with the addition of a conceptually simple system for allowing robots to transmit and hence share parts of their internally modelled behaviour with each other. The core proposition of this chapter is that if we could build such a system, we would then have a model of robot-robot storytelling. That model might then be used to experimentally explore a range of interesting questions, for example on narrative-based social learning or the relationship between the narrative self and shared narrative.

A. F. T. Winfield (✉)
Bristol Robotics Laboratory, University of the West of England, Bristol, UK
e-mail: alan.winfield@brl.ac.uk

R. Walsh, S. Stepney (eds.), *Narrating Complexity*,
https://doi.org/10.1007/978-3-319-64714-2_4

2 Internal Models and Dennett's Tower of Generate and Test

An Internal Model is a mechanism for internally representing both the system itself and its current environment. An example of a robot with an Internal Model is a robot with a *simulation* of itself *and* its currently perceived environment, inside itself. A robot with such an Internal Model has, potentially, a mechanism for generating and testing what-if hypotheses; i.e.:

1. *what if* I carry out action x..? and, . . .
2. of several possible next actions x_i, *which* should I choose?

Holland (1992, p. 25) writes: "an internal model allows a system to look ahead to the future consequences of current actions, without actually committing itself to those actions". This leads to the idea of an Internal Model as a *consequence engine*—a mechanism for anticipating the consequences of actions. Dennett, in his book *Darwin's Dangerous Idea* (Dennett 1995), elaborates the same idea in what he calls the Tower of Generate-and-Test, a conceptual model for the evolution of intelligence that has become known as Dennett's Tower. Dennett's tower is a set of conceptual creatures each one of which is successively more capable of reacting to (and hence surviving in) the world through having more sophisticated strategies for generating and testing hypotheses about how to act in a given situation.

The ground floor of Dennett's tower represents *Darwinian creatures*; these have only natural selection as the generate-and-test mechanism, so mutation and selection is the only way that Darwinian creatures can adapt—individuals cannot. All biological organisms are Darwinian creatures. On the first floor we find *Skinnerian creatures*, a subset of Darwinians, which can learn, but only by generating and physically testing all different possible actions then reinforcing the successful behaviour—providing of course that the creature survives. On the second floor Dennett's *Popperian creatures* have the additional ability to internally model the possible actions so that some (the bad ones) are discarded before they are tried out for real. A robot with an Internal Model, capable of generating and testing what-if hypotheses, is thus an example of an artificial Popperian creature within Dennett's scheme. The ability to internally model possible actions is of course a significant innovation.

On the third floor of Dennett's tower, a sub-sub-subset of Darwinians, are *Gregorian creatures*. In addition to an internal model, Gregorians have what Dennett refers to, after Richard Gregory, as *mind tools*—including words, which they import from the (cultural) environment (Dennett 1995, p. 378). Conceptually therefore Dennett's Gregorians are social learners.

In the field of intelligent robots, specifically addressing the problem of machine consciousness (Holland 2003), the idea of embedding a simulator in a robot has emerged in recent years. Such a simulation allows a robot to internally try out (or 'imagine') alternative sequences of motor actions, to find the sequence that best achieves the goal (for instance, picking up an object), before then executing

that sequence for real. Feedback from the real-world actions might also be used to calibrate the robot's internal model. The robot's embodied simulation thus adapts to the body's dynamics, and provides the robot with what Marques and Holland call a 'functional imagination' (Marques and Holland 2009).

Bongard et al. (2006) describe a 4-legged starfish-like robot that makes use of explicit internal simulation, both to enable the robot to learn its own body morphology and control, and notably allow the robot to recover from physical damage by learning the new morphology following the damage. The internal model of Bongard et al. models only the robot, not its environment. In contrast, Vaughan and Zuluaga (2006) demonstrate self-simulation of both a robot and its environment in order to allow a robot to plan navigation tasks with incomplete self-knowledge; their approach significantly provides perhaps the first experimental proof-of-concept of a robot using self-modelling to anticipate and hence avoid unsafe actions. Zagal et al. (2009) describe self-modelling using internal simulation in humanoid soccer robots; in what they call a 'back-to-reality' algorithm behaviours adapted and tested in simulation are transferred to the real robot.

All of the examples cited here describe robots capable of generating and testing what-if hypotheses using simulation-based internal models; in Dennett's scheme they are all Popperian robots.

3 A Generic Internal Modelling Architecture (for Safety)

Simulation technology is now sufficiently well developed to provide a practical basis for implementing the kind of Internal Model required to test what-if hypotheses outlined above. In robotics, advanced physics and sensor based simulation tools are commonly used to test and develop, even evolve, robot control algorithms before they are tested in real hardware. Examples of robot simulators include Webots (Michel 2004) and Player-Stage (Vaughan and Gerkey 2007). Furthermore, there is an emerging science of simulation, aiming for principled approaches to simulation tools and their use (Stepney et al. 2018).

Figure 4.1 outlines an architecture for a robot with an Internal Model in which the model is used to test and evaluate the consequences of the robot's next possible actions. Note that the machinery for modelling next actions is relatively independent of the robot's controller; the robot is capable of working normally without that machinery, albeit without the ability to generate and test what-if hypotheses. The what-if processes are not in the robot's main control loop, but instead run in parallel to override the Robot Controller's normal action selection if necessary; acting in effect as a safety governor by inhibiting unsafe actions.

At the heart of the architecture is the Internal Model (IM). The IM is initialised from the Object Tracker-Localiser and accepts, as inputs, candidate actions from an action generator. For each candidate action, the IM simulates the robot executing that action, and generates a set of model outputs ready for evaluation by the Consequence Evaluator. The Internal Model and Consequence Evaluator loop through each

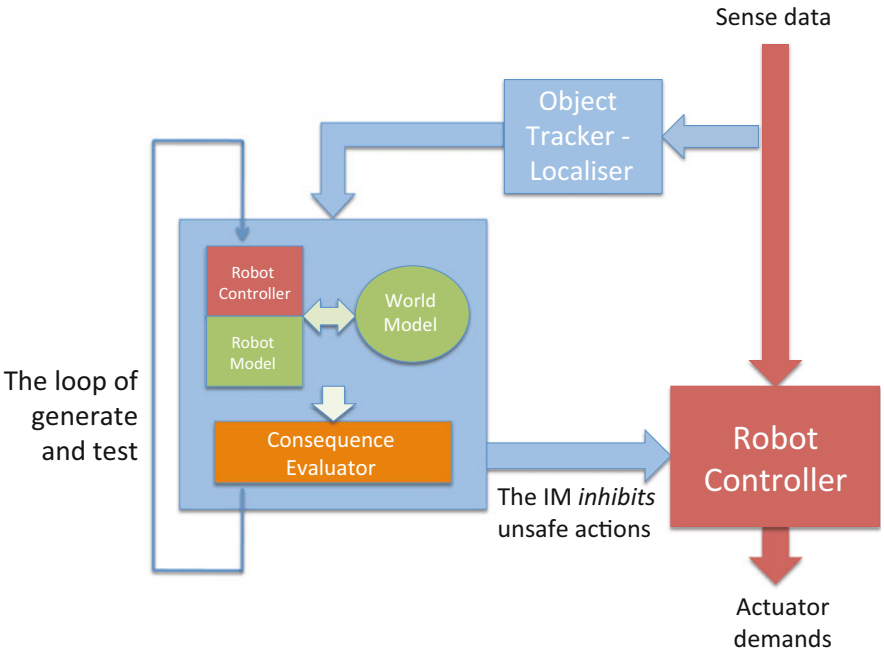

Fig. 4.1 A Control System Architecture for Safety. The Robot Control dataflows are shown in red (right); the Internal Model and its dataflows in blue (left)

possible next action; this is the loop of generate and test. The IM's simulator comprises three components: a World Model, Robot Model and Robot Controller; the latter is an exact duplicate of the real Robot Controller. The World Model is a simplified model of the robot's environment, including the robot's position and pose in it, at the present moment. Only when the complete set of next possible actions has been tested does the Consequence Evaluator send, to the Robot Controller, actions it assesses to be unsafe.

We have implemented the simulation-based internal model outlined here in a system of e-puck mobile robots and, with an additional logic layer demonstrated robots with simple ethical behaviours (Winfield et al. 2014), and robots with improved safety in dynamic environments (Blum et al. 2018). That system was able to generate and test 30 next possible actions every 0.5 second.

4 An Embodied Computational Model of Storytelling

Dennett's Tower describes an evolutionary drive toward internal modelling, allowing what-if generation and testing strategies for action. Let us explore the idea that these several what-if narratives are constructed fictions: they haven't

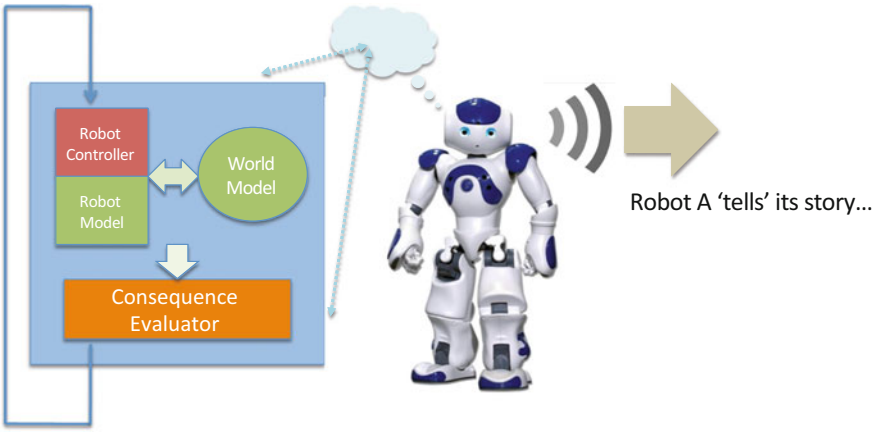

Fig. 4.2 Robot A, the storyteller, 'narrativises' one of the 'what-if' sequences modelled by its generate-and-test machinery. First an action is tested in the robot's internal model (left), second, that action—which is not executed for real—is converted into speech and spoken by the robot

happened; most will never happen. Dennett's Popperian creatures thus, in principle, have the cognitive machinery for the creation of fictional narratives. If we allow them to 'tell' those stories then they become Gregorian creatures.

Assume that we have two robots, each equipped with the internal modelling machinery outlined above. Let us also assume that the robots are of a similar type, in other words they are conspecifics. Within Dennett's framework each robot is a Popperian creature; it is capable of generating and testing next possible actions. Let us now extend the robots' capabilities in the following way. Instead of simply discarding ('forgetting') an action that has been modelled and determined to be a bad action, the robot may transmit that action to another robot.

Figure 4.2 illustrates robot A 'imagining' a what-if sequence, then narrativising that sequence. It literally signals that sequence using some transmission medium. In practice we could make use of any number of signals and media: Morse code via wireless, or body movements intended to be visually interpreted, for instance. But, since we are building a model and it would be very convenient if it is easy for human observers to interpret the model, let us code the what-if sequence verbally and transmit it as a spoken language sequence. Technically this would be easy to arrange since we would use a standard speech synthesis process. Although it is a trivial narrative robot A is now able to *both* imagine and then literally *tell* a story, and because that story is of something that has not happened, it is a *fictional* narrative.[1]

Robot B is equipped with a microphone and speech recognition process—it is thus able to listen to robot A's story, as shown in Fig. 4.3. Let us assume it is programmed to 'understand' the same language, so that a word used by A signifies the same part of the what-if action sequence to both A and B. Providing the story has

[1]Here we assume a simple ontological approach to what is fictional narrative.

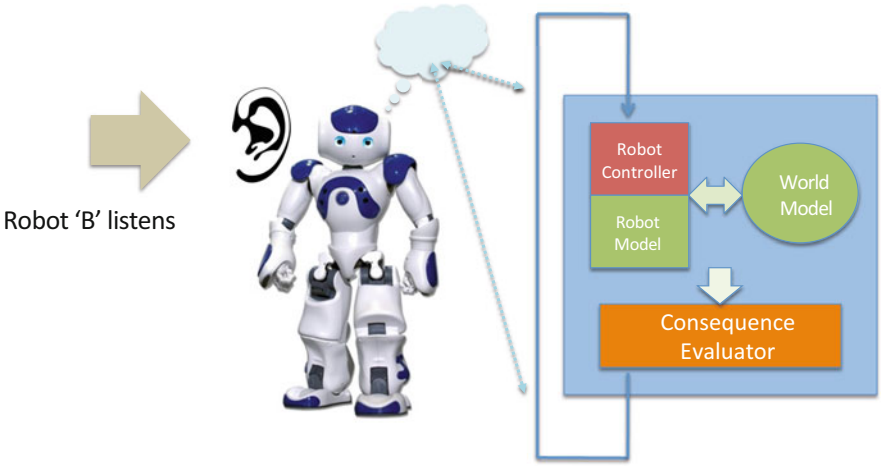

Fig. 4.3 Robot B, the listener, uses the same 'what-if' cognitive machinery to 'imagine' robot A's story. Here the robot hears A's spoken sequence, then converts it into an action which is tested in B's internal model

been heard correctly then robot B will interpret robot A's story as a what-if sequence. Now, because robot B has the same internal modelling machinery as A—they are conspecifics—it is capable of 'running' the story it has just heard within its own internal model. In order that this can happen we need to modify the robot's programming so that the what-if sequence it has heard and interpreted is substituted for an internally generated what-if sequence. This would be easy to do. But, once that substitution is made, robot B is able to run A's what-if sequence (its story) in *exactly* the same way it runs its own internally generated next possible actions, simulating and evaluating the consequences. Robot B is therefore able to 'imagine' robot A's story.[2]

In this model we have, in effect, co-opted the cognitive machinery for testing and discarding unsafe actions for imagining, or internally experiencing, heard stories. By adding the machinery for signalling and signifying internally generated sequences (narratives)—the machinery of semiotics—we have transformed our Popperian robots into Gregorian robots. Thus we have an embodied computational model of storytelling.

[2]Where is the meaning? It could be argued that when the listener replays the story in its IM (functional imagination) that *is* meaning.

5 What Could We Learn from This Model?

How does narrative emerge from interactions with the world? If we provide the robots outlined above with a context—a physical environment with physical features and, perhaps, safety hazards that they can move around in and explore—then, at a fundamental, level we are providing our robots with something they can tell stories about.[3] The physical act of moving through and exploring their environment, together with the cognitive act of running the internal model of Fig. 4.1, provide the robots with a rich set of 'imagined' what-if actions to share with each other using the model outlined above. There are practical details to resolve. For instance, how does a robot 'decide' when to tell a story? We might, for instance, trigger this action simply when the two robots find themselves in close proximity; if they are sharing a relatively limited space this could happen quite often. Another question is how does a robot decide whether to tell or to listen—the roles of robots A or B in Figs. 4.2 and 4.3? A simple mechanism might be to default to listening, but if a robot hears nothing for a randomly chosen number of seconds, then it switches to telling. A third question is how does a robot decide which of the several what-if actions tested in its internal model to tell? Here we could use the robot's evaluation of the consequences of those what-if actions; the one with the highest risk for instance might be the candidate for telling: "if I had continued to walk forward I would have fallen into a hole".

The 'robots gossiping' experiment outlined here would provide rich data for analysis. Perhaps most interesting would be to examine which simple stories are told and their relationship to the storytelling robot's current location in the world and the physical features in it. Equally interesting would be to look 'inside the head' of the listening robot and compare the way those heard narratives are 'imagined' from the different perspective[4] of the listener, given that its current position in the world is different. A simple extension to this experiment would be to provide robots with the ability to modify their internal models on the basis of heard stories so that, for example, the listener robot would add a 'potentially dangerous hole' to its world model. We would then have narrative-based social learning.

There are several further directions we could take these ideas.

First, consider the machinery for signalling and signifying narratives—the language. In the experiment outlined above this machinery is fixed and pre-programmed. If instead we introduce some plasticity so that robots can, for instance, either invent new signals or modify existing signals, for new features encountered in the environment, then we open the possibility for an emergent

[3] In the model set out here the context is the here and now. But of course the story could be used to create a different context for the listener, i.e., to initialize its World Model the story could begin: "Imagine you are standing by the . . ."

[4] Note that the listener's world model will be different to the storyteller's, since the objects and their locations in the world model are initialised by each robot's object tracker/localiser (Fig. 4.1) as it moves through the world.

robo-semiotics. While the idea of robo-semiotics is not new (Ziemke 2003) there are deep open questions on the cultural evolution of language (Steels 2011). The model outlined in this essay might allow us to address these questions in a new way by experimentally studying the transition from Popperian to Gregorian creatures.

Second, consider the potential for adding autobiographical memory structures to the robots. It would be relatively easy for a robot to build a memory of everything that has happened to it, but of much greater interest here is to integrate the autobiographical memory into the internal model, perhaps leading to what Conway (2005) describes as a self-memory system (SMS). Two experimental possibilities are of particular interest. One is that when an episode from the autobiographical memory is retrieved it is then rehearsed in the internal model, so memory recall becomes re-imagining. Another is that the autobiographical memory allows the storyteller robot to string together a series of recalled (and now re-imagined) actions into a longer narrative sequence.[5] Each robot, even though they are in a shared environment and with shared encounters, will have a unique personal narrative. Arguably each robot would then have, at least in some minimal sense, a developing narrative self.

Third, consider the relationship between the narrative self and shared narrative, i.e., the storytelling component of culture. In previous work the author has experimentally explored robots able to learn socially, by imitation. Because the imitation was embodied, imitation was imperfect and hence imitated actions—in this case short sequences of moves (dances)—mutated as they went through successive generations of imitation (Winfield and Erbas 2011). We call this noisy social learning. That work demonstrated behavioural evolution and the emergence of new behavioural 'traditions' in a robot collective; we also explored the impact of memory in the persistence of these traditions (Erbas et al. 2015). The robots of that work did not have simulation-based internal models.

Consider now the possibility that we allow several robots to learn socially from each other using the experimental models outlined in this essay, in particular narrative-based social learning and the narrative self. We then free run the experiment so that robots are able to gossip and re-tell heard stories, which then evolve and change over multiple successive retellings. We would then have an embodied computational model for exploring the emerging relationship between narrative self and shared narrative.

Acknowledgements The title of this chapter is a quote from the late Richard Gregory. In 2006 when discussing the possibility of emergent robot culture with the author, Richard Gregory declared: "when your robots start telling each other stories, *then* you'll really be onto something". The work of this chapter is partially funded by EPSRC grant reference EP/L024861/1.

[5]Note also that there is no reason that same machinery couldn't be used for the sharing of 'historical' narratives, rather than fictional, i.e., what actually happened to robot A, rather than what it imagines but didn't enact.

References

Blum C, Winfield AFT, Hafner VV (2018) Simulation-based internal models for safer robots. Front Robot AI 4:74

Bongard J, Zykov V, Lipson H (2006) Resilient machines through continuous self-modeling. Science 314(5802):1118–1121

Conway MA (2005) Memory and the self. J Mem Lang 53(4):594–628

Dennett D (1995) Darwin's dangerous idea. Penguin, London

Erbas MD, Bull L, Winfield AFT (2015) On the evolution of behaviors through embodied imitation. Artif Life 21(2):141–165

Holland JH (1992) Complex adaptive systems. Daedalus 121(1):17–30

Holland O (2003) Machine consciousness. Imprint Academic, Upton Pyne

Marques H, Holland O (2009) Architectures for functional imagination. Neurocomputing 72(4–6): 743–759

Michel O (2004) Webots: professional mobile robot simulation. Int J Adv Robot Syst 1(1):39–42

Steels L (2011) Modeling the cultural evolution of language. Phys Life Rev 8:339–356

Stepney S, Polack FAC, Alden K, Andrews PS, Bown JL, Droop A, Greaves RB, Read M, Sampson AT, Timmis J, Winfield AFT (2018) Engineering simulations as scientific instruments. Springer, Heidelberg (in press)

Vaughan RT, Gerkey BP (2007) Really reused robot code from the player/stage project. In: Brugali D (ed) Software engineering for experimental robotics. Springer, Heidelberg, pp 267–289

Vaughan RT, Zuluaga M (2006) Use your illusion: sensorimotor self-simulation allows complex agents to plan with incomplete self-knowledge. In: Nolfi S et al (eds) From animals to animats 9 (SAB 2006), LNCS, vol 4095. Springer, Heidelberg, pp 298–309

Winfield AF (2014) Robots with internal models: a route to self-aware and hence safer robots. In: Pitt J (ed) The computer after me: awareness and self-awareness in autonomic systems. Imperial College Press, London, pp 237–252

Winfield AF, Erbas MD (2011) On embodied memetic evolution and the emergence of behavioural traditions in robots. Memetic Comput 3(4):261–270

Winfield AF, Blum C, Liu W (2014) Towards an ethical robot: internal models, consequences and ethical action selection. In: Mistry M, Leonardis A, Witkowski M, Melhuish C (eds) Advances in autonomous robotics systems (TAROS 2014), LNCS, vol 8717. Springer, Heidelberg, pp 85–96

Zagal JC, Delpiano J, Ruiz-del Solar J (2009) Self-modeling in humanoid soccer robots. Robot Auton Syst 57(8):819–827

Ziemke T (2003) Robosemiotics and embodied enactive cognition. SEED – Semiotics. Evol Energy Dev 3(3):112–124