

A Modelling Framework for Functional Imagination

Hugo Gravato Marques¹ and Owen Holland² and Richard Newcombe³

Abstract.

Imagination is generally regarded as a very powerful and advanced cognitive ability. In this paper we propose a modelling framework for what we call functional imagination: the ability of an embodied agent to simulate its own behaviors, predict their sensory-based consequences, and extract behavioural benefit from doing so. We identify five key components of architectures for functional imagination, and claim that they may be both necessary and sufficient. We outline a typical architecture, explain the flow of control within it, and describe a typical testing scenario using nested physics-based robot models. We also show how malfunctions within such an architecture may produce effects reminiscent of those found in certain human pathologies.

1 INTRODUCTION

Imagination has been regarded in Western philosophy as a very useful and significant cognitive ability. Prominent thinkers like Aristotle, Descartes, Hume, Kant and Sartre have all made contributions to the subject. In spite of this, however, little light has been shed on the mechanisms underpinning imagination [37]. One of the reasons why this might be so is that very often discussions about imagination are in fact discussions about imagery. It is true that the concepts cannot be completely dissociated, but it is important to be clear about the differences between them.

Imagery usually refers to an internally pictured object or situation ([38]); historically, it may be mentioned either in a phenomenological or a representational context. Phenomenologically, imagery is referred to as being capable of triggering experiences (or sensations) that resemble our experiences of daily life. This phenomenon is often called *quasi-perceptual* experience. The imagery of a dog can trigger a range of sensations one might have when one is actually close to a dog: the feeling of touching its rough coat, the sound of barking, or the dog-like smell. Representationally, imagery is strongly associated with an image-like representation ([38]). Imagination, on the other hand, is the process responsible for producing imagery. All references to imagination are references to things that are not present to the senses. However, some experiences of non-existent reality are not produced by imagination (e.g. the experience of an after-image or *phosphenes*). In addition, Penfield has shown that coherent complex experiences can be

generated artificially simply by inducing currents in neurons located in specific areas of the brain ([26]).

Several theories have tried to explain the process of imagination through the phenomenological and representational aspects of imagery. The theories of Descartes and Hume are examples of such attempts. Descartes's dualist theory relied on a mystical soul to explain the process of imagination. Hume eliminated the reference to an immaterial entity by transforming the workings of the mind into a mechanical system ruled by laws of association, where each idea (in Hume an idea is basically an image) is brought to consciousness through the principles of resemblance, causality and contiguity.

We have discussed elsewhere several reasons why these two theories failed to explain the workings of human imagination ([16]). Here, it will suffice to say that it would be very hard to model the process of imagination solely in terms of the phenomenological or representational aspects of imagery simply because we are not yet at a stage of being able to access them scientifically. For this reason, we propose to focus on a third, and more recent, aspect of imagery - the neuroscientific aspect. We will try to identify what happens in the body and in the brain while imagery and imagination are occurring.

In neuroscience, several experiments suggest that roughly the same areas of the brain are active when a subject is sensing (or acting) overtly and when he doing so covertly. Although it is far from being universally accepted by neuroscientists, several experiments suggest the involvement of the early visual cortex during visual imagery [13]. In one experiment, subjects were asked to close their eyes and to visualise a set of previously seen striped patterns [12]. Results show that areas as early as area 17 and 18 of the visual cortex were active during imaging. In the same set of experiments the normal functioning of area 17 was disturbed in order to investigate the role it plays in imaging. Results showed that, by disrupting the normal operation of area 17, performance on the imagery (and perceptual) task was impaired.

Furthermore, recent studies of motor imagery suggest that motor imagery is functionally and anatomically related to motor execution. An fMRI study of finger movements showed significant activation of the SMA (supplementary motor cortex) and the PMC (premotor cortex) during both execution and imagery [14]. In the same study, the M1 (primary motor cortex) and S1 (somatosensory cortex) showed less activation during imagined finger movements. These results were confirmed by a similar study carried by Porro [27].

Psychological studies have also shown a striking relation between overt and covert behaviour. Shepard and colleagues

¹ University of Essex, UK, email: hgmarq@essex.ac.uk

² University of Essex, UK, email: owen@essex.ac.uk

³ University of Essex, UK, email: ranewc@essex.ac.uk

have shown that the time taken to manipulate objects mentally seems to be linearly dependent on the number and extent of movements (or operations) made [30]. Subjects were asked to see whether, by folding a piece of paper in various indicated ways, two of its edges could be brought together. The results indicated that the time to reach a conclusion is dependent on the complexity of the folding process. The same sort of conclusion was reached in a mental rotation experiment where subjects were asked whether a given 3D object can be rotated to match another object [31]. The results indicated that the time to mentally rotate the image was proportional to the angle of rotation.

In addition, there also seems to be a connection between muscular activity during imagery and overt perception. In a study aiming at comparing eye movements during overt viewing and visual imagination, subjects were asked to look at the same irregularly-checked diagram four times (the diagram was rotated by 90 degrees between trials) while their eye movements were being recorded [2]. The eye movements were also recorded while subjects were imagining each of the four patterns. The results showed a close relationship between eye movements during overt and covert viewing suggesting also that eye movements reflect the content of what is being seen and what is being imagined. Similar results have been observed in different sensory modalities. An experiment testing the capabilities of subjects to imagine a smell have shown that it is very difficult (if not impossible) to image a smell without overtly sniffing [1].

All these results, and many more, suggest a strong relationship between imagination and the body, which highlights the possible relevance of embodiment theories to help explain this aspect of human cognition [40], [3]. However, it is only recently that ideas of embodiment have penetrated into Artificial Intelligence (AI). Previous attempts to capture the apparently abstract nature of human thought in AI were implemented within a symbolic framework. The General Problem Solver is an early and clear example of such an attempt (see for example [24], [25], [23]). In the General Problem Solver a group of operators could manipulate a collection of logical or mathematical expressions in order to find a solution to a given problem (e.g. a mathematical proof, or finding an analogy). The basic idea behind these reasoning systems relied on evidence from introspection and verbal protocols that humans seem to be able to manipulate propositions in order to form plans of action that can then be executed overtly. In spite of the failure of the Physical Symbol System Hypothesis to find support in neuroscience, logic-based approaches continue to dominate AI, but the lack of success in dealing with real world systems is finally turning attention towards the ideas of embodiment that are close to achieving dominance in cognitive science.

In this research programme, we want to use experimental results from neuroscience and psychology in order to produce a qualitative model (or candidate architecture) of human imagination. We do not aim to capture every single cognitive ability in which imagination might be involved, nor to account all the phenomena that seem to be related to imagination, such as dreaming, free-associative thought, etc. For this reason we will focus on what we call functional imagination: the mechanism that allows an embodied agent to simulate its own behaviours, predict their sensory-based consequences,

and extract behavioural benefit from doing so [17] (see below). Through the construction of increasingly architectures from simple components we hope to be able (1) to show that they work, (2) to establish parallels between the way they work and experimental evidence from humans (or other animals), and (3) to investigate malfunctions in the models caused by missing or defective components and compare them, if possible, with disorders found in humans.

The remainder of the paper is structured as follows: in Section2 we will summarise some work in AI and robotics relevant to the topic of imagination; in Section3 we will outline the concept of functional imagination; in Section4 we will present one of our models for functional imagination; in Section5 we will describe an experiment using a complex and dynamic physics-based simulator to study the model; in Section6 we will make some qualitative comparisons between our model and some of the experimental data shown above; finally, in the Section7 we will make some concluding remarks.

2 CURRENT RESEARCH

Some research in AI is clearly relevant for a theory of imagination. Shanahan, for example, is creating architectures based on Baar's Global Workspace Theory that allow inner rehearsal of actions by a (computationally) embodied agent [29]. From the learned associations between sensorimotor patterns of neural activation, a simulated agent is able to extract the consequences of certain actions through inner rehearsal and select actions that lead to rewarding behaviour.

Ziemke and colleagues have been exploring the sharing of sensorimotor structures for driving a simulated Khepera robot around a room in the absence of sensory information. They implemented a wall following algorithm for driving a Khepera robot around a room in order to discretize the environment into different categories (e.g. corners, corridors, etc). At the same time they trained a recurrent neural network to predict the next category given the current one. Then by feeding the neural network with its own predictions they showed that the robot was able to 'imagine' itself driving around the room [33].

Using a broadly similar approach, Stein introduced MetaToto, an upgraded version of Mataric's robot Toto [19], which was able to navigate in an unknown environment and add new locations (nodes) to a dynamic map (graph). MetaToto was capable of goal-driven navigation using known landmarks; more interestingly for our concerns, the robot was able to move to new and unknown locations using very crude descriptions of the environment [32]. By reusing the mechanisms for sensing and acting, MetaToto was able to generate sensory representations of what it would be like to be in those places, and was thus able to find its way to unknown locations.

Mel's Murphy, a real robot equipped with an arm and a video camera, was able to solve grasping problems using visual imagery [22]. The robot worked in two modes. In the first mode it moved its arm around until it found a way to grasp an object. During this training stage, associations were created between the movements of the arm and the image of the arm and the object recorded from the camera. After the connections were established Murphy was able to 'imagine' the grasping of objects using only its (visual) imagery capabilities.

Embodiment is central to all these research projects. From a disembodied perspective Thaler claims to have invented a ‘Creative Machine’ that can perform discovery and invention at the human level in fields as diverse as drug invention, car design, dance steps, musical compositions, etc. [36]. The Creativity Machine has two main components: an Imagination Engine for generating new ideas and an Alert Associative Center, for evaluating the ideas coming from the Imagination Engine. The Imagination Engine is an Artificial Neural Network (ANN), which can be trained on some body of knowledge and then perturbed internally with just the right amount of noise. Creative ideas are supposed to come from the associations made during training combined with the right amount of noise added to the response of the ANN. Unfortunately Thaler’s claims are very difficult to establish or assess from published data. For example, we could not find the mathematical parameters of the network providing the ‘right’ perturbation needed for potential ideas to arise, or any details on the way the evaluator actually operates. Other problems are the lack of external references to support claims such as: ‘[the] architecture emulates the thalamo-cortical loop in the brain (e. g., the seat of intelligence and consciousness) rather than blind [search]’ [10].

3 FUNCTIONAL IMAGINATION

As mentioned before, in our project we are focusing on building architectures that can exploit neuroscientific data for producing architectures for functional imagination. We define functional imagination in the context of artificial embodied agents as the mechanism that allows an agent to simulate its own behaviours, predict their sensory-based consequences and extract behavioural benefit from doing so (see [17]). By behavioural benefit we mean an increase in reward or utility achieved by using internal simulation. Here, we present what we claim to be five necessary and sufficient conditions for the presence of functional imagination in any embodied agent. We will briefly discuss each condition in order to introduce some of the components in our models. We have demonstrated sufficiency elsewhere using a working implementation of a minimal architecture where the 5 conditions were included [17]; in this paper we will concentrate on the actual operation of a simple architecture.

Condition 1: Sensorimotor-based prediction

An embodied agent should be able to predict the consequences of its actions in terms of sensory-based activations. This idea has been advanced by [7] [8] and [9] and offers a possible explanation of neural activations in the sensory and motor areas of the brain during covert behaviour. This condition implies the existence of sensory-motor mechanisms as well as a mechanism for predicting the sensory consequences of a motor action. In control theory such a mechanism is called a forward model. In general a forward model is a mechanism that predicts the next state of any system (the plant) given its current state and the current action. Forward models have been argued to be very basic mechanisms that evolved initially for anticipation in motor control [7]. However the new idea is that, if detached from the external sensory data, the

forward model can then predict the consequences of an action and substitute for the incoming sensory signal with its predicted value. If in addition to this an agent is able to select an action and inhibit it from overt execution, then the agent would be endowed with a sort of virtual world which could be detached from the external world, and in which actions could be tried covertly and their consequences predicted without further external information [6]. Dennett argues that an animal endowed with such a virtual world - a ‘Popperian creature’ - would have an evolutionary advantage over other creatures, because it would be able to try various risky ‘hypotheses’ of action without putting itself in real danger [4].

Condition 2: Goals

An agent must be able to execute goal-related behaviour. By goal-related behaviour we mean simply the ability to generate motor commands as a response to an internal state that might be changed as a result of the execution of those commands. This could be as simple as searching for food in response to hunger. In this situation the internal state of the agent is hunger (or some representation of the need for food), its goal is to reduce or eliminate hunger, the target of its action is food, and its behaviour is foraging.

McFarland distinguished between goal-directed, goal-achieving and goal-seeking behaviour [20]. A goal-directed system is one where the behaviour is guided by reference to an explicit internal representation of the goal to be achieved; for example, an explicit representation of the required percentage of stomach filling to be achieved. A goal-achieving system is one that can recognize the goal once it is arrived at (or that can at least change its behaviour once it reaches the goal), but where the process of achieving the goal is determined solely by the environmental circumstances. For example, an animal would keep foraging and eating until the stomach was full, when the signal from the full stomach would cause a switch in or cessation of the behaviour. Finally a goal-seeking system is one that is designed to approach the goal without the goal being explicitly represented within the system. A good example of this is a scheduling system that allocates a certain time slot for some particular behaviour (say, eating); when that time slot ends, some other behaviour is triggered.

For functional imagination, the goal is not required to be explicitly represented in the way required by a goal-directed system. Nevertheless, the goal needs to be recognized once the agent has arrived at it. The reason for this is that the usefulness of internal simulation must be measured in relation to a goal. Without the presence of a goal (be it implicit or explicit) internal simulation loses its functional value because there is no way of establishing whether it arrived at a useful result or not; it then becomes something closer to day-dreaming or the free association of ideas.

Condition 3: Evaluation

It is very hard (if not impossible) for an agent to behave in every situation in a way that maximizes its chances of survival and/or reproduction. For example, through lack of appropriate cognitive abilities, a dog might fail to identify the usefulness of a stick for taking food out of an otherwise inaccessible cage. If an agent was always able to behave optimally

according to some desired measure there would be no need for functional imagination. It is the possibility that the agent might produce behaviours that fail to achieve its goals that makes the role of functional imagination relevant. An agent therefore must in some way be able to evaluate its current state (be it real or imagined), which implies at least the capacity to distinguish whether a goal is fulfilled or not. As explained above this is a minimum requirement of any agent capable of either goal-directed or goal-achieving behaviour. Evaluations might be binary - stating simply whether a goal was achieved or not - or might take a range of values indicating the degree of satisfaction of the goal according to some measure (say, energy expenditure).

Condition 4: Action selection

An agent must be able to select actions (or motor responses) for internal simulation. Animals have a number of different tasks that have to be performed in order to enable their survival and reproduction (e.g. eat, drink, mate, etc). This means that they must be capable of producing different and appropriate behavioural responses in order to fulfil each task. This action selection is also necessary for dealing with simulated actions. If an animal was able to perform only one possible action, the usefulness of imagination would be restricted to the decision of executing the action or not, according to the evaluation of the consequences of the simulated action. In all other cases, functional imagination requires the agent to be able to try different actions in the same situation. As a minimum, this demands that the same action should not be repeated even though the state remains the same; a simple mechanism implementing inhibition of return [11] can fulfil this requirement.

Condition 5: Selection of sensorimotor-based state

An agent must be able to imagine situations that are not tightly tied to its current context. Selecting the scenario within which internal simulation is performed is essential to allow the agent to set its internal state independently of its current state. This would also allow the agent for example to simulate what it would have happened in a past situation if some other actions had been taken - enabling reflection and enhanced planning [34]. In addition, different states are produced during internal simulation as a result of simulating different actions and it will be useful - for example, in multi-step planning - for the agent to be able to select the state within which actions will be simulated. Without this mechanism an agent would be restricted to scenarios based only on its current state.

4 ARCHITECTURE

In our project to date we have implemented a variety of architectures. The reason for this is simply that we suspect (and therefore we want to show) that increasing imaginative ability comes at the expense of an increasing number and variety of components in the architecture. In order to differentiate and categorize these architectures we have created a taxonomy. We differentiate between architectures that reuse the same

sensory-motor structures for both overt and covert behaviour (economical architectures) and architectures that use copies of those structures for covert behaviour (duplicated architectures). We differentiate between architectures that overtly trigger the first solution they find for a certain problem (reactive architecture) from architectures that are capable of applying the best solution found within a given time (rational architecture). We also differentiate between architectures that can simulate only one step ahead (single-step architectures) and those that can cope with several steps ahead (multi-step architectures). Finally, we distinguish between architectures that retain and use memories of previous plans (memorizing architectures) and architectures that have to search afresh for a plan every time a given problem appears (memoryless architectures). Due to space limitations, we will present only one architecture here, an example which in our taxonomy would be classified as reactive, single-step, economical, and memoryless).

4.1 Architecture components

The architecture we will describe here is shown in Figure 1. When producing a model of functional imagination, one of the first questions to be answered is how the system can distinguish reality from fiction (what is being imagined) [28]). In our architectures we use a switch mechanism both to implement and to capture this distinction. As can be seen in Figure 1 the switch mechanism affects both the sensory and the motor systems. When the switch mechanism is set to ON the sensory system uses the information coming from the real world; when the switch is set to OFF the sensory system uses the information provided by the forward model. In addition, when the sensory system is set to ON the motor actions are executed overtly, while when it is set to OFF any motor action is inhibited from overt execution. In between the sensory and motor systems there are two selection mechanisms: one to select an action for execution, and another for selecting the feature (target) at which the action will be directed.

The short-term memory connected to sensory system (*State0 STM* mechanism) allows the agent to set the scenario (sensory state) that will be used as the starting point for the plan. In this implementation this state is the sensory state at the time the switch is set to OFF, but in other architectures this is not necessarily the case. In addition, because this is a single-step architecture, the state stored in this memory will be the starting point for the simulation of every action tried; it must replace the sensory states produced by the forward model after each simulated action.

The sensory mechanism is connected to an evaluation mechanism which allows the agent to evaluate its current state in relation to its current goal. In addition, the evaluation mechanism determines whether the current action and feature selection policies should be changed or not. A policy is a mapping from sensory states to the actions or plans that the agent ought to perform ([35]). If the evaluation is positive then the target and action policies are changed in order to make the decisions that led to the rewarding state more salient and more likely to be chosen in the future. If, on the other hand, the evaluation is negative the policy should be changed in order to allow for other actions to be preferentially selected. In this architecture a plan entails only one action because it is a

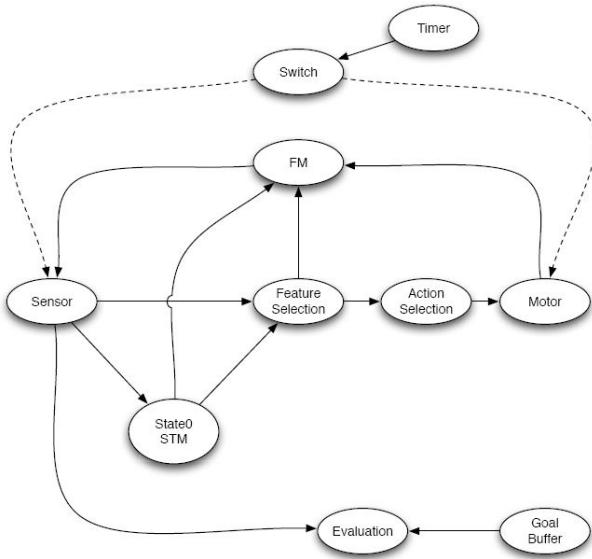


Figure 1. The architecture implemented.

single-step architecture.

Finally, we can see that the switch is connected to a timer. This allows the architecture not only to allocate the amount of time it will spend searching covertly for an action or plan, but also to interrupt the covert simulation (by setting the timer to OFF) in response to any unexpected event that might happen elsewhere in the system.

4.2 Control Flow

In addition to the components of the architecture we also present here the control flow of the activities in the system (see Figures 4 and 5). The architecture is divided into two main functional groups: one responsible mainly for controlling the switch (Figure4) and another responsible for controlling the behaviour (overt and covert) of the agent (Figure5). Each group works on its own internal loop but one is capable of influencing the other. A good analogy here is the way two threads in an operating system can run their own internal loops independently, but each influences the other by (for example) setting common variables.

In the implementation of the architecture the agent starts by being engaged in overt behaviour under the control of the sensory and motor policies until a goal is set in the system. In the current implementation the goal is set manually, but in an agent fully embodied in the world the goal could arise automatically - for example, the need for food, or for a mate. Once a goal arises in the system the agent performs the necessary steps for entering simulation mode: it stores the current state in the *State0 STM* mechanism, sets the timer to the amount of time allocated for solving that specific problem, and sets the switch mechanism to OFF, which inhibits the sensory and motor components from communicating with the external world. In this implementation the timer is set to a fixed time, but in an agent fully embodied in the world the time allocation for solving a problem should be dependent on some measure of significance of the goal in the current context

of the agent. Time allocation is an essential ability of any embodied agent that must fulfil multiple goals in order to survive [21] and the allocation of time to simulate and imagine is just an extension of that idea. After the initialization is complete this functional group does not do anything until the time for solving the problem expires.

Once the switch is set to OFF the architecture responsible for the behaviour stops whatever action was running and evaluates the current sensory state. From this point on this sub-architecture simulates actions covertly until the switch is turned ON. The simulation of one action starts by loading the sensory state stored at the time the problem arose. As mentioned above, the reason for this is that the architecture presented here is a single-step architecture (see above). After loading the state, the agent selects a target and an action and simulates the result of executing that action using the forward model. Once the action is complete the agent evaluates the resulting state; if the state achieves the current goal then the agent sets the timer to OFF and sets a flag indicating that it found a solution for the goal; otherwise it simulates another action.

Once the timer is OFF (either because the agent found a solution for its problem or because the time for solving the problem expired) the architecture responsible for controlling the switch sets the switch to ON. This will force the sensing mechanisms to receive information from the real world. In an embodied system, if no solution was found during the simulation period the agent should choose what to do next based on its current context (e.g. take more time to simulate, or perhaps do nothing until the situation changes); here, the architecture simply terminates stating that no solution was found. If, however, a solution has been found, the policies for the target and action selection are reinforced to bias them in favour of executing the action overtly.

Once the overt execution of the plan (one action) terminates, the final state of the (real) world is evaluated. If the evaluation is positive the problem is solved; otherwise, the architecture states that the plan was unsuccessful. Here, as before, the current context of the agent and the nature of the problem should determine what happens next.

For reasons of clarity, there is a part of the diagram that is not included in Figure5, which is connected to the decision point "Action complete" (see Figure2). This part deals with the problem of what happens if a goal cannot be achieved because the action in the plan was for some reason unsuccessful (e.g. perhaps due to a sudden change in the world). If an action during the execution of a single step or multi step plan was unsuccessful the plan becomes obsolete and the architecture terminates in a state of 'plan unsuccessful'. Once again, it should be the current context of the agent that determines what to do next.



Figure 2. Part of the work flow that was not included in Figure 5.

5 EXPERIMENT

In order to test our architecture, we used our physics-based humanoid simulator - SIMNOS [5]. In SIMNOS's body the skeletal components are modelled as jointed rigid bodies, with spring-damper systems at each joint. The rigid body limbs are fully contactable surfaces which allow the robot to interact with its environment. Each muscle is modelled as a single parallel spring-damper system with asymmetrical conditioning of the spring and damper constants, in order to allow the muscle to produce force only when is contracted (for more details see [5] and [18]). The robot has a monocular visual system (first-person view) that allows it to capture coloured images of its environment in a way similar to a camera mounted on a real robot. This simulated camera also provides the agent with the distance of each projected pixel.

For this experiment we used two instances of SIMNOS running in parallel; one to capture the interactions between the *real* agent and the *real* world, and the other to capture the covert interactions of the agent (the result produced by the forward model in the internal architecture). We will call the former the *real agent* and the latter the *virtual agent*. A similar approach where a second instance of a simulator is used as an internal model of the first instance can be found in [39].

5.1 The task

In our experimental setup, the real agent has in front of it a blue object and a red object on a table top (see Figure3). As can be seen in Figure3, the same general scene is loaded into the virtual agent, but the objects are not included. The real agent can visually explore the real environment by moving its head around. Every time an unknown object is found in the real world the virtual environment is updated by placing an object of the same colour in the appropriate position. The objects are distinguishable by their colour.

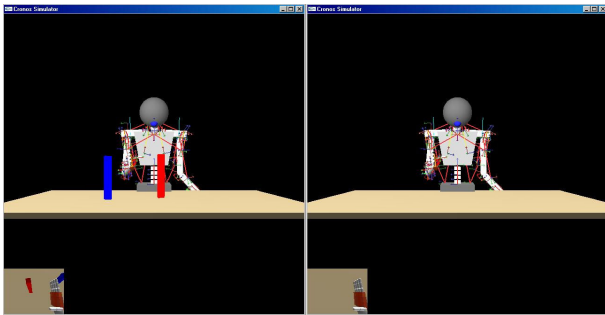


Figure 3. The experimental setup: the *real* world on the left and the virtual world on the right. At the beginning of the task the agent's virtual world (right) does not contain objects to interact with; they are added as the agent explores the real world.

Before we set the goal manually, the agent was given some time to explore its environment - enough to find the two objects and update its virtual world. The goal the agent was required to achieve was to move the red object further than a certain fixed distance. In order to solve the task the agent was endowed with two pre-programmed behaviours: one that allowed it to grasp an object, and another that allowed it

to grasp an object and throw it forward. The goal distance threshold was set to a value that could usually be exceeded when the agent executed the throwing behaviour on any of the objects. Once the goal was set, the architecture ran as described above until one of the end states was reached.

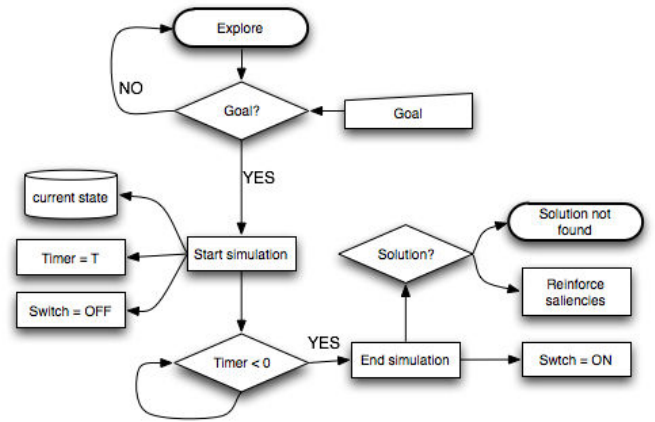


Figure 4. Control flow of the architecture responsible for controlling the switch.

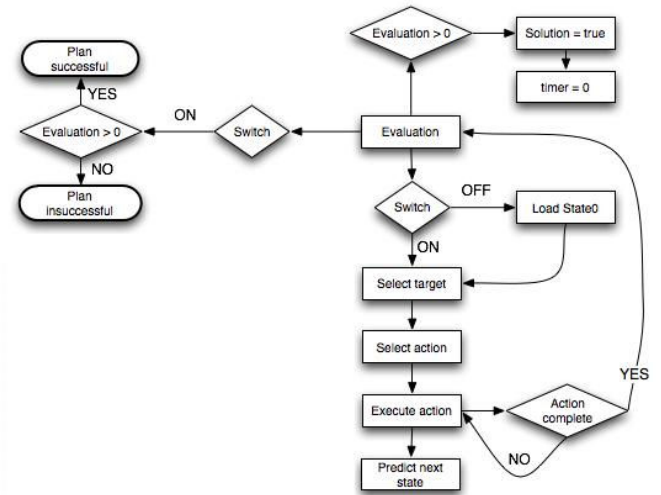


Figure 5. Control flow of the architecture responsible for the behaviour of the agent (overt and covert).

5.2 Selection mechanisms

We adopted a reinforcement learning strategy in order to form the target and action selection policies appropriate for reaching the goal. The reinforcement learning algorithm had to arrange that the red object and the throwing behaviour should be more salient than the blue object and the grasping behaviour respectively. This would allow the agent to throw the

red object forwards and exceed the distance threshold. In this single-step architecture the target and action saliences were modelled using two arrays, one for each selection mechanism. Because there were two objects, the feature salience array contained two values, one for each object. The same applied to the action salience array.

The salience arrays were initialised with low random values. Every time a negative evaluation was received (meaning that the goal had not been achieved) the feature and the action just selected were punished by decreasing the salience of each by a fixed amount. During the search, the probability of selecting the most salient item (feature or action) was set to a value of 0.75⁴. This meant that 25% of the times the item selected was the less salient one.

5.3 Results

The results of the experiment showed that the agent was reliably capable of solving the problem overtly when given enough time to search covertly. The behaviours implemented are not certain to be successful, and occasionally fail. For example, the agent might fail to grasp an object, or might drop the object when trying to throw it. Uncertainty is a feature of complex and dynamic environments (such as the real world, and also our simulator) and this is actually a good strategy for testing our model; it shows that it is capable of coping with deviations from perfection. In addition, the architecture was also able to detect when a solution was not found within the time slot allocated.

6 DISCUSSION

6.1 Activation of sensorimotor areas during covert behaviour

We started this paper by mentioning some experimental results showing the relations between overt and covert behaviour at the level of brain and behaviour. One of the themes was the appropriate activation of sensory and motor areas during imagery. In our model we have shown that executing actions covertly necessarily activates the sensorimotor mechanisms active during overt behaviour.

6.2 Overt and covert behaviour

Other results have shown that the time for performing a mental rotation depends linearly on the angle of rotation, and that the time taken to imagine folding a piece of paper increases with the number of operations that need to be performed. In our architecture, the time that it takes to imagine grasping a target depends on the distance of the target. Here, one could argue that by using a second instance of the same simulator as the forward model the time to reach a target overtly and covertly should actually be the same. This is true if one assumes the simulators run at the same speeds, which in our

⁴ This value is almost irrelevant in the current implementation as there are only two choices per selection mechanism. The only situation we want to avoid is that of a cycle where *feature1* and *behaviour1* are selected first, then *feature2* and *behaviour2*, and then back again to *feature1* and *behaviour1*. The use of a stochastic method for selecting the most salient behaviour ensures the diversity of the behaviour.

implementation they do. However, the linear dependence between the distance to the target and the time to grasp it would happen in any sort of model that could covertly produce data that is qualitatively similar to the real data. At the moment we are at a stage where we will start implementing the architectures in our real humanoid robot. Once this is done we will have more data to prove our point. For the same reason, time that it would take to imagine the execution of several behaviours would be dependent on the number of behaviours that need to be executed. Here, we only present data regarding a single behaviour, but this can be shown in the implementations of our multi-step architectures.

6.3 Defects and pathologies

The simple architecture we have described has a key component - the switch. It can malfunction in a number of ways, and some of these have intriguing parallels to some human pathologies. For example, it is supposed to be OFF during episodes of internal simulation; if it fails to operate, the imagined behaviour will be executed in reality. In humans, motor commands produced during dreaming are normally inhibited, but in some people, popularly called sleepwalkers, actions corresponding to dream elements are occasionally carried out, and several cases of murder have been defended in the American courts by claiming that the alleged murderer had been unconsciously acting out a dream. One eventual aim of this project is to systematically damage the architectures produced and to log the system pathologies, with the aim of comparing these to standard databases of human pathologies. Correspondences will not amount to proof that our models capture the processes of human imagination, but the comparison will at least reveal something about the goodness of fit.

6.4 Simultaneous overt and covert behaviour

We have argued elsewhere [15] that an architecture that reuses its sensory and motor systems for overt and covert behaviour cannot act covertly at the same time as it acts overtly. In fact, the question of whether we reuse the exact same neural structures for overt and covert behaviour or whether we use copies of those systems (which must be localized in the vicinity of the sensory and motor areas for the experimental results to hold) is an open question. We have already implemented architectures that deal with this problem, and allow an agent to act covertly and overtly at the same time and we hope to publish the results in the near future.

7 CONCLUSION

In this paper we have proposed a modelling framework for what we call functional imagination: the ability of an embodied agent to simulate its own behaviors, predict their sensory-based consequences, and extract behavioural benefit from doing so. We have identified five key components of architectures for functional imagination, and claim that they may be both necessary and sufficient. We have outlined a simple architecture, explained the flow of control within it, and described a

typical testing scenario using nested physics-based robot models. We have also speculated about how malfunctions within such an architecture may produce effects reminiscent of those found in certain human pathologies. This is ongoing work, as yet in its early stages, but it holds some promise of leading to a systematic synthetic approach to understanding the problem of imagination.

8 ACKNOWLEDGEMENTS

We would like to thank the Portuguese FCT for the PhD fellowship to Hugo Gravato Marques and to the EPSRC (GR/S47946/01) for funding the CRONOS project.

REFERENCES

- [1] M. Bensafi, J. Porter, S. Pouliot, J. Mainland, B. Johnson, C. Zelano, N. Young, E. Bremner, D. Aframian, R. Kahn, and N. Sobel, 'Olfactomotor activity during imagery mimics that during perception', *Nature Neuroscience*, **6**, 1142–1144, (2003).
- [2] Stephan Brandt and Lawrence Stark, 'Spontaneous eye movements during visual imagery reflect the content of the visual scene', *Journal of Cognitive Neuroscience*, **9**, 27–38, (1997).
- [3] Ronald Chrisley and Tom Ziemke, 'Embodiment', in *Encyclopedia of Cognitive Science*, 1102–1108, Macmillan Publishers, (2002).
- [4] Daniel Dennett, *Darwin's Dangerous Idea*, Harmondsworth, Allen Lane The Penguin Press, 1995.
- [5] David Gamez, Richard Newcombe, Owen Holland, and Rob. Knight, 'Two simulation tools for biologically inspired virtual robotics', in *Proceedings of the IEEE 5th Chapter Conference on Advances in Cybernetic System*, pp. 85–90, Sheffield, (2006).
- [6] Rick Grush, 'An introduction to the main principles of emulation: motor control, imagery, and perception', Technical report, UC San Diego, (2002).
- [7] Rick Grush, 'The emulation theory of representation - motor control, imagery, and perception', *Behavioral and Brain Sciences*, **27**, 377–442, (2004).
- [8] Germund Hesslow, 'Conscious thought as simulation of behaviour and perception', *Trends in Cognitive Science*, **6**(6), (2002).
- [9] Owen Holland and Rod Goodman, 'Robots with internal models: A route to machine consciousness?', in *Machine Consciousness*, ed., Owen Holland, Imprint Academic, Exeter, UK, (2003).
- [10] Imagination Engines Incorporated. Iei's patented creativity machine, 2005. [Online; accessed 27-August-2007].
- [11] Laurent Itti and Christof Koch, 'A saliency-based search mechanism for overt and covert shifts of visual attention', *Vision Research*, **40**, 1489–1506, (2000).
- [12] Stephen Kosslyn, 'The role of area 17 in visual imagery: Convergent evidence from pet and rtms', *Science*, **284**, 167–170, (1999).
- [13] Stephen Kosslyn, Giorgio Ganis, and William Thompson, 'Neural foundations of imagery', *Nature Reviews Neuroscience*, **2**, 635–642, (2001).
- [14] Martin Lotze, Pedro Montoya, Michael Erb, Ernst Hülsmann, Herta Flor, Uwe Klose, Niels Birbaumer, and Wolfgang Grodd, 'Activation of cortical and cerebellar motor areas during executed and imagined hand movements: An fmri study', *Journal of Cognitive Neuroscience*, **11**(5), 491–501, (1999).
- [15] Hugo Marques and Owen Holland, 'Minimal architectures for embodied imagination', in *In Proceedings Brain Inspired Cognitive Systems (BICS2006)*, (2006).
- [16] Hugo Marques and Owen Holland, 'Architectures for imagination: why models matter', (2007). Unpublished.
- [17] Hugo Marques and Owen Holland, 'Architectures for embodied imagination', *Neurocomputing*, (2008). Submitted.
- [18] Hugo Marques, Richard Newcombe, and Owen Holland, 'Controlling and anthropomimetic robot: A preliminary investigation', in *Proceedings of ECAL2007*, Lisbon, (2007). Springer Verlag.
- [19] Maja Mataric, 'Integration of representation into goal-driven behaviour based robots', *IEEE Transactions on Robotics and Automation*, **8**(3), 304–312, (1992).
- [20] David McFarland, 'Goals, no-goals and own goals', in *Goals, No-Goals and Own Goals: A Debate on Goal-Directed and Intentional Behaviour*, eds., Alan Montefiore and Denis Noble, Unwin Hyman Ltd, London, (1989).
- [21] David McFarland and Thomas Besser, *Intelligent Behavior in Animals and Robots*, The MIT Press, 1993.
- [22] Barlett Mel, 'Murphy: A robot that learns by doing', in *Neural information processing systems*, American Institute of Physics, New York, (1988).
- [23] Allen Newell and Simon Herbert, 'Gps, a program that simulates human thought', in *Computers and Thought*, eds., Edward Feigenbaum and Julian Feldman, McGraw-Hill, (1963).
- [24] Allen Newell, J. Shaw, and Herbert Simon, 'Report on a general problem-solving program.', Paris, (1959). Proceedings of the International Conference on Information Processing.
- [25] Allen Newell, J. Shaw, and Herbert Simon, 'A variety of intelligent learning in a general problem solver', in *Self Organizing Systems*, eds., Yovits and Cameron, Pergamon Press, (1960).
- [26] Wilder Penfield, 'Some mechanisms of consciousness discovered during electrical stimulation of the brain', *Proceedings of the National Academy of Sciences*, **44**(2), 51–66, (February 15 1958).
- [27] Carlo Porro, Maria Francescato, Valentina Cettolo, Mathew Diamond, Patrizia Baraldi, Chiava Zuiani, Massimo Bazzocchi, and Pietro Prampero, 'Primary motor and sensory cortex activation during motor performance and motor imagery: A functional resonance imaging study', *The Journal of Neuroscience*, **16**(23), 7688–7698, (1996).
- [28] Jean-Paul Sartre, *Imagination: A Psychologic Critique*, The University of Michigan Press, 1962.
- [29] Murray Shanahan, 'Cognition, action selection, and inner rehearsal', *Proceedings IJCAI 2005 Workshop on Modelling Natural Action Selection*, 92–99, (2005).
- [30] Roger Shepard and Christine Feng, 'A chronometric study of mental paper folding', *Cognitive Psychology*, **3**, 228243, (1972).
- [31] Roger Shepard and Jacqueline Metzler, 'Mental rotation of three-dimensional objects', *Science*, **171**, 701–703, (1971).
- [32] Lynn Stein, 'Imagination and situated cognition', Technical Report 1277, MIT AI Lab, (1991).
- [33] John Stening, Henrik Jacobson, and Tom Ziemke, 'Imagination and abstraction of sensorimotor flow: Towards a robot model', in *Proceedings AISB2005 Symposium on Next Generation Approaches to Machine Consciousness*, pp. 50–58, Hatfield, UK, (2005).
- [34] Susan Stuart, 'The binding problem: Induction, integration and imagination', in *AISB2005: Proceedings of the Symposium on Next Generation Approaches to Machine Consciousness*, Hatfield, UK, (2005).
- [35] Richard Sutton and Andrew Barto, *Reinforcement Learning*, The MIT Press, 1998.
- [36] Stephen Thaler, 'Neural networks that autonomously create and discover', *PC AI*, (1996).
- [37] Nigel Thomas, 'Imagining minds', *Journal of Consciousness Studies*, **10**(11), 79–84, (2003).
- [38] Nigel Thomas, 'Mental imagery', in *The Stanford Encyclopedia of Philosophy*, ed., Edward N. Zalta, (2005).
- [39] Richard T. Vaughan and Mauricio Zuluaga, 'Use your illusion: Sensorimotor self-simulation allows complex agents to plan with incomplete self-knowledge', (September 2006).
- [40] Tom Ziemke, 'What's that thing called embodiment?', in *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, Mahwah, NJ, (2003). Lawrence Erlbaum.