

# Learning to React to Abstractions: Accumulating Adaptations in a Humanoid Embodiment

Christopher Larcombe, Anthony Morse, Angelo Cangelosi

School of Computing and Mathematics  
University of Plymouth, Drake Circus  
Devon, PL4 8AA, UK

E-mail: christopher.larcombe@plymouth.ac.uk

**Abstract**—Human beings and several other living organisms are capable of acquiring a diverse repertoire of adaptive behaviours or skills, through interaction with an appropriate environment. Based on observations of human embodiment and existing cybernetic theory, an operational description of this form of ‘scalable’ adaptive behaviour is derived. An articulated mechanism using the principles identified is implemented and used to control the humanoid robot *iCub*. The experimental physical embodiment is tested with a number of environments. Preliminary results demonstrate a limited form of emergent behavioural growth and corresponding ‘task’-non-specificity: the *iCub* is able to cumulatively learn *multiple* ‘tasks’, by chaining together sequences of primitive ‘reactions’, to ‘abstractions’.

## I. INTRODUCTION

The past decade of research in the interdisciplinary fields of epigenetic<sup>1</sup> robotics [1], [2] and autonomous mental development [3] indicate a growing interest in what has been termed ‘open-ended learning’ [3]–[5] and ‘ongoing development’ [6]–[9]: a broad topic, exploring the science of living organisms and engineering of life-like artificial systems, capable of continuously learning behaviours or skills in a cumulative, open-ended manner. Such systems are often also referred to as task-independent or task-non-specific (learning) systems [3], [5], [8].

Recent robotics work in this area has focused on the involvement of motivation, in particular *intrinsic motivation*, where the development of a learning agent is driven or regulated by curiosity [8], [10], novelty [11], ability to anticipate [12], or other similar measure intrinsic to an exploratory learning process (see [5], [8] for a review).

There appears to be a general consensus forming that an intrinsic motivation mechanism, or self-regulating developmental program, is necessary for ‘open-ended learning’, or at least fundamental to it. In a review of curiosity and motivational systems in *artificial agents*, Stojanov and Kulakov recognise that curiosity can drive a *task-non-specific* learning system “to do something rather than nothing” [5, p.1]; assuming that in the absence of any goal or task-specific structures, coherent behaviour cannot otherwise occur. The authors also cite Dennet, who claims that “curiosity –epistemic hunger– must drive any powerful learning system” [13].

While these three related assumptions are not refuted in this paper, using an alternative approach and methodology we aim to demonstrate both *ongoing* behavioural growth and a limited *task-non-specificity* in a situated system, without the use of *intrinsic* motivation. This may provide an additional perspective, complementary to the existing research in this area.

We also propose a distinction between two forms of open-ended learning in *task-non-specific* systems: **1**) where development of behaviours or skills progress through a *series of stages* or multiple levels, increasing in complexity or sophistication as described in [3], [8], [10], [12]; and **2**) where multiple behaviours or skills are accumulated indefinitely *without* necessarily proceeding through such progressive stages. The latter describes this work, where the emphasis is on task-non-specific behavioural growth (indefinite accumulation of behaviours), rather than incremental complexity. It may be regarded as open-ended in a different sense, or perhaps to a lesser degree.

Acknowledging the vision of open-ended learning in recent work, Prince et. al. [6, pp.1] define the term *ongoing emergence* as a description of “behavioural growth”, referring to “the continuous development and integration of new skills”, exhibited when a number of criteria are met including ‘incorporation of new skills with existing skills’, ‘autonomous development of values and goals’, ‘stability of skills’ and ‘reproducibility’. While the research presented in this paper was not intended to meet these criteria, since several are met it may be considered closer to a demonstration of ongoing emergence than open-ended learning, in the first sense (stage-progressive) defined above.

### A. Overview

Building on the cybernetics work of W. Ross Ashby (see Section I-B) and second-order cybernetics of H. Von Foerster [14], this approach emphasises the dynamics and physical characteristics of whole embodiments, taking into account both the embodiment of the observer and our shared terrestrial environment. In particular, the principle of “accumulating with adaptation” is used, which resembles the concept of cumulative learning. Based on this principle and several tentative observations of human embodiment, a set of constraints are identified and used to derive several definitions and the concept

<sup>1</sup>Also called developmental robotics and ontogenic robotics.

of ‘accumulating *reactions-to-abstractions*’ in an articulated system (see Section II). The system is capable of learning to flexibly re-use a set of behaving parts, called ‘interacting-parts’, in multiple perceptual contexts (states of abstraction).

This general model is then instantiated forming an embodiment-specific but ‘task’-non-specific dynamical system (see  $\Theta^D$ , Section III-B), which is subsequently implemented in software and used to control the humanoid robot *iCub* [15]. Over a series of experiments it is shown that the *physical* embodiment synthesised is capable of demonstrating a limited form of environmentally-determined (epigenetic) ‘behavioural growth’ (see Section IV).

### B. The Cybernetics of W. Ross Ashby

In the 1950s, a British psychiatrist and pioneer of cybernetics W. Ross Ashby, published a sophisticated yet elegant framework [16] accounting for adaptive behaviour, and the adaptive nature of living organisms. Simple mechanisms were presented in order to communicate what Ashby considered to be the mechanistic essence and *physical* principles involved; the adapted behaviour of an organism was equated with the concept of a system in equilibrium with an environment [17], reacting appropriately to disturbances, in order to regulate ‘essential variable(s)’ associated with survival.

One core concept was that of *ultrastability*, demonstrated by the *ultrastable system* (see Figure 1, left). In the ultrastable system, stable behaviour  $R$  in an environment changes at random, if and only if occasional random changes occur in a *step-function*  $S$ , acting as a parameter to the behaviour. The step-function is stable if and only if the *essential variable* is homeostatically regulated. When the essential variable moves beyond its ‘given limits’, random periodic changes occur to  $S$ . Therefore, a behavioural interaction only endures if the essential variable is appropriately regulated, i.e., the system as a whole is both stable and adaptive.

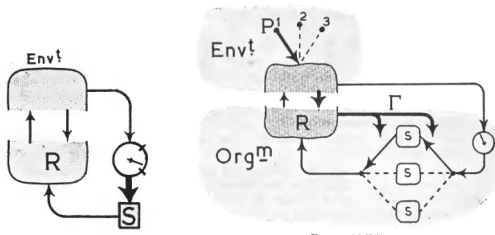


Fig. 1. The Ultrastable System (left) and Accumulator of Adaptations (right). Diagrams taken from ‘Design for a Brain’ [16]. The arrows indicate ‘immediate effect’.

The accumulator of adaptations (see Figure 1) was used by Ashby as a didactic device in order to demonstrate the benefit of “adapting with accumulation”, when an adaptive system faces a changing environment that repeats itself, requiring corresponding changes in behaviour. The parameter  $P$  occasionally changes between three states, requiring adaptation to three environments (three different behaviours). A gating mechanism  $\Gamma$  ensures that a different step-mechanism is ‘active’ in each environment, such that adaptations are retained

on returning to a ‘previous’ environment. The mechanism was presented in contrast to the ultrastable system (see Figure 1, left), which would have to re-adapt to every change in such an environment.

## II. THEORY

This section introduces the primary contribution of this paper: the concept of **1)** accumulating *reactions-to-abstractions* (see Figure 2, left); and **2)** doing so in an articulated system of independently controlled (de-coupled) parts, each accumulating reactions-to-abstractions, sharing a common ‘context of abstraction’ (Figure 2, right), i.e., the same set of abstractions. Each component of the diagram is justified by a particular embodiment-specific constraint, which will be stated in the sub-sections following Section II-A. The latter concept is represented in the form of a general model  $\Theta$ , the *articulated abstraction-reaction accumulator* (AARA).

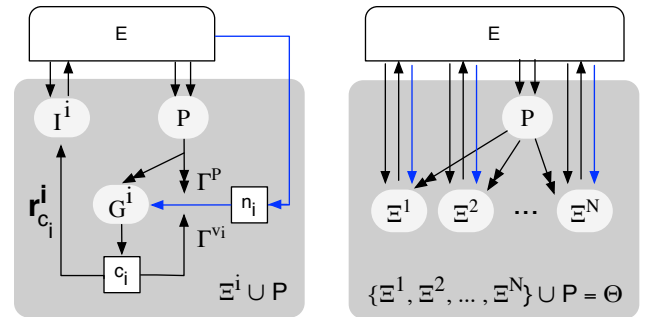


Fig. 2. The (Articulated) Abstraction-Reaction Accumulator,  $\Theta$  – Each unit  $\Xi^i$  accumulates adaptations independently for an interacting-part  $I^i \in \Xi^i$  (left).  $N$  adaptive abstraction-reaction (sensorimotor) loops, modulated by a (shared) ‘perceptual environment’ or *set of abstractions*  $P$  (right).

The theory includes the following working definitions at the time of writing:

- *Behaviour*: overt (sequence of) motion(s) in an observable environment (abstracted by an observer)
- *Reaction*: a recurring covert state, *modulating* behaviour.
- *Abstraction*: a recurring covert state, co-occurring with a (variable) *set* of environmental states.
- *Reaction-to-an-abstraction*: a covert state, correlated with the hypothetical future co-occurrence or non-co-occurrence of a particular *reaction* and *abstraction*.

### A. Operational explanation

Observable *behaviour* is split into  $N$  independently observable ‘interacting-parts’  $I^i$  (where  $i \in [1..N]$ ), that each exhibit a total of  $|r^i|$  recurrent (repeating, recurring) behavioural interactions with  $E$ . The observable behaviour of each part is *modulated*, rather than fully determined, by a *reaction-variable*  $c_i \in r^i$ , which can be in one of  $|r^i|$  different states. In other words, the variable  $c_i$  specifies or indicates the occurring (current) *interaction* for its corresponding interacting-part.

In any given moment, and environment, a dynamic but limited set of interactions are possible for each interacting-part. The environment provides normative feedback indicating the

‘appropriateness’ of the current interaction, via perturbation to  $n_i$ , a *normativity variable*. Adaptation of  $G^i \in \Xi^i$  and thus  $\Xi^i$  occurs if and only if  $n_i \neq 0$ , gated by the ‘state of abstraction’  $P$  (what is ‘perceived’) and reaction  $c_i$  taking place.

The double-arrows to and from  $P$  indicate multiple independent channels of communication (effect), expressing the fact that there are *many* de-coupled abstractions. If  $G^i$  contains a set of states greater than or equal to the number of abstractions  $|P|$  multiplied by the total reaction-states of that part  $|r^i|$ , there is necessary and sufficient variety in the system to form a unique association between *every* distinct reaction and abstraction. If this is the case, and assuming  $G^i$  is appropriately structured, the double-gating  $\Gamma^P, \Gamma^{c_i}$  of  $P, c_i$  on  $n_i$  effecting  $G^i$ , can be directed to a unique sub-set of states in  $G^i$ . Such states can act as adaptive reactions-to-an-abstractions, by regulating the (future) effect of  $P$  on  $c_i$  accordingly, using the additional channel of communication from  $P$  to  $G^i$ .

Making the assumption that *the future will be like the present*<sup>2</sup> [18], when a normativity perturbation to unit  $i$  occurs,  $G^i$  can be modified so that the reaction occurring at the time is more or less likely to occur again in the future, in that particular *state of abstraction*, i.e., the perceptual, morphological or environmental context. Hopefully, in a manner appropriate to avoiding *negative* disruptions to  $n_i$  in future situations.

#### Constraint I: Articulated Morphology

Every human morphology is typically considered to be an articulated morphology by definition, featuring a finite number of ‘moveable joints’. However, the human nervous system, by observing its morphological environment, is also able to abstract ‘parts’ making use of multiple ‘joints’, such as ‘arms’ and ‘legs’. Closer observation reveals that each part actually consists of many other parts nested in a holarchical fashion, such as ‘hands’ containing ‘fingers’, and ultimately ‘cells’, etc, forming a dynamical hierarchy [19].

Each dynamic part abstracted can be represented by a *degree of freedom* or dimension of variety in a *description* of that morphology at a particular level of description, e.g., a set or range of *joint* angles, muscle tensions, *limb* postures or positions. We can call this a ‘variable’; for a part  $i$ , at a single level of description, its (current) state may be represented by the *variable*  $c_i$ . The range or set of states  $c_i$  can take may be represented by the vector (state-space)  $r^i$ , i.e,  $c_i \in r^i$ .

Since each part in actuality is not a ‘fixed’ structure, but a dynamic process in itself, a fixed state or posture may be better described as a *stable interaction*, resistant to perturbation. Therefore, for a part  $i$ , a large set of interacting variables  $I^i$  may describe the ‘part’, *modulated* by  $c_i$  (Figure 2:  $c_i \rightarrow I^i$ ). The current ‘state’  $c_i$  therefore indicates the ‘current interaction’ in  $I^i$ . In addition, since each part is both a sub-system of our shared terrestrial environment and a specific morphology,  $I^i$  is coupled to an environment  $E$  (Figure 2:

$E \leftrightarrow I^i$ ). The set of variables  $E$  describe the morphology and its terrestrial environment, including the observer(s).

#### Constraint II: Recurrent Interactions

While implicit having introduced *Constraint I*, for clarity it should be made explicit that interactions repeat themselves, or recur over time. For the part  $i$ , the ‘current interaction’  $c_i$  will return to previous states (those of  $r^i$ ), and thus  $I^i$  will demonstrate recurrent behaviour. To the observer, this will appear as familiar movements, postures or joint-states.

#### Constraint III: Concurrent Interactions

We can observe that when a human body restructures itself (changes) over time due to thermodynamic laws or otherwise, the sequence of states a single joint, limb or ‘part’ may take is not arbitrary, nor is it random. Beyond infancy, the limbs assume different states in a noticeably coordinated manner producing ‘appropriate behaviour’. For example: walking, sitting, pointing, waving, etc., all demonstrate multiple *constraints* in changing states of particular joints or limbs, in relation to each other: *co-occurring* sequences of ‘part’-states that *recur*.

There is also a clear association between particular parts and particular ‘behaviours’. For example, waving and pointing involve the arm and hand, while walking and sitting involve the lower limbs. It is possible to wave while walking or sitting, because the set of parts relevant to waving (hand, arm) is disjunct from the set of parts relevant to both walking and sitting (lower limbs). Similarly, it is possible to point while walking or sitting. However, it is *not* possible to wave while pointing, or sit while walking. In the latter cases the corresponding sets overlap *and* require *different* part states. This recurrent combinatorial nature of behaviour is most apparent in humans, but can be observed in other animals.

In order to reflect this constraint in AARA, a single interacting-part  $I^i \in \Theta$  will typically contain the necessary and sufficient variables to describe *a set* of interacting sub-‘parts’ that taken together frequently form *a set* of recognisable interactions, e.g., multi-joint ‘postures’ or movements, in the manner previously exemplified. For example, an ‘arm’ part  $a$  in interaction-unit  $\Xi^a = \{I^a, r^a = \{1, 2\}, c_a, \dots\}$  may have two reaction-states: ‘held out’ ( $c_a = 1$ ), where the variables of  $I^a$  move to a stable equilibrium where the arm is extended; and ‘waving’ ( $c_a = 2$ ), where  $I^a$  enters a cyclic motion<sup>3</sup>. This would be de-coupled from another interacting-part, say  $I^b$ , associated with the hand, in another interaction-unit.

#### Constraint IV: Reactions in a Perceptual Context

The observation can also be made that such recurring interactions tend to occur more frequently in or with certain ‘environmental situations’, within or beyond the morphology. In living organisms environments or environmental features are said to ‘cause’ or provoke certain reactions, and those reactions have an affect on the environment, bringing it to a new state

<sup>2</sup>There appears to be no reason why an abstraction cannot also be an ‘anticipation’, i.e., associated with an *anticipated* set of (future) environmental states.

<sup>3</sup>Alternatively, there may be three reaction-states, and waving could be produced by oscillating between two of those three reaction-states: e.g.,  $c_a = 2$  and  $c_a = 3$ .

where further reactions will be provoked. However, it has been shown that what is sensed, perceived, or *abstracted* and subsequently ‘reacted to’ in human beings and other complex organisms is certainly distinct from ‘the environment’ [14], i.e., “the map is not the territory” [20]. Rather, sensorimotor correlations are modulated by a ‘sensory surface’ [21, p.161], and behaviour occurs within a *sensorimotor loop*.

In AARA this is represented by a channel of communication from  $E$  (see *Constraint I*) to each interacting-part  $I^i$ , which each affect  $E$  through an independent channel of communication. The effect of  $E$  (‘the territory’) on the interacting-parts is *not* immediate (see Figure 2, right); rather, the effect is via a shared set of *abstractions*, a ‘perceptual environment’ of coupled variables  $P$  (‘the map’). More specifically, for an interaction-unit  $\Xi^i = \{c_i, I^i, G^i \dots\}$ , an arrow from  $P$  to  $c_i$ , via a sub-system (set) of variables  $G^i$  (see Figure 2, left), represents the notion that *what is abstracted* will affect the reaction-state  $c_i$ . In other words, the state of  $G^i$  and what is ‘perceived’ by the system determine the reaction, which modulates the observed interaction. Abstractions will therefore co-occur with reactions, depending on  $G^i$ .

#### Constraint V: Adaptive Reactions in a Normative Context

This constraint is more speculative, based on the observation that certain recurring interactions of certain parts (*Constraint II*) are ‘appropriate’ and ‘inappropriate’ in multiple contexts (environments), and thus in different ‘states of abstraction’ – and further, that the co-occurrence and non-co-occurrence of abstractions and reactions (and thus observed behavioural interactions) are not fixed, but adaptive, depending on this normative context, i.e., whether or not an interaction is appropriate.

The first difficulty arises in grounding the terms appropriate and inappropriate. This will be left open. However, we do recognise that living organisms, including humans, regulate ‘essential variables’ correlated with survival [16], [17] and seek to maintain their autopoiesis [21], or self-creating identity. For simplicity we will assume that each part or unit  $\Xi^i$  has an independent ‘adaptation signal’ or *normativity* variable  $n_i$ , based on the appropriateness<sup>4</sup> of the interaction  $I^i \in \Xi^i$ . While this may *not* be a reasonable assumption, in terms of the dynamics,  $n_i$  is primarily a state correlated *with adaptation*, or change in  $G^i$ . For example, in a human being,  $n_i$  for a ‘hand’ or ‘arm’ part may be a set of states correlated with several essential variables monitoring skin temperature, around the hand. The perception of ‘fire’, a complex set of abstractions, may occur with an extended arm and open hand.

The sub-system  $G^i$  *determines* the reaction-state, and is modulated by  $P$ . This sub-system contains variables that change when learning (adaptation) takes place, that alter the relationship between abstraction (perception) and (re)action.

<sup>4</sup>In relationship to the ultrastable system,  $n_i$  (indicating *innappropriateness* over neutrality) would change state only fleetingly, when a step-change occurs to the step-function. However, there is no analogy when  $n_i$  indicates *appropriateness* over neutrality, as this possibility was not taken into account by Ashby.

When  $G^i$  is affected by  $n_i$ , adaptation takes place. The arrows labelled  $\Gamma^P$  and  $\Gamma^{c_i}$  indicate that a gating effect of  $n_i$  on  $G^i$  takes place, modulating the effect of the normativity variable on  $G^i$ . In other words, the effect of the feedback from the environment on  $G^i$  depends on the current reaction (behavioural interaction) and current state of abstraction (perception).

### III. DEMONSTRATION METHOD

The *humanoid* robot iCub [15], superficially resembling the human form, can also be described as an articulated system; the morphology is therefore suitable for demonstrating the concept of accumulating reactions-to-abstractions and an example implementation of AARA (see Figure 2). However, since the iCub is not an autopoietic system and is thus lacking in essential variables, for convenience *only* the normative context is provided by a human tutor. Normativity perturbations are applied *manually*, analogous to changes in the step-function of an ultrastable system (see Figure 1, Section I-B) when homeostasis of an essential variable is challenged. The concept of ‘positive’ normativity is also introduced, based on the observation that adaptation also often coincides with restabilisation of essential variables<sup>5</sup>.

#### A. Environment

In the experimental setup the ‘environment’ to the AARA-system includes both the morphology (iCub robot and coupled computer systems), a human tutor, a table, sloping ramp and two coloured rubber balls: a blue and a green ball. An experimental implementation  $\Theta^D$  of the general AARA  $\Theta$  was instantiated in software on the computer system, within the software development environment *Aquila* [22].

In the experimental environment (see Figure 3), the iCub robot is supported in a fixed position behind a table, upon which a sloping ramp is placed, also in a fixed position. When a ball is placed on the ramp or released above it, under the effect of gravity the ball rolls towards the end of the ramp, to the left of the iCub. The ramp is capable of holding two balls simultaneously.

During the experiment the tutor ensures that the coloured balls when stationary remain within a set of positions at different times, depending on the phase of the experiment (the ‘environment’). For convenience, the positions have been labelled (see Figure 3): POS1, on the table in front of the iCub; POS2, above the table in front of the iCub (above the ramp); POS3, on the ramp, at the end of the ramp; POS4, on the ramp, behind POS3; POS5, above the table to the right of the iCub; POS6, on the table to the right of the iCub.

The iCub is able to open and close its right hand, and move its right arm between 4 different positions, corresponding to POS1, POS2, POS5 and POS6. The arrangement of the environment is such that a ball can be grasped in each of these positions, if the arm and hand are coordinated appropriately. A ball can only enter POS4 if the other ball is at the end of

<sup>5</sup>The superficial resemblance to operant conditioning also provides a familiar narrative, making interpretation of the results easier. However, we hope this will not be a distraction from what is being demonstrated.

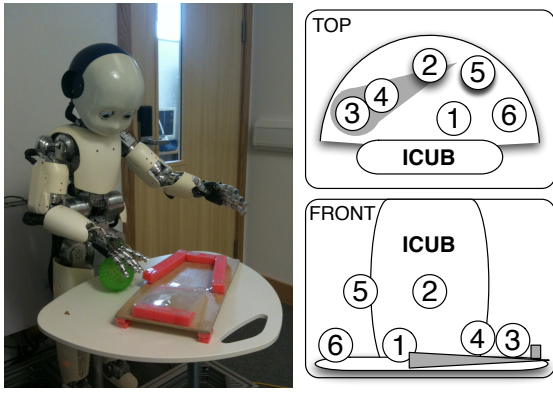


Fig. 3. The experiment setup, a green ball in POS1 (left). Showing POS1-POS6 from above, and facing the iCub (right).

TABLE I  
ENVIRONMENT-TASK MAPPING

E1	<i>iCub</i> move green ball from POS1 to POS3 <i>tutor</i> move green ball from POS3 to POS1 ( <i>tutor</i> remove blue ball)
E2	<i>iCub</i> take blue ball from POS5 to POS4 <i>tutor</i> give iCub blue ball at POS5 <i>tutor</i> remove blue ball from POS4 ( <i>tutor</i> keep green ball in POS3)
E3	<i>tutor</i> remove green ball from POS3 <i>iCub</i> open hand, move hand (arm) over POS1 <i>tutor</i> move green ball to POS1 ( <i>tutor</i> keep blue ball in POS3)

the ramp, at POS3. A ball released at POS2 will move down the ramp, towards POS3.

*Multiple Environments:* The behaviour of the tutor cycles between three different modes, pertaining to three recurring ‘environments’ to the AARA-system  $\Theta^D$ : E1, E2 and E3. Within each environment certain states will recur requiring particular responses, or ‘appropriate’ behaviour by  $\Theta^D$ . The goal of the tutor is to teach the iCub three ‘tasks’ through a process of incremental learning, where one task is taught in each environment (see Table I). When a behaviour is not appropriate to the tutor, a NP is administered. While one ‘task’ appears to be taught in each environment (see Table I), each ‘task’ and environment can be broken down into sub-‘tasks’ and sub-‘environments’, defined at the level of abstraction and reaction. The break-down into three environments is primarily for convenience, and somewhat arbitrary;  $\Theta^D$  is learning sequences of appropriate reactions to abstractions, which are observed as sequences of appropriate behaviours in coordination with a changing environment.

An experimental run consists of the following sequence of environments, necessary and sufficient to demonstrate accumulation of adaptation: E1  $\rightarrow$  E2  $\rightarrow$  E1  $\rightarrow$  E2  $\rightarrow$  E3  $\rightarrow$  E1  $\rightarrow$  E2  $\rightarrow$  E3 ( $\rightarrow$  E4). The tutor is free to pass from one environment to the next when the ‘task’ corresponding to that environment has been completed 4 times, without any normativity perturbations (when the behaviour is adapted). In E4, the tutor can cycle between E1, E2, and E3 freely.

## B. AARA Configuration

The demonstration AARA  $\Theta^D$  is configured to suit the experimental embodiment: the iCub morphology and its environment (see Section III-A). Two interaction-units  $\Xi^A \in \Theta^D$  and  $\Xi^H \in \Theta^D$  are used to control the *right arm* and the *right hand* of the iCub respectively, sharing a common ‘perceptual system’  $P \in \Theta^D$ .

*Perceptual System:* The perceptual system  $P$  contains a vector  $\mathbf{p}$  of 390 ‘abstraction elements’, where each element (vector element) has a value between 0 and 1. Of the 390 abstractions, 6 are proprioceptive, based on joint-encoder values, and the remaining 384 are visual, abstracting the presence of two different hues (192 for blue and 192 for green) in areas of the RGB image received from the right-eye camera of the iCub. The image is partitioned into 192 segments, where the value of each segment (abstraction element) is equal to the percentage of pixels in that segment of the target hue.

*Motor System:* Since the morphology has two ‘parts’ which may be independently controlled, two interacting-parts are used:  $I^A$  and  $I^H$ . The right arm of 7 joints, which can move between 4 different positions, is modulated at time  $t$  by the reaction-state variable  $c_A(t) \in \mathbf{r}^A$ , i.e.,  $c_A \in \{1, 2, 3, 4\}$ , which indicates the current (active) reaction-state. The selection of each state moves the arm to a different position. The set of variables  $I^A \in \Xi^A$  therefore relate to the motion of the arm, which is directly observable and not fully determined by  $c_A$  or  $\Theta^D$ . The state of the right hand (8 joints), described by  $I^H$  is similarly modulated by the reaction-state  $c_H \in \{1, 2\}$ , which can take one of two states, corresponding to an open and closed hand respectively.

*Gating Mechanism:* Each interaction-unit  $i$  in  $\Theta^D$  must also have a gating mechanism  $G^i$ , coupling the normativity variable  $n^i$  and abstractions  $\mathbf{p}$  to the reaction-state variable  $c_i$  (see Figure 4). This mechanism is responsible for the appropriate selection of the state of  $c_i$ , such that the normativity variable  $n_i$  is regulated at a value of zero, i.e., to minimise future disruption to the normativity variable.

For each interaction-unit  $i$ , a vector of ‘activations’  $\mathbf{a}^i$  is used to represent the relative appropriateness of each potential state of  $c_i$ . Therefore for  $\Theta^D$ ,  $|\mathbf{a}^A| = |\mathbf{r}^A| = 4$  and  $|\mathbf{a}^H| = |\mathbf{r}^H| = 2$ . The greater the activation value, the more likely the corresponding reaction-state is to be selected over another competing reaction-state within a unit  $i$ :

$$c_i(t) = \begin{cases} \mathbf{r}^i_{m_i(t)} & \text{if } l_i(t) \geq 1 \vee t = 0 \\ c_i(t-1) & \text{if } l_i(t) < 1 \wedge t > 0 \end{cases} \quad (1)$$

where  $m_i(t)$  is the index of the reaction-state with the maximum activation, i.e.,  $m_i = \arg \max_x (a_x^i)$ , and  $l_i$  is the *lability* of unit  $i$ . The lability  $l_i$  of each unit  $i$  acts as a timing mechanism, enforcing ‘trials’ of different reaction-states of sufficient duration to receive feedback (normativity perturbations). When a lability variable increments from 0 to a value of 1, a change in reaction-state can occur (see Equation 1) and the lability subsequently resets to a value of 0. The lability variable  $l_i$  of unit  $i$  will increase if and only if there

is an activation variable greater than that of the activation of the current reaction-state:

$$l_i(t) = l_i(t-1) + \mu \left( a_{m_i(t)}^i(t) - a_{z(t)}^i(t) \right) - \lambda \quad (2)$$

where  $t > 0$ , constant  $\mu = 0.01$  and linear decay constant  $\lambda = 0.01$ . When a liability variable reaches zero no further decay occurs.

Finally, activations at time  $t$  are calculated based on the state of abstraction and a learned weight matrix  $W^i(t)$  as follows:

$$a_j^i(t) = \sum_{k=1}^{|\mathbf{p}|} p_k(t) w_{k,j}^i(t) \quad (3)$$

where  $w_{k,j}^i(t)$  is a real-valued weight relating the shared ‘abstraction element’  $k$  (perceptual element  $p_k$ ) to the potential reaction-state  $1 \leq j \leq |\mathbf{r}^i|$  of unit  $i$ . In  $\Theta^D$ , there are therefore two weight matrices: a  $|\mathbf{p}|$ -by- $|\mathbf{r}^A|$  matrix  $W^A$  and a  $|\mathbf{p}|$ -by- $|\mathbf{r}^H|$  matrix  $W^H$ . Weights are updated at time  $t$  depending on the occurrence of a normativity perturbation as follows<sup>6</sup>:

$$w_{k,z(t)}^i(t) = \begin{cases} w_{k,z(t)}^i(t-1) + n_i(t) & \text{if } p_k(t) \geq \alpha \\ w_{k,z(t)}^i(t-1) & \text{if } p_k(t) < \alpha \end{cases} \quad (4)$$

where  $t > 0$  is the time-step,  $\alpha = 0.3$  is an ‘association-threshold’,  $z(t)$  indexes the current reaction-state at time  $t$ , i.e.,  $\mathbf{r}_{z(t)}^i = c_i(t)$ , where  $k \in [1..|\mathbf{p}|]$  and  $n_i(t) \in \{-0.5, 0, 0.5\}$ . The normativity variable  $n_i(t)$  will equal zero unless a perturbation is applied at  $t$ , in which case a weight will be modified corresponding to each active abstraction element (with an activity exceeding  $\alpha$ ) and the current reaction-state. In other words, the effect of  $n_i$  on  $W^i$  is gated by both  $c_i$  and  $\mathbf{p}$  (see Figure 4).

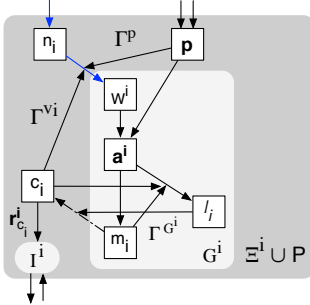


Fig. 4. Structure of interaction-unit  $\Xi^i$  and  $P$  of  $\Theta^D$ , used in the experiment; with implementation-specific gating mechanism  $G^i$ .

#### IV. DEMONSTRATION RESULTS

The experiment described in Section III was repeated a total of 5 times, with activations initialised to random values between  $-0.5$  and  $0.5$ . Of the 5 experimental runs, which showed only minor variations, time-series data captured from one of the runs is graphed in Figure 5. The experimental run shown lasted approximately 8 minutes and 27 seconds (5070 time-steps), demonstrating adaptation of the AARA-system  $\Theta^D$  in multiple ‘environments’ (as in Section III-A).

<sup>6</sup>Alternative, without threshold:  $w_{k,z(t)}^i(t) = w_{k,z(t)}^i(t-1) + n_i(t)p_k(t)$

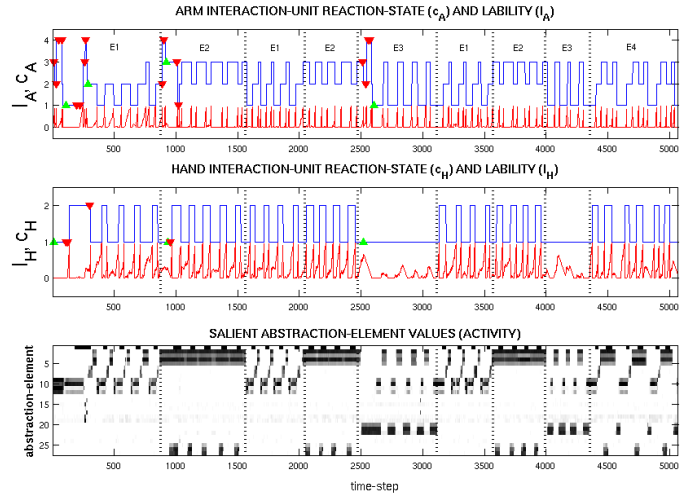


Fig. 5. Top ( $\Xi^A$ ): arm current reaction-state  $c_A$ , arm liability  $l_A$  - Middle ( $\Xi^H$ ): hand current reaction-state  $c_H$ , hand liability  $l_H$  - Bottom (associated elements in  $\mathbf{p} \in P$ ): activity of abstraction elements.

During the experiment a total of 27 normativity perturbations were administered by the tutor (7 positive, 20 negative), over the three environments: E1, E2 and E3. Figure 5 shows the liability and reaction-state during the experiment, for both the arm and hand interaction-units. The graph also shows the activity of 27 (of the 390) abstraction elements that were associated with reactions at some point during the experiment, i.e., active ( $\geq 0.3$ ) when a normativity perturbation (NP) occurred<sup>7</sup>. The unique pattern of abstractions made in each environment is apparent when comparing the activity of the abstraction elements between environments.

The down and up triangles on the graphs indicate NPs, showing that adaptation occurred during the first occurrence of each environment as expected. Upon the second occurrence of each environment, no NPs occurred as no further adaptation was necessary, i.e., behaviour was ‘appropriate’, as adaptations were successfully retained (accumulated) without significant interference. It is also apparent from the results that only one ‘trial’ (repetition) was necessary to teach each ‘task’, i.e., on the second ‘trial’ within each environment behaviour was appropriate. In other runs two trials were necessary in some cases, but adaptation always occurred.

##### A. Adapting to E1

When the experimental run began, the hand was open ( $c_H = 1$ ) and the arm was extended out and to the right of the iCub ( $c_A = 3$ ), positioning the hand just above POS5. The green ball was at POS1, indicated by activation of the four abstraction elements 9 to 12 on the lower graph (Figure 5). Since the hand was already positioned appropriately for grasping the green ball, at time-step 8 the tutor ‘rewarded’ the hand interaction-unit  $\Xi^H$ , i.e.,  $n_H(8) = 0.5$ . This resulted in

<sup>7</sup>The total number of NPs was not equal to the number of associated abstraction elements in other runs; typically the number of perturbations will be greater if reactions-to-abstractions overlap between ‘tasks’ destructively.

the following weights increasing by a value of approximately 0.5:  $w_{124,1}^H$ ,  $w_{125,1}^H$ ,  $w_{136,1}^H$ ,  $w_{137,1}^H$  only, since only four abstraction elements had an activity exceeding the threshold of 0.3 (elements 10, 11, 12, 13 in Figure 5, bottom). This modification increased the activation value  $a_1^H$ , resulting in the stability of the reaction-state  $c_H = 1$  (open-hand), in the abstraction-context ‘green ball in POS1’.

Shortly after ‘rewarding the hand’, the tutor observed the arm was not in an appropriate position for grasping the green ball, and applied a negative NP ( $n_A(13) = -0.5$ ). Similarly, since the state of abstraction had not changed significantly this also altered four weights associated with the same abstraction elements, but gated to reaction-state  $c_A = 3$  of  $\Xi^A$ . Following this perturbation the activation value  $a_3^A$  decreased below the activations of the other reaction-states, resulting in instability; the lability increased to 1 and a bifurcation occurred at time-step 16. At this point the reaction-state changed ( $c_A(16) = 2$ ) and the arm subsequently moved over the ramp (a bifurcation occurred in the interacting-part  $I^A$ ).

The tutor continued to apply NPs to the arm-unit until the appropriate reaction-state was selected, and the open hand moved over the ball at time-step 91, in an position appropriate for grasping. The hand of the iCub partially obscured the ball from vision when this occurred, resulting in a change in abstraction. Three of the four abstraction elements that were previously active then fell below threshold. At time-step 107 a positive NP was applied to the arm. The tutor then recognised the iCub was in an appropriate position to grasp the ball, administering a negative NP to  $\Xi^H$ . Following the perturbation, nothing appeared to change, so another was applied. At time-step 123  $l_H$  began increasing and a bifurcation occurred at time-step 138, at which point the reaction-state changed, and the hand *began* to close. At time-step 181 the hand had finished closing, resulting in a change in the state of abstraction: a proprioceptive abstraction element activated (element 1 in Figure 5) indicating closure of the hand.

Observing the closed hand, the tutor applied a series of negative NPs to  $\Xi^A$  until the arm reached the appropriate position above the ramp ( $c_A = 2$ ), where the ball was held in POS2. The arm-unit was then ‘rewarded’ and the hand was ‘punished’, so it would open, dropping the ball. The negative perturbation to  $\Xi^H$  formed an association between the ‘closed hand’ abstraction-element and three other abstraction-elements associated with ‘green’ in an area of the image associated with POS2. As a result of the changes to the weights and thus activations, the lability of the hand-unit began to increase and the hand opened. The ball was released at POS3 and moved down the ramp towards POS4, once again changing the state of abstraction.

### B. After Adaptation to E1

At time-step 328, when the ball reached POS4 at the end of the ramp, the tutor recognised that the first task had been completed, concluding one ‘trial’ of ‘task-1’ in E1. Since no trials had yet been completed without NPs, the tutor continued teaching the first task, i.e., the environment remained in the

state E1. The green ball was moved back to POS1 at time-step 358 and another ‘trial’ began. When the green ball reached POS1 at time-step 361, the iCub subsequently moved its arm towards the appropriate position ( $c_A(363) = 2$ ,  $c_A(364) = 1$ ). When the hand closed around the ball, due to  $l_H$  reaching a value of 1, the context of abstraction changed and the lability of the arm-unit began increasing. At time-step 424 the reaction-state of the arm-unit changed to  $c_A = 2$ , the appropriate reaction, bringing the ball to POS2 once again. At this point, the hand opened as before and the ball returned to POS3 on the ramp. This was repeated a further three times, before the tutor decided to teach the second task (E2).

In E2, the tutor responds to the green ball in POS1 differently: instead of moving the ball back to POS1, it is left in POS3. After teaching the second task (E2) involving the blue ball (see Table I), and observing 4 trials performed without correction, the environment returns to E1, where the tutor places the green ball back in POS1. At this point the iCub appropriately moves its arm and hand back to the first position above POS1, grasps the ball, moves it to POS2, opens its hand, and the ball falls to POS3. The behaviour is appropriate as the previous adaptations have been retained.

At time-step 2475, a third environment is presented to the AARA-system. Adaptation takes place as before, and behaviour in the other environments is tested following adaptation. The behaviour is once again found to be adapted immediately, and no further normativity perturbations are necessary. At time-step 4354, in E4 the tutor decides to cycle through each environment 3 times, allowing one trial of each task to be completed in each environment. As anticipated, the behaviour of the iCub is appropriate.

## V. DISCUSSION AND CONCLUSION

The concept of accumulating adaptations was applied in a humanoid context, taking inspiration from human embodiment. Several constraints were identified and used to justify the structure of a novel class of mechanism: the articulated accumulator of reactions-to-abstractions (AARA). In this embodiment-specific mechanism abstractions form a shared ‘perceptual environment’ of recurring states that can be reused in time to modulate the behaviour of multiple de-coupled ‘interacting-parts’. Each interacting-part is capable of producing a number of different recurring ‘behaviours’, which co-occur in potential coordination with the recurrent ‘behaviours’ of other interacting-parts, creating a large combinatorial domain of potential interactions. This is possible because each part can be independently adapted and modulated – each part acting as an independent accumulator of adaptations and sensorimotor loop.

An interacting-part is parameterised by a unique set of reaction-states, which can be associated with a particular set of abstractions. For each interacting-part, only one reaction-state of a finite set is ‘active’ or selected at once, co-determining the observed behaviour of that part with the morphological or trans-morphological environment. Due to the shared ‘perceptual environment’ of abstractions, an articulated

morphology implementing AARA can coordinate two or more interacting-parts by associating their respective reaction-states with the same set of abstractions. By associating multiple sets of abstractions with the same reaction-states, one particular reaction(-state) can be re-used in many contexts, as or part of *many* different ‘behaviours’; this provides a growing repertoire of basic behaviours, where interactions are re-used in multiple environments, individually or in coordination with other interactions, in a combinatorial manner.

This capacity was demonstrated to a limited extent in the iCub experiment, allowing the system to accumulate multiple behaviours, despite the fact that the set of abstractions and reactions remained fixed. In a more complex implementation of AARA it seems reasonable that the abstractions and reactions could also be learned, extending the domain of potential interactions, but this was shown to be unnecessary for *basic* ongoing behavioural growth.

The number of behaviours the system is capable of learning is also extended by combining behaviours in time, to synthesise new (meta-)behaviours composed of sequences of behaviours. For example, the experiment results demonstrate in E1 a green ball in a particular position that led to an arm movement positioning the hand above the ball, which led to a change in abstraction, which caused the hand to close, which led to a change in abstraction, which caused the arm to move above the ramp, which led to a further change in abstraction, causing the hand to open, etc. Sequences of behaviour emerge in the mechanism when abstraction leads to reaction, which changes the state of abstraction, which changes the reaction, etc. Complex chains or sequences of multiple reactions to multiple abstractions provide an *additional* domain of potential behaviours, that can be ‘grown’. Since the behaviour or task of ‘moving the green ball from POS1 to POS3’ is an emergent phenomena, this could therefore be termed *synergic* behavioural growth.

The demonstration of AARA in this paper was limited to a single configuration of a single implementation, exposed to a limited set of environments. The *scalability* of the system was therefore not explicitly investigated. While the implementation is by its design ‘task’-non-specific, the particular reaction-morphology couplings used constrain the number and variety of ‘tasks’ the system is *capable* of learning. Of this relatively large sub-set, only the cumulative learning of three specific ‘tasks’ were demonstrated here, corresponding to three dynamic environments. However, the same system with an identical initial configuration to that shown should be capable of learning many other similar tasks (behaviours-in-environments), from the aforementioned sub-set. Furthermore, at any point in its development re-adaptation can occur to a new set of tasks, or further task-specific behaviours can be accumulated.

The observation was also made that as the total learned behaviours and corresponding number of ‘tasks’ increases, the reactions-to-abstractions accumulated tend to overlap both positively (constructively) and negatively (destructively) between tasks and environments. The latter is equivalent to

re-adaptation, where previous adaptations are lost; while the former is equivalent to ‘generalisation’, where novel adapted interactions occur, or adaptation is accelerated due to previous adaptations. This constraint clearly limits the scalability of this approach, and will be a topic of investigation in future work.

#### ACKNOWLEDGMENT

This work was funded by the EU FP7 ITALK project.

#### REFERENCES

- [1] M. Lungarella, “Developmental robotics,” *Scholarpedia*, vol. 2, no. 8, p. 3104, 2007.
- [2] M. Asada, K. Hosoda, Y. Kuniyoshi, H. Ishiguro, T. Inui, Y. Yoshikawa, M. Ogino, and C. Yoshida, “Cognitive developmental robotics: a survey,” *Autonomous Mental Development, IEEE Transactions on*, vol. 1, no. 1, pp. 12–34, 2009.
- [3] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Thelen, “Artificial intelligence. Autonomous mental development by robots and animals.” *Science*, vol. 291, no. 5504, pp. 599–600, Jan. 2001.
- [4] G. Baldassare, M. Mirolli, F. Mannella, D. Caligiore, E. Visalberghi, F. Natale, V. Truppa, G. Sabbatini, E. Guglielmelli, F. Keller *et al.*, “The IM-CLEVeR project: Intrinsically motivated cumulative learning versatile robots,” in *9th International Conference on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, 2009, pp. 189–190.
- [5] G. Stojanov and A. Kulakov, “On curiosity in intelligent robotic systems,” 2006.
- [6] C. Prince, N. Helder, and G. Hollich, “Ongoing emergence: A core concept in epigenetic robotics,” pp. 63–70, 2005.
- [7] D. Blank, D. Kumar, L. Meeden, and J. Marshall, “Bringing up robot: Fundamental mechanisms for creating a self-motivated, self-organizing architecture,” *Cybernetics and Systems*, vol. 36, p. 2005, 2005.
- [8] P. Oudeyer, F. Kaplan, and V. Hafner, “Intrinsic motivation systems for autonomous mental development,” *Evolutionary Computation, IEEE Transactions on*, vol. 11, no. 2, pp. 265–286, 2007.
- [9] A. F. Morse, J. De Greeff, T. Belpaeme, and A. Cangelosi, “Epigenetic robotics architecture (era),” *IEEE Transactions on Autonomous Mental Development*, vol. 2, no. 4, pp. 325–339, 2010.
- [10] P.-Y. Oudeyer, “Intelligent adaptive curiosity: a source of self-development,” pp. 127–130, 2004.
- [11] A. Barto, S. Singh, and N. Chentanez, “Intrinsically motivated learning of hierarchical collections of skills,” in *Proceedings of the 3rd international conference on development and learning (ICDL 2004)*, Salk Institute, San Diego. Citeseer, 2004.
- [12] D. Blank, D. Kumar, L. Meeden, and J. Marshall, “Bringing up robot: Fundamental mechanisms for creating a self-motivated, self-organizing architecture,” *Cybernetics and Systems*, vol. 36, no. 2, pp. 125–150, 2005.
- [13] D. Dennett, *Kinds of minds*. Basic Books, 1996.
- [14] H. Von Foerster, “On constructing a reality,” *Understanding understanding: Essays on cybernetics and cognition*, pp. 211–228, 1988.
- [15] G. Metta, L. Natale, F. Nori, G. Sandini, D. Vernon, L. Fadiga, C. von Hofsten, K. Rosander, M. Lopes, J. Santos-Victor, A. Bernardino, and L. Montesano, “The icub humanoid robot: An open-systems platform for research in cognitive development,” *Neural Netw.*, vol. 23, pp. 1125–1134, October 2010.
- [16] W. R. Ashby, *Design for a Brain*. London: Chapman & Hall, 1960.
- [17] —, “The nervous system as physical machine: With special reference to the origin of adaptive behavior,” *Mind*, vol. 56, no. 221, pp. 44–59, 1947.
- [18] H. Von Foerster, “Perception of the future and the future of perception,” *Instructional Science*, vol. 1, no. 1, pp. 31–43, 1972.
- [19] T. Lenaerts, D. Chu, and R. Watson, “Dynamical hierarchies,” *Artificial Life*, vol. 11, no. 4, pp. 403–405, 2005.
- [20] A. Korzybski, *Science and sanity: An introduction to non-Aristotelian systems and general semantics*. Inst of General Semantics, 1994.
- [21] H. R. Maturana and F. J. Varela, “The Tree of knowledge,” 1998.
- [22] M. Peniak, A. Morse, C. Larcombe, S. Ramirez-Contla, and A. Cangelosi, “Aquila: An open-source gpu-accelerated toolkit for cognitive robotics research,” 2011.