Salient Features and Snapshots in Time: An Interdisciplinary Perspective on Object Representation

Veronica E. Arriola-Rios[1,*], Zoe P. Demery[2,*],
Jeremy Wyatt[1], Aaron Sloman[1], and Jackie Chappell[2]

[1] School of Computer Science, University of Birmingham,
Edgbaston, Birmingham, B15 2TT, UK
{v.arriola-rios,j.l.wyatt,asloman}@cs.bham.ac.uk
[2] School of Biosciences, B15 2TT, UK
{zxd878,j.m.chappell}@bham.ac.uk

Abstract. Faced with a vast, dynamic environment, some animals and robots often need to acquire and segregate information about objects. The form of their internal representation depends on how the information is utilised. Sometimes it should be compressed and abstracted from the original, often complex, sensory information, so it can be efficiently stored and manipulated, for deriving interpretations, causal relationships, functions or affordances. We discuss how salient features of objects can be used to generate compact representations, later allowing for relatively accurate reconstructions and reasoning. Particular moments in the course of an object-related process can be selected and stored as 'key frames'. Specifically, we consider the problem of representing and reasoning about a deformable object from the viewpoint of both an artificial and a natural agent.

Keywords: Representations, Learning, Exploration, Cognitive Agents, Animal Cognition, Deformable Objects, Affordances, Dynamic Representation, Salient Features.

1 Introduction

The cognitive architecture of any animal or machine (jointly 'agents') has limits, so it cannot contain a perfect model of the dynamic external and internal world, such as about all matter, processes, affordances, or more abstract concepts, like 'mind' or 'spirit'. Every agent receives a particular amount of data through its sensors. How useful that data is in the short or long term depends on the environmental conditions, how accurately the data might be processed into information, and the agent's behavioural response. Frequently, an agent should maximise the amount of meaningful, relevant information it can obtain about its surroundings, while minimising the energy expended, but this is highly dependent on the nature of the agent [1]. This applies not just to a static snapshot of time, but also to a constantly changing world with a past, present and future, where being able to predict events, or select between alternative actions without actually trying

* Joint first authors.

G. Dodig-Crnkovic and R. Giovagnoli (Eds.): *Computing Nature*, SAPERE 7, pp. 171–184.
DOI: 10.1007/978-3-642-37225-4_10

them, may be useful for the agent. So in these circumstances, what are the most useful elements for the agent to store and process in its cognitive architecture and how may they best be coded? Principally, we propose that when an agent gathers information through its senses, often it may form object representations supported by exploration[1].

To date in the field of animal cognition (AC), there has been surprisingly little systematic, quantitative research on exploration, and how it could support learning mechanisms in different agents (see [3] for more discussion). What research there is, has largely been on humans and focussed on Bayesian network learning (e.g. [4]). Among the non-human animal researchers, the focus has been on *what* the different cognitive capacities of different species are, rather than *how* they actually process information to achieve those capacities [5]. For example, the 'trap-tube task' is a typical litmus test for causal understanding of gravity (e.g. [6]). It has revealed a lot about many species, but it is just a binary measure of whether an individual can complete the task or not. No one has fully investigated why one individual can succeed at the task, while another fails – is it something about their different exploratory strategies? Moreover, although quite complex-looking actions can often be performed by agents with simple mechanisms and small neural architectures (e.g. [7]), they may not be able to *generalise* these actions to other similar, but novel circumstances. Thus in this paper, we are concerned with more complex, flexible agents. Another area consistently ignored in AC, but one which may provide answers, is how the senses support exploratory learning (e.g. [8]).

It is a blossoming area in Artificial Intelligence (AI) however. Robots force us to explicitly define the model design, suggesting concrete, testable hypotheses for AC. However, we believe there is not yet a robot/simulation that can flexibly abstract concepts, or generalise knowledge to new situations. AI has looked at different learning mechanisms in isolation with relative success, but few projects have tried combining them into one agent (e.g. [9]). Therefore, AC behavioural experiments can provide realistic biological constraints and inspire more integrative cognitive modelling in AI.

We would like to propose that when exploration of objects occurs for forming representations, it is not always random, but also *structured*, *selective* and *sensitive* to particular features and salient categorical stimuli of the environment. Also that it can follow through three stages of theory formation – the forming, the testing and the refining of hypotheses [10]. Each hypothesis may be specific to a particular group of affordances or processes ('exploratory domains'), but they may also be generalisable to novel contexts. We introduce how studies into artificial agents and into natural agents are complementary [10], by comparing some findings from each field.

First, we will take a top-down approach to explore what some of the general environmental constraints imposed on an agent's system when internalising the world around it may be. Then we will look at some of the possible mechanisms to solve these problems, particularly in the visual domain of object representation. There are several

[1] Cognition does not always rely on internal representations and the degree of detail in any internal representation can vary greatly depending on the situation. For instance, there can be a lack of detail especially when the environment can largely control an agent's behaviour, such as in flocking behaviour or in using pheromone trails. Here alternative, but complementary, mechanisms may be more relevant, such as emergency or embodiment, but in this paper we will not consider these cases[2].

methodological problems in computer vision research, including recognition, tracking and mental imagery [11]. Within robotics, we present an approach where simulations of real objects, calibrated from real-time environmental data, can be used as artificial mental imagery. We have exploited a combination of key features from image analysis, computer graphics and animation, as well as aspects of physical models, to generate an internal representation of a deformable object with predictive capabilities. Finally, we will consider the degree of ecological validity of this model by comparing it with AC behavioural findings about parrots, who are notoriously exploratory and playful throughout their lives.

2 Requirements for the Agent-Environment Interaction

An agent interacting with its surrounding environment often combines perception and analysis with action. It can also be driven by its goals, which can be quite explicit, like foraging for survival, or particular problem-solving tasks. Or they can be quite implicit, such as to gather information by apparently random exploratory behaviour. Shaw [12] suggests, *"The chief end of an intelligent agent is to understand the world around it."* Here, the word 'understanding' implies the agent's ability to make predictions about the world. For this to take place, the agent should be able to detect the consistent properties or salient features in its environmental niche. These properties allow a link to form between the agent and the environment. We will now consider what some of these primordial properties might be (see also [13]).

2.1 Redundancy

Given the inherent limitations of the agent, it will only be possible for it to gain a partial understanding of its surroundings[2]. This partial understanding may not allow the agent to make perfect predictions for all environmental events, so it cannot always be ready to process useful information. As it detects sensory data, it also may not succeed at processing relevant signals. Therefore, we expect there may be errors and inexactitudes at different levels of the agent's perceptual or analytical processes. It may thus be useful for its system to be able to tolerate this margin-of-error. Some agents often have more than one mechanism to find things, solve problems, or to perform actions. The agent could just react according to different layers of data filtering, or it could use one or a combination of different learning mechanisms [10]. While qualitatively different, all of these mechanisms produce similar, valid results. In this sense, we call these different possible mechanisms 'valid ways', and say they are 'redundant'. Therefore, redundancy allows the agent to ignore irrelevant data, or to reconstruct faulty perceptions from new perceptions that convey the same information.

2.2 Consistency

When multiple methods are used to collect or analyse data, they can act in a complementary way, and contribute by providing different information. Alternatively, they may

[2] An artificial agent (e.g. a virtual automaton) in a very simple environment can make perfect predictions; but we are not concerned with these cases.

be superfluous; in which case, they confirm previous findings. For an agent, different methods of perceiving or deducing the same thing should be consistent with each other, if there is enough knowledge. An agent that sees a pen while touching it, should gain tactile information in accordance with the position and surface of the image it sees. If there is a fault in the synchronisation between this visual and tactile information, the agent will not be able to properly integrate this information, or accurately describe the object. This principle is present in human mathematics: different methods used to solve the same equation, *must give the same answer*.

2.3 Persistency

For an agent to be able to make relatively accurate predictions about the environment, there should be at least a few unchanging rules in the environment for a significant period of time. These rules are useful for the agent's internal knowledge and learning mechanisms. The strongest examples can be found in mathematics and physics. In order to develop the cosmological theories of physics, it is necessary to assume that the physical laws that rule at the present time on planet Earth, are the same rules that applied during the Big Bang and in galaxies far beyond ours. Agents should respond in the same way to the environment. During complex actions, an agent may change or modify their goals and plans. Even then though, they should make the changes according to a particular, foreseeable pattern, which may be rooted, for example, in their brain structure. If agents do not follow persistent rules, their behaviour is erratic and unpredictable.

2.4 Regularity

This is the predictable presence of previously perceived features or classes of them, due to a fixed relationship between occurrences[3]. There should be persistent patterns in the environment, allowing at least for partial predictions, particularly when an agent is faced with different causal problems. Causality is a manifestation of regularity, where the partaking elements are not always identifiable, but whose manifestation always entails the same consequence. Thus, agents should have mechanisms capable of detecting these patterns to take advantage of them. Then the environment could be categorised using a finite amount of key features linked by predictive relationships, including elements representing continuous features. For example, a small mathematical equation can describe an infinite parabola.

Sequentiality. This is a particular form of regularity, but in a universe with only one temporal dimension, it becomes especially relevant. Sequentiality is the presence of a series of features of two or more elements that are nearly always perceived in the same total or partial order[4]. The first features can be used to identify the sequence and predict either the following features, or the rules set needed to process them. Some examples include: identify a command and know which actions to execute; analyse the first items of a numerical sequence and predict the next; listen to the first notes of a

[3] This can be present in different dimensions, or in a hierarchical structure.

[4] These may not be contiguous and can include cause-and-effect learning.

song and remember how to sing the rest (which was memorised in advance); identify the beginning of a question and prepare to understand it to look for the answer; or listen to the sound of a prey and prepare to chase.

Structure with Partial Order and Layers. There could also be a succession of *sub*-sequences. The connections here would only allow a few options to follow, such as beginnings of other sub-sequences. This forms a branching structure, which becomes layered, modular, and, in some cases, hierarchical [14]. The maximum length of an existing sequence, and the maximum number of branches that can be remembered and manipulated, impose strict limitations upon what the agent can understand, and the types of patterns it is capable of detecting. However, this structure may allow more complex agents to make abstractions, as concepts formed at one stage could be re-used and refined to repeatedly form ever more complex concepts in multiple ways [1]. This allows for progressively specific and parallel processes (e.g. [15]).

2.5 Experience

For small and well-identified tasks, a largely pre-programmed agent may suffice. Little experience may be needed in a relatively static environment, such as where precocial animals, whose behaviour has been almost completely determined by their genome, just need to survive long enough to reproduce. Other agents are often required to adapt to diverse, dynamic environments, where a lot more learning is required (see [1] for greater discussion). The different extractions of relevant information (Section 2.1) would more likely be processed by mechanisms shaped and influenced by experience. The agent should seek out information to reinforce, evolve and, when possible, prove or disprove its current models, particularly if its expectations are violated. Depending on the needs and the competences of the agent, a specific, relevant subset of experiences would allow specific, relevant features of the individual's niche to be captured (e.g. [16]). We believe there is continual extension of these 'branches' or 'blocks of knowledge' throughout the life of a cognitive agent. At different ages or stages of development, an agent should take in different aspects of the same overheard conversation, for instance, or different aspects of the operation of the same tool.

2.6 Where Does This Leave Us?

All of the above described environmental features/constraints together form a structured universe. Parts of this structure may be perceived and understood by artificial and natural agents. The existence of regularities reduces the information needed to describe a part of the environment, as once enough elements and relationships have been identified, the rest can be inferred. Some animals may have the ability to identify 'valid ways' and describe them as 'formalisms'; sets of rules that can warrant good results when sufficient conditions are met [10]. This is essentially how science operates, particularly logic, mathematics and computer science.

Within the field of AI, some formalisms for 'knowledge representation' focus on the association of symbols to entities (i.e. objects, relationships and processes) in a structured way, such as 'Frame Languages' [17]. However others, like 'First Order Logic', incorporate powerful systems of deduction. These symbolic languages are extremely powerful for discrete reasoning, but they may not be particularly appropriate for describing continuous dynamics, or even for making predictions, such as when objects move through an environment. In AI, it is highly relevant to consider the amount and type of knowledge needed before an agent can be capable of processing it. How much does the agent need to know to be able to predict a few movements of different objects? Can that knowledge be learned from experience, or does it need to be pre-programmed?

In certain contexts, the minimum number of necessary elements to complete a description is known as the number of degrees of freedom. For example, given the generic mathematical formula that describe parabolae, only three points are needed to specify a single, infinite parabola. This principle can be directly applied in computer graphics. By making use of algebraic equations, an infinite amount of shapes can be approximated, represented and reconstructed with just a few polynomials [18]. Furthermore, transformations of these shapes can be encoded with mathematical formulae, thus allowing the representation of physical objects and processes; which can be used to implement a form of mental imagery.

Hence, whether the powerful deductive machinery is available in a natural or an artificial agent, it is important to define how we go from representations of continuous transformations, to discrete objects and events. As with the popular phrase, ‘a picture is worth a thousand words’, predicate logic may not be able to naturally represent 3D graphical information in a consistent, complete and compact description. It may be possible, however, to extract logical information from graphical simulations when required for symbolic reasoning. Here we give an example of how this could be achieved in AI by combining traditional animation techniques, computer graphics and physics, with symbolic representations.

We believe this approach may be more rigorous than the standard mechanism used in human brains. Humans can recognise things without being able to draw them [19], or use mental imagery without making exact simulations [11] (while our AI system requires them). This shows how we need to better understand the underlying mechanisms of natural agents processing and representing the world around them. Observations of natural exploration behaviour do provide realistic biological constraints on the design of AI models for object representation. We will investigate these issues in AC by running behavioural experiments on parrots, as our exemplar exploratory and adaptive species. Is there evidence of each of the environmental requirements/regularities described above being attended to by the parrots? Does their exploration behaviour suggest underlying strategies for processing and representing the environment?

3 Designing a Representation

3.1 Using Key Frames to Model Deformable Objects

The study of the perception and understanding of the affordances of deformable objects is particularly appropriate to illustrate the points outlined in the section above. The problem of representing solid objects, their properties and their related processes has been studied in great detail in computer graphics [20], and there have been attempts to generate representations using semantic information [21]. Within the first field, there are several good representations for many different types of shapes, most of them based on meshes, splines or quadrics [18]. The motion of objects is simulated with an approach analogous to traditional cartoon animations. There is a set of key frames, where a 'key frame' is a drawn snapshot in time defining the start and end points of any smooth transition, and all of the frames connecting them are called the 'inbetweeners'.

Currently, key frames are identified and drawn by humans; in traditional animation the most skilled cartoonists are responsible for them. Due to the smoothness of the transition between key frames, it is possible for a less-skilled cartoonist to interpolate the inbetweeners. In computer animation, the control points and curves defining the geometry and colours of the scene are set in the key frames. The transitions between key frames are mainly polynomial interpolations, or continuous mathematical transformations of these control elements [22]. To create realistic animations, movements are often captured from real objects. This is a very slow and expensive process [23]. In an attempt to automate the rendering of realistic movements and the inbetweeners, physics engines have been incorporated into the animation packages. They are also present in real-time virtual environments where interaction with a user takes place.

However, the incorporation of physics changes the dynamics of producing an animation slightly. Instead of interpolating between two key frames, the first key frame is given by a human designer and the simulation stops when a given condition is satisfied, thus automatically generating the inbetweeners and the final key frame. Note that predictive capabilities have been attained, and that the simulation is now required to specify the new parameters of the material. This includes mass, young coefficients or spring stiffness, in addition to the method's criteria, such as integration methods or time steps. Correctly estimating these parameters is a difficult problem.

Furthermore, while the simulations may look plausible to the human eye, they may not be physically accurate, so different models are required to simulate different materials and differently shaped objects. A natural agent's brain faces a similar computational problem, yet evolution largely seems to have solved it in a *qualitatively* different way. Humans can reason and make predictions about features of the world, but we probably do not simulate it in the quantitative way a physics engine does. It is still not completely clear how or what exactly the underlying mechanism is in various natural agents. Behavioural experiments can allow us to *infer* what is going on in an animal's mind. However, interpretation of the data is largely based on assumptions and only allows us to make <u>indirect</u> conclusions. Invasive techniques, such as particular neurophysiological or brain imaging methods, only provide partial information about the *content*, or even about the structure or neural representations, in an animal's mind. Thus, <u>if</u> done correctly, AI simulations can be very illuminating. We suggest that an initial list of problems an artificial agent needs to solve are:

1. Generate an internal representation of real deformable objects in the surrounding environment;
2. Identify key frames of the related environmental processes;
3. Interpolate (continuously or discretely) the links between frames;
4. Use previous knowledge to predict future events.

The automation of the animation process provides one solution for the first three points. Traditional animation techniques, however, cannot address the fourth point. The use of physics models and formal logics can address the two last points, but in this case the agent needs to select and calibrate the right model. It is still debatable whether physics models can correctly approximate all the ranges of processes observed in natural environments, given the inherent limitations of mathematical models to model real, complex deformations. Furthermore, there is still no model that integrates all of the points into one agent. Given the huge variety of possible affordances perceived by humans alone, we expect that some form of learning should be used to generate the model(s), which would provide the interpolating link between key frames and aid the artificial agent in making predictions. However, *which* type of learning mechanism is still open to question.

Here we present the advances of a preliminary, physics-based method, where a general (though not completely accurate) model of deformable objects is used [24], and an artificial agent learns to calibrate and use it in the way described above. The next step is to take the key frame representation of the object and extract symbolic ones from it. Then we need to take functions that describe the transformations, associate a symbol to each, and consider that symbol as referring to a categorised process or action. For several cases, this step should be quite straightforward, since the representation has already been discretised, grounded and categorised [25]. Then the already developed, symbolic-level machinery can be applied. Finally, the overall results can be compared with natural exploration behaviour (e.g. of parrots) for ecological validity. Is there evidence of similar mechanisms in natural systems? Is our model biologically plausible?

3.2 Representing the Object's Shape

Kakariki Experiment I: AC Implications for AI Models. When segregating the world around itself, we believe an agent first needs to identify and represent distinct objects. Then the agent needs to understand what the shape of each object means, i.e. its affordances when it interacts with the rest of the world. What are its physical properties? How can some of these properties be encoded in the memory of the agent? For instance, if two key elements are connected by a known relationship, anything in between is already implicitly represented. Contact points and segments of continuous curves can be approximated by lines and polynomials, and delimited by key points. Under this light, it is natural that an agent would be more interested in these points of discontinuity. Indeed, in our first AC experiment, we found that this does seem to be the case, at least for the New Zealand red-fronted parakeet or 'kakariki' (*Cyanoramphus novaezelandiae*).

We chose kakariki as our model animal species for investigating how agents gather and represent environmental information, as they are neophilic and have a high exploratory tendency throughout their lives. Moreover, as with many other parrot species, they are relatively intelligent and have an anatomy adapted to dexterity and fine object manipulation. We presented a series of novel objects of a range of different rigid shapes to 21 kakariki individually and recorded their exploratory behaviour in detail over a 25-minute period. They spent most of the time exploring the corners and indents of the objects, then areas of high curvature second, over smooth surfaces. We would like to suggest this may be because corners and areas of high curvature are more likely to cue useful properties/affordances about different objects.

Related to this finding, it is interesting to consider in AI how Shaw uses information theory to apply the principle of maximising information and predictive capabilities to an image analysis task, and the first result he finds is an edge detector [12]. Similarly, related to the AC finding on relative importance of areas of high curvature, Ravishankar [26] found that it is easier for an artificial agent to recognise deformed objects by placing emphasis on the bending around points of high curvature. It is further compelling that a piece-wise continuous mathematical function is naturally segmented at its points of discontinuity; corners are discontinuities of the derivative of a function of one dimension; edges are discontinuities of the derivative of a function of two dimensions; while points of high curvature (maxima, minima and inflexion points) are points where the first derivatives are zero. It would seem that the same points that mathematicians deem interesting are playing a major role in both natural and artificial agents, as features for object segmentation, categorisation, tracking and, possibly, prediction. Therefore the use of mathematical curves to approximate deformable objects is highly illustrative.

AI Model I: Modelling the Sponge. In one dimension, a way of approximating a continuous curve is by a succession of lines. In two or more dimensions, shapes can be approximated by **meshes**, where each element is ‘flat’ and defined by nodes and edges. Triangular and hexagonal meshes are widely used. Alternatively, quadrics and polynomials of two or three degrees can be used. They are flexible enough for representing most humanly distinguishable continuous shapes. Polynomials have been used to form **splines**, which are defined by a small set of control points. They can be used to interpolate as much detail of a shape as desired, since the polynomials are continuous, while the connections between them can be discontinuous (e.g. [18]). This is why we considered meshes and splines for our model.

As an experimental example, our model analysed the process of deforming a sponge. In general, compliant materials have the potential to change their shape in nearly an infinite amount of unpredictable ways, therefore understanding deformable objects poses a particularly interesting challenge for both artificial and natural agents. Unlike rigid objects, it is not possible to know in advance all of the elements required for representing the deformation. How many splines or elements in a mesh is required, or what are its degrees of freedom? For some specific objects under controlled circumstances, these possibilities can be restricted, as in medical research with human organs [27]. However, an agent that interacts with an environment populated with unrestricted deformable objects, requires a more general solution. One approach is to automatically generate a hierarchical mesh to represent a few objects in a virtual environment, which adapts as an object deforms [28]. However, this has not yet been directly tried in robotics, where an internal representation needs to match objects in the external environment. This continues to remain an open question even in the AC literature – what would the agent do if an object becomes deformed to a shape unforeseen by the initial representation?

As a tentative first step towards solving this problem, we looked at modelling a spheric robotic finger pushing against a sponge. Please note we are not claiming this model replicates animal vision or reasoning, but it may provide a building block from which to work from. The movement of the robotic finger was blocked by a pencil

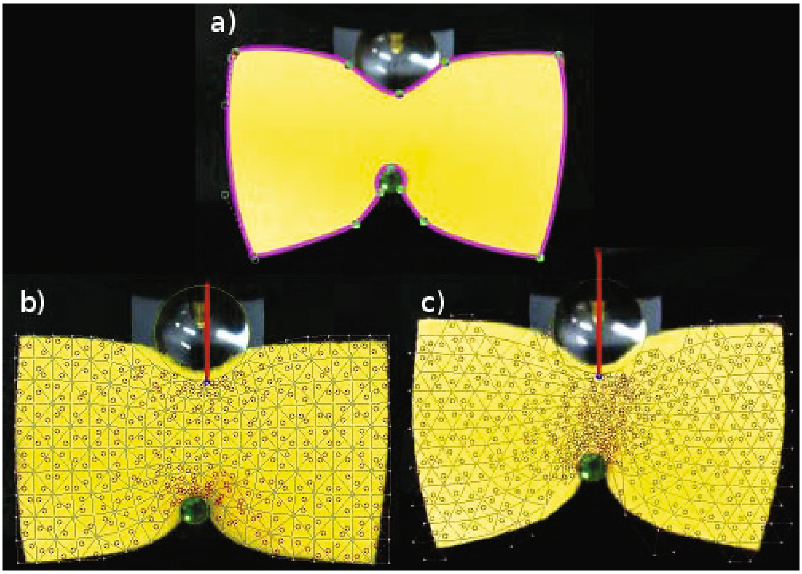

Fig. 1. Top view of an experiment where a robotic finger (sphere at the top) pushes a sponge, perpendicular to its widest axis, against a pencil that serves as an obstacle (green cap). **a)** The contour of a deformed sponge approximated by a series of splines, with the control points placed by a human. **b)** The sponge represented by a rectangular mesh, generated in the first frame before deformation; the mesh configuration was predicted by the physics model. **c)** Hexagonal mesh, similar to (b).

directly opposite. The finger performed a continuous movement against the sponge, while a camera and a force sensor registered the interaction. Figure 1 illustrates the use of splines and meshes to approximate the contour and surface of the sponge as it became deformed.

3.3 Representing the Related Processes

Kakariki Experiment II: More Implications for Models. Once the agent can generate a representation of any object shape it may detect, we believe the next step is for it to understand the related physical processes in the environment. It should identify the key elements and unite them with appropriate functions. How does the object become deformed when interacted with in the world?

We first considered this in the natural dimension in a second AC behavioural experiment. We presented the same kakariki with five cubes of different deformabilities in a random order over five trials over different days. As in the previous experiment, in each trial we allowed them 25 minutes to interact with the objects as they chose and recorded their exploration behaviour in detail.

As we predicted in [10], they initially explored the two extremes the most (i.e. the most rigid and the most deformable cube), but their exploratory 'focus' or 'strategy' changed. So in the second and third trial, the cube of the 'median' or intermediate deformability was explored significantly more than all of the other cubes. Then in the final two trials, the cubes the next interval along (i.e. the second-most deformable cube and the second-most rigid cube) became more of a focus for the kakariki's exploration. In conclusion, the exploration strategy seems to change with time, perhaps as more experience and progressively more specific knowledge is gained about the deformability of objects and different object categories. We would like to suggest that the kakariki had a exploration strategy that allowed them to gain more information about the *process* of deforming an object.

AI Model II: Modelling the Deformation. Simultaneously, we wanted to consider what the design of this internal strategy/learning mechanism could be for an artificial agent. In the AI example of deforming the sponge, the following key frames can be identified:

1. **The finger starts moving.** At this point the force sensor detects only some noise, but the command to move has been given and the vision (camera) begins to detect changes between frames, i.e. that the position of the finger is changing. Thus, the first key frame would contain the finger separated from the sponge and the pencil.
2. **The finger touches the sponge.** At this point the force sensor detects an abrupt increase in one direction. Visually, collision detection routines begin to detect a contact between the circle (i.e. the finger), and one or two triangles in the mesh (i.e. the sponge).
3. **The finger stops moving.** No more changes are detected.

Notice that these coarse key frames are the frames where things change in a very noticeable manner. It is possible to connect frames 1 and 2 by using a function that describes the simple linear translation of the circle (finger). Between frames 2 and 3, the same translation function applies to the circle, but also the physics model gets activated to deform the mesh (sponge) as the circle pushes it. These two functions can predictively describe the observed movements. At frame 3, no function or model is required anymore, because the execution of the command is over and there is no more movement. The scene has ended. From this perspective, the whole process/action can productively be segmented into smaller actions. The internal representation of each frame can be formed by tracing back the activation and deactivation of the required *mechanisms*. Now each segment can be re-represented by a single symbol. The whole sequence can be described as something like: displace finger; push sponge; stop. The agent can then choose between thinking of the command it executed (e.g. translate), or the changes in the sponge (detected through vision or touch), or combinations of both.

There are precedents to doing this type of segmentation, such as in the work by [21]. Here the agent, *Abigail*, analyses a simple circle-and-sticks simulation of ping-pong. Even for this highly simplified world, it was not trivial to unequivocally detect the points of discontinuity that establish the beginning and end of an action. However, Siskind was not quite using our concept of segmentation in modelling, which is just an extension of the idea of a polynomial connecting two control points. Even though the use of splines to approximate curves is a widely used technique, there is not a general technique that can automatically generate a spline from scratch to approximate any curve. It is a brand new research field; to investigate the use of models for interpolation between frames, segmenting and understanding actions.

4 Conclusion

By studying both artificial and natural agents, we can provide a fuller account of how, when necessary, an individual can efficiently represent objects and their related processes in the environment from the huge number of sensory signals they receive. In this light, we can also consider what the requirements posed by the external environment may be upon the finite brain of the agent. Thus, we have briefly discussed two behavioural experiments on parrot exploration of novel objects to give us an insight into what the biological constraints might be on an AI model for representing deformable objects.

Specifically, we have described how a selection of key elements from the environment could be used as a basis for an object representation, and considered possible underlying exploration strategies for gathering information by observing natural behaviour. These key elements are connected through functions, which indicate how to obtain the value of other points. The same mechanism could be used to represent processes and actions, by identifying key frames, and finding the correct physics model to interpolate between frames. It is possible to segment a complex interaction between the agent and the environment into individual actions, by detecting: the commands given; discontinuities in the sensory signals; and the intervals of application of each mechanism. Each of these individual actions could then be represented by symbols. These symbols are grounded in the environment through the selected key elements. It is straightforward to use these symbols for traditional problem-solving tasks, as in [14]. We have further provided evidence that natural agents seem to similarly focus their exploration behaviour on key environmental elements, such as corners, edges and areas of high curvature. Likewise, at least with parrots, individuals seem to attend first to extreme exemplars of particular object properties, including deformability/rigidity, but this exploration strategy becomes gradually refined with time. However, we cannot yet confirm if this parrot exploration is due to similar underlying mechanisms as those presented in our AI model. In conclusion, we have presented an interesting *preliminary* analysis of some of the forms of object representation that may be useful to intelligent natural agents in certain contexts, and demonstrated these capabilities in working computer models.

Acknowledgements. We would like to thank our colleagues in the Intelligent Robotics Lab and the Cognitive Adaptations Research Group for their help throughout this project. This work was made possible by our funders, the CONACYT, Mexico, the ORSAS Scheme and the BBSRC, UK.

References

1. Chappell, J., Sloman, A.: Natural and artificial meta-configured altricial information-processing systems. International Journal of Unconventional Computing 3(3), 211–239 (2007)

2. Calvo, P., Gomila, T. (eds.): Handbook of cognitive science: An embodied approach. Elsevier, London (2008)
3. White, R.W.: Motivation reconsidered: The concept of competence. Psychological Review 66(5), 297–333 (1959)
4. Schulz, L., Sommerville, J.: God does not play dice: Causal determinism and preschoolers' causal inferences. Child Development 77(2), 427–442 (2006)
5. Thomas, R.: Investigating cognitive abilities in animals: unrealized potential. Cognitive Brain Research 3(3-4), 157–166 (1996)
6. Visalberghi, E., Tomasello, M.: Primate causal understanding in the physical and psychological domains. Behavioural Processes 42, 189–203 (1997)
7. McCrone, J.: Smarter than the average bug. New Scientist 190(2553), 37–39 (2006)
8. Demery, Z.P., Chappell, J., Martin, G.R.: Vision, touch and object manipulation in Senegal parrots *Poicephalus senegalus*. Proceedings of the Royal Society B 278(1725), 3687–3693 (2011)
9. Markram, H.: The blue brain project. Nature Reviews Neuroscience 7(2), 153–160 (2006)
10. Demery, Z., Arriola-Rios, V.E., Sloman, A., Wyatt, J., Chappell, J.: Construct to understand: learning through exploration. In: Proceedings of the International Symposium on AI-Inspired Biology, pp. 59–61 (2010)
11. Mesulam, M.: From sensation to cognition. Brain 121, 1013–1052 (1998)
12. Shaw, J.M.: Unifiying Perception and Curiosity. PhD thesis, University of Rochester, Department of Computer Science, The College Arts and Sciences. Rochester, New York (2006); Supervised by Professor Dana H. Ballard
13. Chappell, J., Demery, Z.P., Arriola-Rios, V., Sloman, A.: How to build an information gathering and processing system: Lessons from naturally and artificially intelligent systems. Behavioural Processes 89(2), 179–186 (2012)
14. Arriola, V.E., Savage, J.: Knowledge Acquisition and Automatic Generation of Rules for the Inference Machine CLIPS. In: Gelbukh, A., Kuri Morales, Á.F. (eds.) MICAI 2007. LNCS (LNAI), vol. 4827, pp. 725–735. Springer, Heidelberg (2007)
15. Karmiloff-Smith, A.: Beyond Modularity: a developmental perspective on cognitive science. MIT Press, Cambridge (1999)
16. von Bayern, A.M., Heathcote, R.J., Rutz, C., Kacelnik, A.: The role of experience in problem solving and innovative tool use in crows. Current Biology 19(22), 1965–1968 (2009)
17. Minsky, M.: A framework for representing knowledge. In: Winston, P. (ed.) The Psychology of Computer Vision, p. 81. McGraw-Hill, NY (1975)
18. Montagnat, J., Delingette, H., Ayache, N.: A review of deformable surfaces: topology, geometry and deformation. Image and Vision Computing 19(14), 1023–1040 (2001)
19. Barber, B.: How to Draw Everything. Arcturus (2009)
20. Hoffmann, C.M., Rossignac, J.R.: A road map to solid modeling. IEEE Transactions on Visualization and Computer Graphics 2, 3–10 (1996)
21. Siskind, J.M.: Grounding language in perception. Artificial Intelligence Review 8, 371–391 (1994)
22. Parent, R. (ed.): Computer Animation. Morgan Kaufmann (2009)
23. Menache, A.: Understanding Motion Capture for Computer Animation. Elsevier, London (2011)
24. Teschner, M., Heidelberg, B., Muller, M., Gross, M.: A versatile and robust model for geometrically complex deformable solids. In: Proceedings of Computer Graphics International (CGI 2004), pp. 312–319 (2004)

25. Steels, L.: The symbol grounding problem has been solved. so what's next? In: de Vega, M. (ed.) Symbols and Embodiment: Debates on Meaning and Cognition, Oxford University Press, Oxford (2008)
26. Ravishankar, S., Jain, A., Mittal, A.: Multi-stage contour based detection of deformable objects. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 483–496. Springer, Heidelberg (2008)
27. McInerney, T., Terzopoulos, D.: Deformable models in medical image analysis: a survey. Medical Image Analysis 1(2), 91–108 (1996)
28. Morris, D., Salisbury, K.: Automatic preparation, calibration, and simulation of deformable objects. Computer Methods In Biomechanics And Biomedical Engineering 11(3), 263–279 (2008)