Proceedings of the Towards a Comprehensive Intelligence Test (TCIT): Reconsidering the Turing Test for the 21st Century Symposium,
Aladdin Ayesh, Mark Bishop, Luciano Floridi, and Kevin Warwick (Ed.), at the AISB 2010 convention, 29 March – 1 April 2010, De Montfort University, Leicester, UK

Actions and Observations – Demonstrating Aspects of Understanding in a Simple World

Chris White[1], **David Bell**[1]

Abstract. This paper aims to take a preliminary step in representing understanding in robotic agents at a basic level. Through the use of the 'Webots' platform, the experiment investigates reasoning through a single layer of abstraction. In creating a link between an action and an observation it is possible to reason about objects in space. Using symbolic grounding and a small scale interrogation of an agent we can build a small simulation of a basic Turing test. With the data generated, it is hoped that in further research through the use of classification, an agent can rebuild a story of its actions through its 'observer' and recreate a state space of endeavours. More importantly, the first steps are taken to develop a theory and accomplish more complex systems of reasoning and understanding in autonomous agents.

1 INTRODUCTION

By common consent, the scenario for the experience of interactions between humans and an external, physical environment is a 3-d phenomenal space extending around our bodies. We get information from that environment as we pick up signals through the detection of different forms of energy by means of our eyes, ears and other sense organs.

On the other hand it is also useful to think of an internal environment - the internal representation of experience used by the human agent. One of the Grand Challenges of the 21st century is to address the somewhat disturbing (apparent) fact that this internal representation and associated physical (brain) states – the qualitatively experienced world - seem to be completely different from the external physical environment where the original phenomena and actions took place (This is sometimes referred to as The Hard Problem of Consciousness).

To try to get a handle on this mystery we can isolate some components of this extensive scenario and see if we can emulate them within artificial agents or other artefacts, always remembering that scaling up in size and complexity is not trivial. As a simplification of this emulation, one interesting thing we can focus upon and 'isolate' is the mapping from sense data to mental constructs and some processed versions of these (compiled hindsight). Intuitively this is a key component of our systems for understanding. We can foster beliefs about the external physical world which are express imprints of experiences. We can also summarise these and accumulate categorisations of our experiences and their patterns. But most importantly in the present context, we can display our understanding by responding to queries that justify our decisions, actions, and conclusions by tracing them to grounded linkages to the real world and our actions therein.

We acknowledge that this leaves a lot of aspects of 'awareness' and 'understanding' unaccounted for – some examples being: social categorisation (Durkheim and Mauss) [1]; how we decide upon methods of logical, internal representations for our knowledge and thoughts (Weng) [2]; and the physical, neural mechanisms of representation and those whereby we translate from the external world to the internal world (Hameroff) [3]. The rich features of the external world can be curtailed in various ways for the purpose of experimentation. We can limit what the programmer knows about the external environment in which the agent works when programming, and the degree of how simple or how uncontrolled or how immutable or how unforeseeable the environment is. It is important, however, that the true state of the agent is represented properly, and that the behaviour generated by it is distilled correctly. In the present study we look at some conspicuous aspects of understanding, primarily through categorisation of behaviours and other experiences, and isolate them in order to see if we can emulate them in a robot.

Living and some non-living creatures are sensorimotor systems… The objects in the world come in contact with their sensory surfaces. That sensorimotor contact "affords" (Gibson's term) some kinds of interaction and not others…. Categories are kinds, and categorization occurs when the same output occurs with the same kind of input, rather than the exact same input… (Harnad)[4].

We believe that this takes a small contribution towards addressing some of the issues raised by Harnad [5] about grounding and the Turing test, while seeking to make an artefact with some recognizable kinds of understanding.

2 LIMITING THE ENVIRONMENTAL SCENARIO

We take understanding to be the ability to address an idea and rationalise about it - to learn about (e.g. to categorise) objects and events, summarise them, and drag them through an abstraction layer to the brain to form a concept. Our conceptual framework is simple: the agent's body carries out actions with no understanding of what they represent. Concurrently the 'mind' is processing what the body is doing linking to a corresponding table of abstractions that map to real world objects and events. To model a few of these actions and observations would indicate a sort of understanding as outlined by Jin and Bell (2003).

[1] School of Electronics, Electrical Engineering and Computer Science, Queens University Belfast, 18 Malone Rd, Belfast, BT9 5BN
Email: {cwhite06, da.bell}@qub.ac.uk.

Jin and Bell's account suggests that a level of understanding can be achieved by symbolic grounding. The functioning of the mind is separated into functions carried out by an actor and an observer, respectively. The actor pursues a project expressing its mood while engaging in activities. The observer records this scenario and links the abstract viewing to its own symbolic grounding. For example, the actor sporadically enters rooms while leaving a colour trail. This colour trail changes depending on what activity the actor is engaging in and if the actor is interrupted. The observer has symbolic grounding in that it can link colours to moods and areas to activities. While the actor is following a predefined path, the observer is able to reason through a single layer of abstraction, what the actor is doing and experiencing in terms of mood and activity.

3 THE ENVIRONMENT

The physical, phenomenal environment consists of a working space with 3 'rooms'. The agent can pick up attributes of the environment via sensors and with the help of a built-in observer, answer factual (sense data values) and explanatory questions about its experiences. Each activity is sought depending on what mood the agent is in. Each activity also has the ability to change the actor's mood.

The rooms are coloured differently and represent 3 distinct activities an actor can engage in (Fig.1). The agent that will be used in the final demonstration is the Khepera robot platform, and in the present preliminary investigation this was modelled by the software application Webots (Fig.2.).

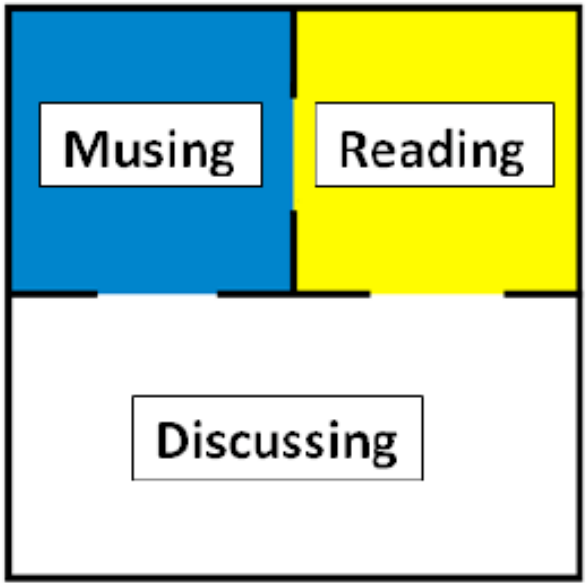

Figure 1. Overhead view of the environment, the rooms and the activities each room represents

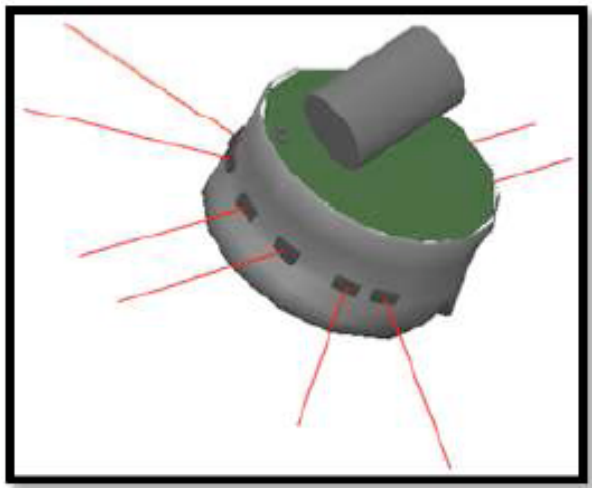

Figure 2. The Khepera agent as modelled by the Webots platform

4 THE EXPERIMENTAL DESIGN AND IMPLEMENTATION

The experiment aims to illustrate a design built on Jin and Bell's account of understanding [6]. The actor (Khepera) will follow a predefined path while leaving a colour trail based on its activity, mood and interrupts. The actor's predefined path changes with activity and interrupt. This means that if it is in a good mood, the actor will seek to read, however, if interrupted before the actor can get to the reading area, the mood and therefore the goal will change as detailed in Table 1.

The colour trail is then analysed and the pixel colour is fed to the observer. The observer can consult its memory to link the colour with a mood as shown in Table 1. The same methodology is applied to activity reasoning. The actor can constantly feed its physical location via GPS coordinates to the observer without ever understanding what it means. The observer takes this data and, again through grounding in its memory, is able to link certain boundaries with different activities as shown in Table 2. This is achieved through grounding on the observer's part; GPS coordinates of the actor as well as knowledge of room boundaries.

Finally an extra dimension is added to the experiment. We introduce the role of a 'tester' of this simple world. The tester has the ability with the touch of a button to invoke the interrupt on the actor as detailed earlier.

Table 1. How interrupts affect colour trace (mood) and the goal of the actor

Initial		After Interrupt	
Colour	**Goal**	**Colour**	**Goal**
Green	Reading	Grey	Musing
Grey	Musing	Red	Discussing
Red	Discussing	Green	Reading

Table 2. The mood data that is symbolically grounded in the agents 'Observer' memory. For example, the observer can always conclude a bad mood by the abstract link to the colour red.

Initial Colour Trail	Abstract Link
Green	Good Mood
Grey	Fair Mood
Red	Bad Mood

Table 3. The GPS data that is symbolically grounded in the agents 'Observer' memory. For example, the observer can always conclude an agent is reading when in the specified geographical area.

Room Coordinates	Abstract Link
$(0 \leq x \leq 0.45)$ AND $(0 \leq y \leq 0.45)$	Reading
$(-0.45 \leq x \leq 0)$ AND $(0 \leq y \leq 0.45)$	Musing
$(-0.45 \leq x \leq 0.45)$ AND $(-0.45 \leq y \leq 0)$	Discussing

5 Goals

The goals of this experiment are as follows:

1. Investigate the feasibility of representing Jin and Bell's simplified theory of understanding in a suitable practical platform.
2. Having successfully implemented a framework for this theory is it possible to simulate agent reasoning on a basic level? This aspect gives a two-fold question in the preliminary investigation:
    a. Can the observer successfully link colour to the mood of an actor and thus tell the tester what mood the actor is in?
    b. Can the observer successfully link physical location to GPS coordinates sent by the actor and thus tell the tester what activity the agent is engaging in?
3. Given this scenario, is there enough richness in the groundwork that has been laid to further develop this scenario? For example, can the observer reason on how an interrupt affects the mood and activity goals sought. Can the observer anticipate how an activity will affect the mood of the actor and intervene?

6 RESULTS

Results have been promising. The scenario outlined by Jin and Bell has been created approximately on the Webots platform (Fig. 3.). A trace has been successfully set up to show the actors colour trail (Fig. 4.) and a basic memory bank has been created to help the observer link, through abstraction, moods and activities based on colours and coordinates. A basic graphical user interface has been created to allow the tester to interrupt the actor at will and also ask the observer questions (Fig. 5.).

Observer correctly identified the actor's moods at each query point as well as the activity area the actor is currently engaged in. These results are preliminary in a wider research project, so it is important to assess them with regard to what basis they form for understanding and what groundwork they have laid to take this project to the next level. It is clear that by introducing classification into our analysis we can mine the output data and pose more meaningful questions to our agent. The prospect of this is very promising and it is anticipated a rich state space will be created from this data.

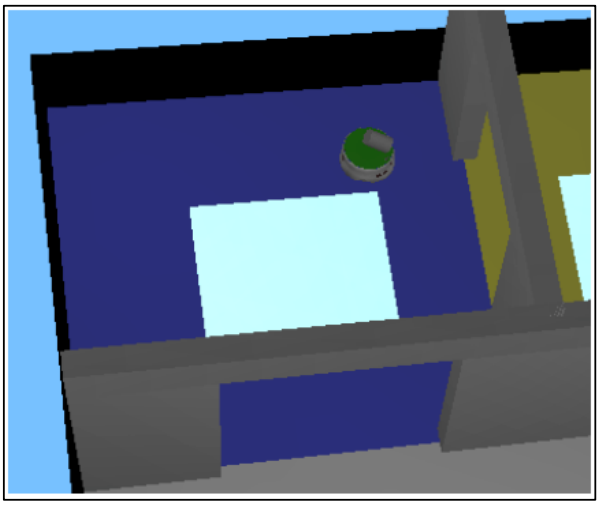

Figure 3. The agent in action on the Webots platform. The agent is currently in the musing area.

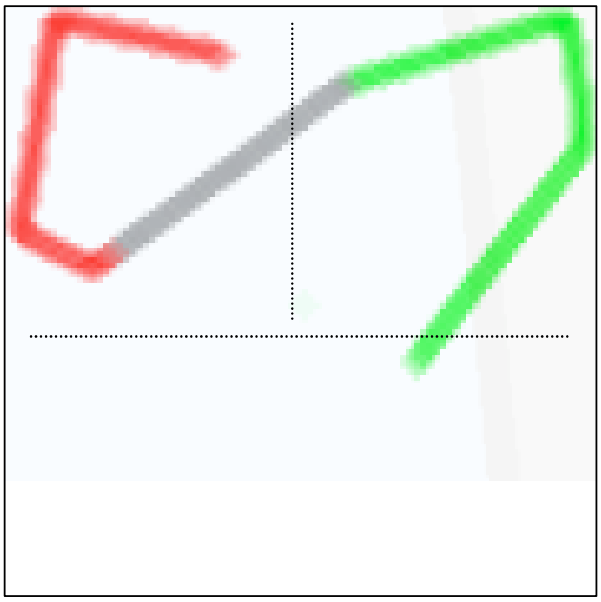

Figure 4. The trace of the agent. This trace shows the agent initially starting in the discussing area with a good mood, as the agent moves into the reading area, mood does not change. Mood changes as the agent moves to the musing area to a fair mood and then to a bad mood signified by the red colour.

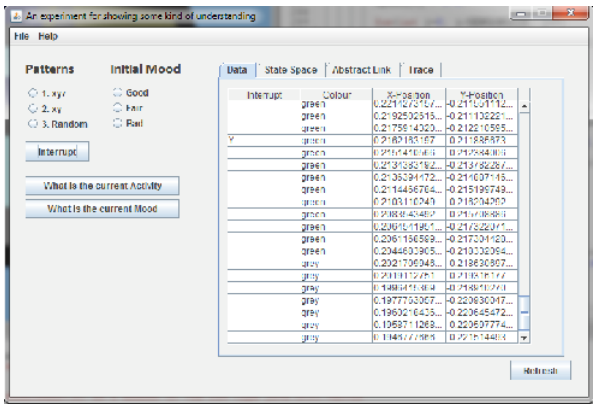

Figure 5. The basic user interface detailing the tester's options to interrupt and ask questions. Also shown is the data that the actor feeds to the observer for recording in its memory.

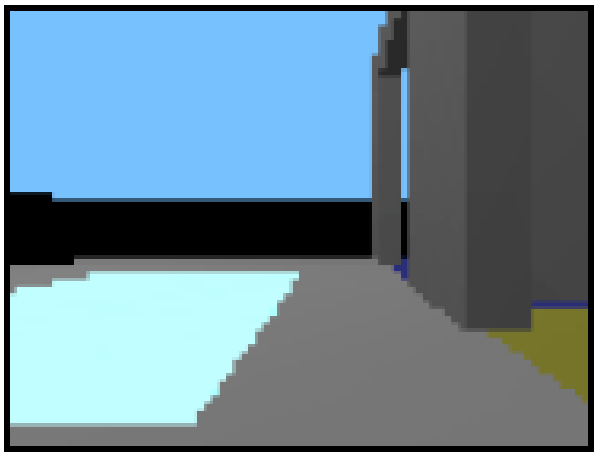

Figure 6. The view from the camera on board Khepera. While not in use in this experiment, it is clear to see that future experiments could reap the rewards of having a real 'observer' on board.

7 CONCLUSIONS & FUTURE WORK

It is hoped to extend the user interface for the 'tester' to ask deeper questions such as 'how do you know you are reading' and 'why do you want to read'. It is intended that the observer will draw from its large memory bank and make links back to the abstract data the actor had passed on as in previous straightforward questions while also incorporating rough sets and classification to create associations and rules which will help form a state space.

A camera has been attached to the agent (Fig. 6.) but is defunct in this preliminary experiment. It is hoped that utilising this camera will give a true meaning to the 'observation' taking place. This will merge two important fields of artificial intelligence (Image processing and robotics). Early ideas include determining when the actor is engaging in an activity based on imagery rather than GPS coordinates and relaying the captured images back to the observer to make sense of it. The observer will of course have still images in its memory bank to compare with.

There is a big emphasis to recreate the initial state model that the actor was defined to take part in. By building the model up and including how activities and interrupts affect moods, it is believed that this will create understanding on a basic level by linking back observational reasoning to exact predestination.

One final thought with regard to the direction of future research is the interesting topic of reinforcement learning. To have an agent reason similar to this experiment while displaying elevated examples of reinforced learning could be of prime interest in furthering understanding in autonomous agents as well as learning methods. An agent that explains to its tester why it has chosen a certain path through reasoning as well as understanding why it has been rewarded or punished is an exciting prospect. If one were to throw in the ability to affect mood similar to this experiment, it would be interesting to see the results!

REFERENCES

[1] Durkheim E, Mauss M (1903),' Primitive Classification'
[2] Weng J (2005),Muddy tasks and the necessity of autonomous mental development, in Proc 2005 AAAI Spring Symposium Series, Developmental Robotics Symposium, Stanford University.
[3] Hameroff SR (1994), Quantum Coherence in Microtubules: a Neural Basis for Emergent Consciousness, Journal of Consciousness Studies, 1, pp. 91-118.
[4] Harnad S (2003) To Cognize is to Categorize: Cognition is Categorization, ProcUQàM Summer Institute in Cognitive Categorisation, http://www.unites.uqam.ca/sccog/liens/program.htm
[5] Harnad S (1990), The Symbol Grounding Problem, Physica D 42:335-346
[6] Jin Z, Bell D (2003), An experiment for showing some kind of artificial understanding, Expert Systems, Volume 20 Issue 2, Pages 100 – 107.