

Action-outcome contingencies as the engine of open-ended learning: computational models and developmental experiments

Gianluca Baldassarre,
Francesco Mannella,
and Vieri Giuliano Santucci

Institute of Cognitive Sciences and Technologies

Eszter Somogyi
Department of Psychology
University of Portsmouth

Lisa Jacquey, Mollie Hamilton,
and J. Kevin O'Regan
Laboratoire Psychologie de la Perception

CNRS UMR 8242

Abstract—Open-ended learning allows humans and robots to autonomously acquire an increasingly large repertoire of skills, that later can allow them to produce suitable actions to achieve desirable effects in the environment ('goals'). Empirical evidence from developmental psychology suggests that a pivotal mechanism possibly driving open-ended learning is represented by action-outcome contingencies. Here we propose a specific hypothesis, expressed in the form of a blueprint cognitive architecture, that sketches the general mechanisms through which contingency-based open-ended learning might take place. According to this hypothesis, the matching (or distance) between a desired goal and the actual effect produced by the action can be used to drive the learning of both the motor skill used to accomplish the goal and the internal representation of the action outcome. We report here a computational model that implements the hypothesis and we illustrate two developmental psychology experiments related to the presented theory. Overall the model and experiments show the soundness of the hypothesis and represent a start towards validating it experimentally.

I. INTRODUCTION

To learn about its environment, an agent must be able to register the statistics that govern its sensory and motor behaviour. Sensory and motor statistics can be split into two kinds [1]. One kind of statistics might be called sensory-sensory contingencies: these correspond to contingencies between perceptual events. Here however we will restrict our consideration to a second kind of statistics, called *sensorimotor contingencies*. These correspond to contingencies between motor commands and the resulting perceptual events.

Developmental psychologists agree that learning must occur through the detection and the exploitation of sensorimotor (or 'action-outcome') contingencies by the infant's brain. Most of the time this idea is an implicit basis underlying more general and complex theories about infant development, notably those starting from behaviourist roots (e.g., [2]), Piaget's wellknown sensorimotor approach [3], Gibson's perspective on affordances [4] and the subsequent work of Von Hofsten [5], the Dynamic System Theory developed by Thelen and Smith [6], and also the recent Bayesian approach to

development. Despite the richness of these theories there has been little theoretical and empirical work that directly tests the hypothesis of the role of sensorimotor contingencies in development. Some exceptions are the approaches using preferential looking of babies (e.g. [9]), using the so-called 'mobile' paradigm (e.g. [10]) or non-nutritive sucking tests (e.g. [11]). There is also work looking closely at the strategies used to choose which sensorimotor contingencies to sample (e.g. [12]). Contingencies might also support the acquisition of motor competences in order to later achieve useful goals (e.g. [13]), and this might rely on mechanisms shared with other primates [14].

On the theoretical/computational side, an important thread of research relevant for contingency based learning involves the computational and robotics models of *open-ended learning* based on *intrinsic motivations* (IMs) [15], [16]. These are motivations supporting the acquisition of knowledge not having an immediate adaptive utility but which might become useful in later stages under the drive of *extrinsic motivations* supporting a direct adaptive/practical advantage [17]. This literature is highlighting the possibility that IMs support learning by working in close synergy with the detection of action-outcome contingencies later supporting goal-directed behaviour. For example, an artificial agent can learn reaching skills that lead to changes in the features of some objects [18], or learn navigation skills that cause the interaction between two objects [19], or can learn visual exploration routines with a motorised camera if they produce some effects in the environment [20].

Other relevant research threads on the computational side are those related to motor babbling and goal babbling. *Motor babbling*, viewing the development of structured behaviour in children as based on the generation of exploratory body movements [21], has often been used in robotics to learn specific tasks rather than for open-ended learning (e.g., [22], [23]). Instead, another view posits that behaviour is always goal-oriented since its early development [24]. Along this line,

in the computational literature *goal babbling* [25], [26] has been proposed to support the learning of multiple tasks. However, these tasks still belong to specific categories that have to be established beforehand by the researcher (e.g., ‘reaching’ goals/tasks). Moreover, they do not face the problem requiring the agent to disentangle the effects produced by the agent’s action from those happening for other causes, such as the environment dynamics or other agents [20], [27].

Instead, we conjecture that *contingency-based learning encompasses general mechanisms able to: (a) support the selfgeneration of goals, and learning of actions, of any type and at any scale; (2) overcome the problem of disentangling the effects produced by the agent’s actions from those happening for other causes.* These issues are expanded in this paper only at the theoretical level while the computational model we present here focusses on developing some detailed mechanisms showing how contingency-based learning can actually work in practice. To this purpose we focus on one class of tasks (reaching body parts) and do not face the issue of agency with the model. The issue of the generality of contingency-based learning with respect to goal/action types and scale, and the issue of agency, will be faced with future models. These issues are a main motivation for this research and the reason why we think contingency-based learning might be so important for autonomous open-ended learning.

Regarding the mechanisms on which contingency-based learning might rely, we posit in essence that it has at its core some events internal to the agent and reflecting the sensorimotor contingencies happening in the environment. Each of these events is in particular a match between the perceived outcome deriving from action execution and the selected goal that produced the action that caused the outcome. This match signal plays a crucial importance as it supports the selection of the goal on which to focus learning resources and guides the both learning of the outcome representation (later possibly becoming a ‘goal’ when activated by internal processes), and the refinement of the action producing the outcome.

A first contribution of this work is the presentation of a blueprint architecture, usable to guide the design of specific computational models and empirical experiments, directed to capture the key elements of contingency-based learning (Sec. II). A second contribution is an illustration of how the blueprint architecture can guide the construction of specific computational models: this is done by overviewing a model illustrated in detail in a companion paper [28] (Sec. III). The last two contributions are the proposal of two developmental psychology experiments (Sec. IV) that provide initial experimental data in line with the presented theory. The first experiment (Sec. IV-A) tests the very basis of the hypothesis, namely that infants are indeed sensitive to sensorimotor contingencies. The second experiment (Sec. IV-B), closely linked to the computational model, involves placing a small

buzzer on the body of a baby to see if this triggers exploration behaviour directed to it. Sec. V draws the conclusions.

II. THE ACTION-OUTCOME CONTINGENCY HYPOTHESIS AND THE BLUEPRINT ARCHITECTURE

The sensory-motor contingency hypothesis presented here states that the agent learns to map both sensory input related

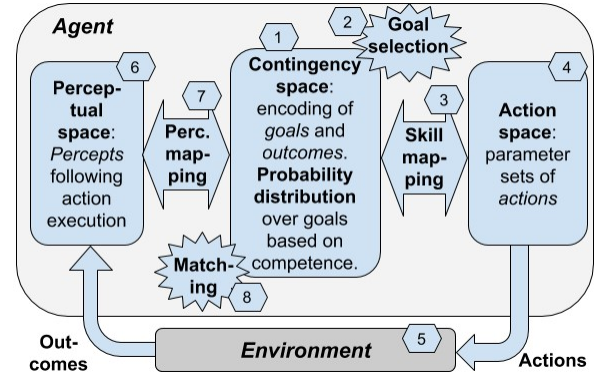


Fig. 1. The blueprint architecture incorporating our hypothesis about the mechanisms underlying open-ended learning of multiple skills. Numbers: sequence of processes happening in one trial of functioning of the system.

to the environment states following action execution (*outcomes*), and the motor behaviour corresponding to the actions, onto a common internal representational space (*contingency space*). Activation patterns within this common space can be triggered both by the sensory input corresponding to the action-outcomes or by internal mechanisms (*goals*) causing the performance of actions. When a correspondence between goals and encoded outcomes takes place within the contingency space (*matching*) the agent self-generates a feedback signalling that an action-outcome contingency between the performed action and the resulting outcome in the environment has taken place. The agent can then compute an estimation of the occurrence of the matching and use it to recursively improve the mapping between the sensory input and the encoded action-outcomes as well as the mapping between the goals and the corresponding motor behaviours.

The blueprint architecture incorporating our hypothesis on the possible mechanisms underlying contingency-based openended learning is sketched in Fig. 1. The architecture pivots on these key elements (the numbers in the figure refer to the architecture processes commented below, but they are also used here to indicate its main elements): (a) *Contingency space*: a pre-wired space which can be discrete or continuous (see number 1 in Fig. 1); (b) *Goals*: internally activated points in the contingency space (number 1); (c) *Skill mapping*: mapping from goals to actions (number 3); (d) *Actions*: behaviour chunks encoded by parameter sets forming points of an *action space* (number 4); (e) *Percepts*: activations of sensors corresponding to action outcomes and forming points

of a *perceptual space* (number 6); (f) *Perception mapping*: mapping from percepts to outcomes, which are encoded in the contingency space (number 7); (g) *Outcomes*: effects of actions in the environment (number 5), and points in the contingency space encoding them (number 1).

We now first illustrate the functioning of this architecture, and then its learning processes. The architecture life is divided in *trials*. For each trial, the following processes take place (see Fig. 1, numbers in hexagons): (1) The contingency space exists before any learning process, so initially it does not have any semantics; (2) At the beginning of each trial the agent *selects and activates a goal* in the contingency space on the basis of a certain probability distribution; (3) The skill mapping, randomly initialised, activates an action; (4) The action is performed in the environment; (5) As an effect of this, the environment returns a percept to the agent (action-outcome); (6) The percept is encoded as a point in the perceptual space; (7) The point in the perceptual space is mapped into a point in the contingency space, the outcome representation; (8) The outcome representation can *match* the goal or not (if the contingency space is discrete), or have a certain degree of similarity with it (continuous contingency space). The matching value for each goal can be used to determine the probability distribution over the goals that allows the agent to select goals on which to focus learning (number 2), for example to select goals for which the competence is low or for which the competence-improvement rate is high.

Regarding the learning processes, the goal-outcome match/mismatch or similarity level (*matching*) can be considered as a measure of the agent's *competence*, i.e. of its capacity to achieve the selected goal based on the performed action. The matching can be used as a *learning signal* to drive the learning processes involving the perceptual and skill mappings. In particular, a successful matching (i.e., a goal-outcome correspondence, or a small distance between goal and outcome) can be used as a reward to adjust the action by reinforcement learning and then to correct the skill mapping. Moreover, the matching signal can also be used to adjust the perceptual mapping so that the outcome matches, or becomes more similar, to the goal. Through these learning processes the contingency space gradually acquires a double semantics. On the skill side, each goal maps to an action, hence acquiring the valence of a 'motor pointer' that allows the agent to trigger suitable actions to accomplish selected goals. On the perceptual side, a goal corresponds to a certain percept representing the world state that the agent desires to accomplish when the goal is selected.

Through these processes, the agent discovers as many as possible outcome states it can produce with its behaviour in the current environment, learns to encode them into the preexisting contingency space, and acquires the motor capabilities needed to accomplish such outcomes when desired (by forming a representation of them and then by internally activating them as goals).

Based on this overview of our hypothesis, it is now possible to go back to the statement proposed in the introduction about the generality of contingency-based learning. First, the mechanism supports the learning of any type of action-outcome (goal) and action (see [29] for the problem of learning any type of outcome). The reason is that both outcomes and actions are encoded and projected onto the contingency space which is neutral with respect to the particular nature and encoding of outcomes and actions. For the same reason, potentially the mechanism might also be used to tackle outcomes and actions at different scales of space/time abstraction, e.g. from 'driving the arm to a certain posture' to 'getting a PhD'. This issue might however possibly require the specialisation of different regions of the contingency space to encode and process knowledge related to different levels of abstraction. Contingency-based learning might also allow the agent to distinguish between the effects it can cause with own action from those happening for other reasons, such as the environment dynamics or other agents. The key mechanism for the attribution of causality to own action with respect to a certain outcome can be summarised by the formula $p(O|A) - p(O|\neg A) > 0$ where p indicates a probability, O the outcomes, A the action, and \neg the negation [7]. To check causality, the blue-print architecture described above could be expanded with a mechanism that checks the presence of own action causality by (a) selecting a goal, (b) checking the probability of the presence of its related world state in absence of action ($p(O|\neg A)$); this is a first element to add to the architecture), (c) perform the action to check its relation with the outcome ($p(O|A)$) based on the matching mechanisms, and then (d) compare $p(O|A)$ and $p(O|\neg A)$ (a second element to add to the architecture). These issues related to the generality of the contingency-based mechanism, considered here only at this theoretical level, should be further investigated with specific computational models in future work.

III. COMPUTATIONAL MODEL

The computational model, designed to study the developmental experiment reported in Sec. IV-B, aims to reproduce the learning processes of a baby that autonomously learns goals and actions to reach and touch parts of its own body surface. For lack of space, here we describe the model at a high level which is however enough to indicate a possible way in which the hypothesis incorporated in the blueprint architecture can be implemented into a specific model. Further details on the system can be found in the companion paper [28].

The baby's body is simulated in a simplified way with a custom simulator reproducing an agent with two 2D 3DoF kinematic arms and a torso (Fig. 2). The life of the system is divided into trials. During a trial, the agent performs an action by setting a sequence of desired postures of the arms: the simulator brings the arms to those postures through standard proportional-derivative controllers. The simulator also

implements 30 touch sensors distributed over the agent's body that activate when touched by one of the two 'hands' of the agent. The activation patterns of these sensors are the outcome states resulting from the execution of the agent's actions and are taken as input by the model.

The first component of the model, the *Outcome Encoder*, encodes sensory touch inputs by clustering them through a self-organising map (SOM; [30]). This clustering serves multiple purposes: (a) reduce the complexity, in particular the dimensionality, of the input space; (b) create similar internal representations to similar inputs; (c) make the system general with respect to the input modalities (e.g., vision might be added to touch); (d) create a discrete representation of the formed clusters that allow the later selection of one (discrete)

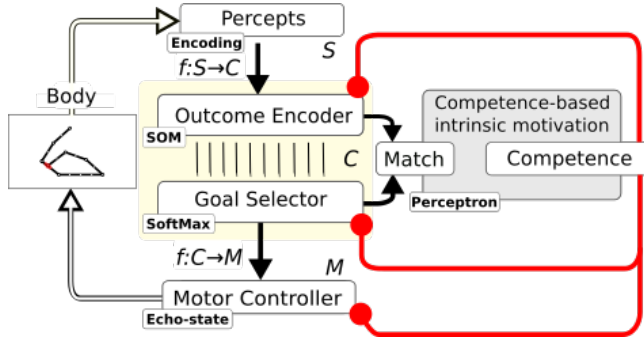


Fig. 2. Architecture of the computational model representing a first implementation of the blueprint architecture hypothesis. The red dotted arrows indicate which component is affected by the competence-based modulation signal during learning. The parallel segments indicate a correspondence between the units of the Outcome Encoder and those of the Goal Selector.

goal. The output layer of the SOM is formed by units (here 25) mapping one-to-one to those of the second component of the system, the *Goal Selector*, and together they form the contingency space. The Goal Selector selects one of its units, encoding a goal, on the basis of a softmax function taking as input a measure of the agent's competence for that goal and generating as output a probability distribution over the (discrete) goal representations. The units selected in the Goal Selector triggers a different motor command if selected. The goal selection process based on intrinsic motivations might work aside other processes for goal selection based on extrinsic motivations [31].

The selected goal activates an action performed in the environment. This is based on a *Motor Controller* component formed by an echo-state neural network [32]. In particular, the output pattern of the Goal Selector is fed into the echo-state network thus biasing its dynamics and causing it to perform a certain action (the read-out layer of the echo-state network is mapped into the arm angles to produce a movement trajectory). The goal fed into the echo-state network is represented through a filter based on radial-basis functions (as usual, these are Gaussian functions that are activated on the

basis of the distance of the selected goal from the equidistant centres of the functions). This would allow the system to interpolate between goals but the current selection of goals, based on a localistic representations goals, should be substituted by one using a continuous representation.

Another component of the system is formed by a *Competence* module computing the competence of the agent for the different goals. This component is formed by a linear neural network that learns to predict the action-outcome contingency success (0/1 event: *Match* signal) related to the selected goal. A contingency success happens when the Encoder and the Goal Selector, that have corresponding units, have the same activation pattern, meaning that the goal (desired outcome) was successfully accomplished. The error between the matching signal and the Predictor output is used to compute the competence level for the selected goal (intrinsic motivation; [33]). The competence measure is used as input to the Goal Selector to select goals with low competence. Given the problem faced, which involves deterministic outcomes and an a-priori knowledge on the maximum achievable performance level (100%), we can use here an IM mechanism based on competence level rather than improvement [34], [35].

Competence also modulates the learning of both the Outcome Encoder and the Motor Controller. The Outcome Encoder learns to encode the perceived outcomes into clusters using the standard algorithms of SOMs [30]. However, this learning process progressively slows down for the goals reaching a higher competence: this protects them against unlearning due to the learning of other goal representations. The Motor Controller learns to perform the correct action by reinforcement learning [36], in particular by storing successful final arm postures leading to a successful matching and by using such postures to train the read-out layer of the echostate network with a standard supervised learning algorithm [32]. Also this learning process progressively slows down for the goals that reach a higher competence, thus protecting the acquired motor skills.

Initially, the activations of the Encoder and Goal generator have a low chance of matching, since the perceptual cluster prototypes and the motor actions have not yet been acquired. With learning, by making use of the agent's competence-based signals, both sensory clusters and actions become increasingly accurate. The key drive of learning is the matching that signals that a contingency has been detected. This contingency in particular involves the activation of a specific goal that triggers the performance of the corresponding action, and this action causes the perception of an outcome – a body touch sensation – belonging to the sensory cluster that corresponds to the goal. All learning processes are performed at each step with the exception of the Goal Predictor that learns at the end of each trial based on the goal chosen at the beginning of the trial and on the matching at the end of the trial. Learning lasts for 8000

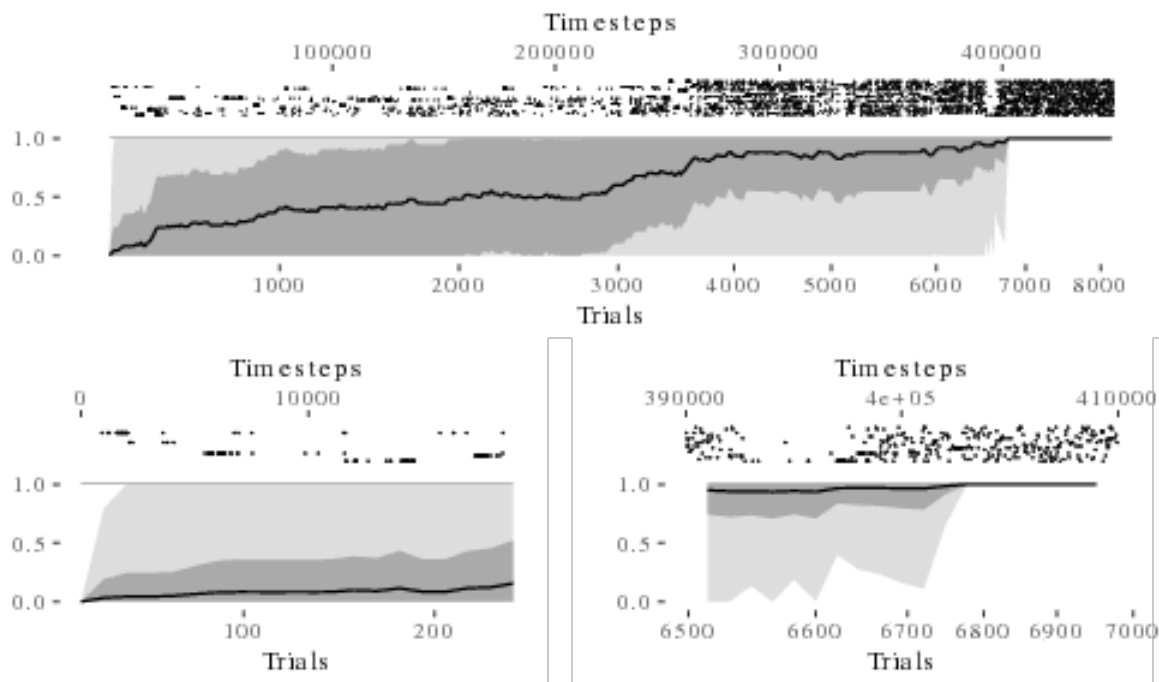
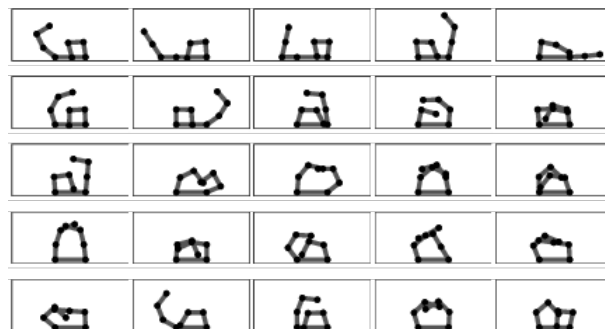
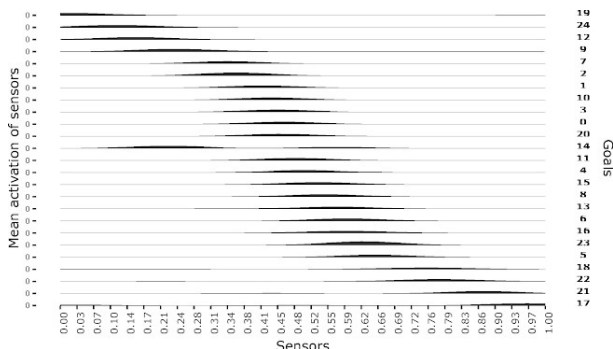


Fig. 3. History of goal competence, measured as the estimate of the Competence component of the selected goal success, plotted over the learning process. Top plot: the black line indicates the average estimate over all 25 goals, the dark gray shadow indicates the standard deviation, and the light gray shadow indicates the worst and best goal/action; the raster plot in the upper part of the graph shows the matching events for each goal, where different rows correspond to different goals. Bottom plots: zoom of respectively the initial phase and convergence phase of the learning process.



trials and each trial ends either with a goal-matching or after a timeout of 100 steps.

The test of the system show that with learning the Outcome Encoder and the Motor Controller converge to an equilibrium where they are closely coupled and the agent manages to accomplish each discovered goal with maximum probability (Fig. 3). In particular, during learning the system tends to repeatedly focus for brief periods of time on different goals and actions until it converges on a stable configuration of them.

At the end of learning, the experienced sensory patterns corresponding to the perceived outcomes are suitably clustered and mapped into the contingency space (Fig. 4), and the points corresponding to them within this space are mapped into the (parameters of) actions able to achieve those outcomes when desired, thus leading to a goal-action-outcome alignment.

Behaviourally, with learning the agent becomes progressively able to produce the actions that lead to successfully accomplish the postures corresponding to the encoded outcomes/goals. This is demonstrated by the fact that after learning if the goal-units of the Goal Selector are externally

Fig. 4. Activation of sensors corresponding to the outcomes and goals of the learned actions. x-axis: body sensors; y-axis: each graph in each row, corresponding to one goal of the SOM, shows the activation level of the 30 body sensors (rows are vertically organised in order of position of the maximum sensor activation from left to right).

activated one after the other, the system produces postures that lead the agent to touch different parts of its body (Fig. 5). Note that in this test the goal units can be thought of as

representing *extrinsic goals* that are activated, say, by a particular need of the agent, for example to reach a particular portion of body to remove a buzzer located on it, as illustrated in the empirical experiments in Sec. IV-B.

Fig. 5. Postures corresponding to goals self-generated by the system and leading the system to touch different parts of its own body.

IV. THE DEVELOPMENTAL PSYCHOLOGY EXPERIMENTS

In order to start to evaluate whether human babies use the type of learning mechanisms described in the previous sections, we first looked to see whether we could demonstrate that babies really are sensitive to sensorimotor contingencies. In a second investigation we checked whether contingencybased learning of body structure could be accelerated by increasing the amount of tactile stimulation that an infant receives. This is what would be predicted from the model, since greater experience with self-touch should facilitate the agent's subsequent ability to reach to a location on the body when this becomes a goal.

A. The contingency experiment

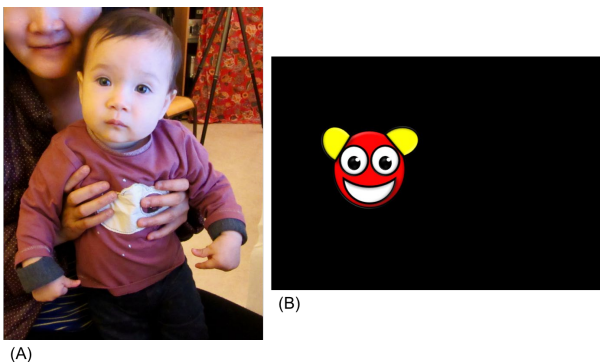


Fig. 6. (A) An 8-month-old infant wearing a bracelet around each wrist. (B) Screenshot of the visual stimulus used. When the infant moved one of its arms, the smiley face moved with speed proportional to the mean acceleration of the infant's arm, changing direction randomly.

In this experiment we equipped babies (aged 4 to 8 months) with bluetooth-enabled bracelets on their wrists that generated a bell sound coupled with movement of a smiley face on the screen, when the infant moved one of its arms (see Fig. 6). For comparison with these 'contingent' infants, we also had a control group of 'non-contingent' infants who were given equivalent audio and visual stimulation, but that was independent of their movements.

We found that, over the 4 minute period of the experiment, infants' arm activity gradually increased, presumably because the infants became excited by the stimulation. However, infants for whom stimulation was contingent on their movements increased their activity more strongly than infants where this was not the case (4 and 6 month olds in Fig. 7). Globally, these results show that infants are sensitive to the sensorimotor contingency linking the arm actions to the smiley motion. Curiously however, we did not find this sensitivity for

infants aged 8 months. This surprising result might be explained by supposing that at 8 months of age, infants are already too old to be interested in our simple contingency. We could obviously test this in the future by further experiments on even older infants. Another interesting result was that infants at all ages seemed not to be able to discover which of the two arms is causing the contingent effect (result not shown in the graphs). We are pursuing further work with more salient contingencies to try to elucidate this surprising negative finding.

B. The buzzer experiment

If, as suggested by the model presented above, infants acquire body know-how through exploration of tactile events, then the more opportunities they have to do so, the earlier they should later be able to rely on this know-how to actually reach for touched locations on their bodies. In order to explore this hypothesis experimentally we designed a longitudinal study to follow the development of infants that received regular non-social tactile stimulation on their body (tactile stimulation condition) and compared the development of their ability to localize touch on their body to the development of infants in a control group that did not receive regular tactile stimulation. Infants in the tactile stimulation condition were not actually trained to reach stimulated areas as movements oriented towards the buzzer were not rewarded or reinforced externally (as opposed to earlier studies, where infants' movements were reinforced by the movement of a mobile above their head, see for instance: Watanabe and Taga, 2009). Infants in this group merely had the opportunity to explore the effect of tactile stimulation on their body and thus learn about their body structure through self-touch.

We visited infants both groups of infants weekly in their homes during the period when the infants were 4 to 7 months of age. In the tactile stimulation condition, each visit consisted of a tactile stimulation session in which buzzers were attached for about 30 seconds to the infant's hands, feet, knees and abdomen, one body part at a time (Fig. 8). In the control condition, infants received identical sessions with the only difference that the experimenter did not actually attach the buzzers to their bodies, only approached the infants with the buzzer. With this design, we controlled for the effect of the mere presence of the experimenter. Ten infants were followed longitudinally in each group.

Our preliminary results confirm our expectations from the model. Starting from age 6 months, infants who received weekly tactile stimulation were better able to reach for the buzzers than the infants who had no stimulation and this difference became significant by 7 months (Fig. 9).

V. CONCLUSIONS

The first contribution of this work was to propose, in the form of a blueprint architecture, a hypothesis on contingencybased mechanisms possibly underlying open-ended learning of multiple goals and actions. According to the

hypothesis, such learning can rely on a contingency space that progressively comes to encode at the same time: goals, that when activated trigger the performance of certain actions, and sensory outcomes, resulting from the performance of those actions. The hypothesis also states that the selection of the goals, and the learning of the two mappings from the outcome sensations to the contingency space and from this space to the actions, can be guided by a competence signal computed on the basis of the internal detection of the contingency. This internal contingency corresponds to the fact that a selected goal, represented in the contingency space, triggers the

performance of an action that causes an outcome in the environment that has an internal representation, within the same contingency space, that matches the goal.

We have here presented a first implementation of this hypothesis using a specific computational model. The model uses a self-organising map (SOM) to encode the sensations produced by actions and an echo-state network to produce actions. The discrete contingency space of the system is represented by the output layer of the SOM mirrored by a layer of units representing goals selected with a probability distribution based on the competence to achieve those goals.

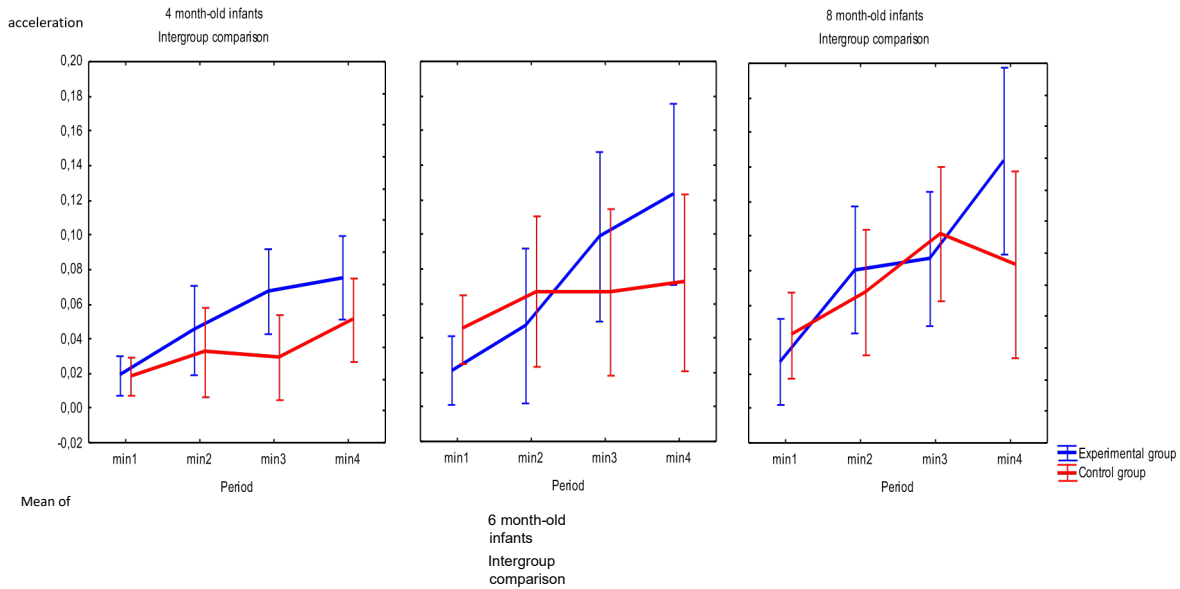


Fig. 7. Activity of infants' arms (measured in multiples of earth's acceleration g) over the 4 minutes of the experiment, for the contingent and non-contingent groups at 4, 6 and 8 months of age. The error bars represent one standard error on either side of the mean. The ANOVA shows that the slopes of the blue (contingent) lines differs significantly from that of the red (non-contingent) lines for the 4 and 6 month old infants.

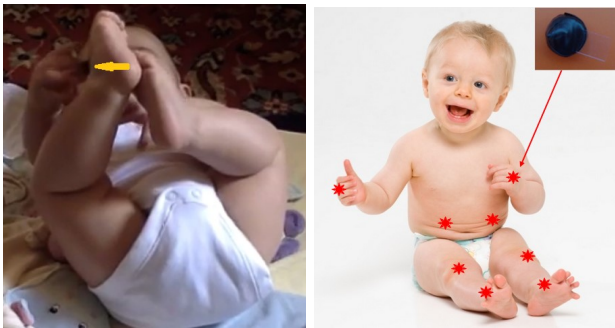


Fig. 8. (Left) A 6-month-old infant collecting a small 'buzzer' (a vibrotactile target, indicated by the yellow arrow) from her right foot during test. (Right) Illustration of the buzzer and the locations stimulated in the 'tactile stimulation' group.

The matching between the selected goal and the actually experienced sensorial action-outcome is used to compute a measure of competence which is in turn used to train the encoding of action-outcomes and of the motor actions. Although preliminary, the tests of the model demonstrate the computational viability of the proposed hypothesis.

The two empirical experiments presented here represented first steps towards making a close link between the computational hypothesis and actual infant behaviour. Unfortunately in our first experiment, which was designed to test the very basis of our hypothesis, namely that sensorimotor contingencies can drive learning in infants, we found it much more difficult than

Body know-how development as a function of tactile experience

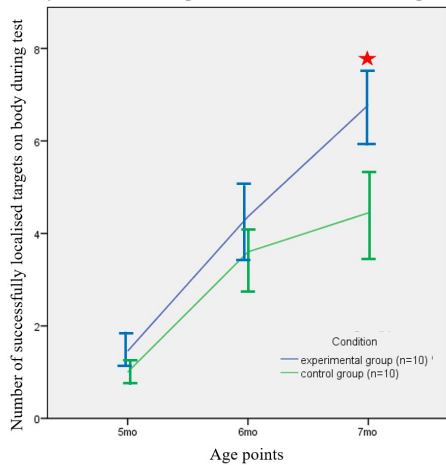


Fig. 9. Ability of infants to reach for buzzers for the experimental (who received weekly tactile stimulation with the buzzer) and control groups at 5, 6 and 7 months of age (n=10 in each group). The error bars represent one standard error on either side of the mean.

expected to find a clear evidence. We are now pursuing further work to improve the paradigm by making the stimuli more salient and the baby more attentive.

The second experiment was also a first attempt to test what would seem like an obvious prediction of the model, namely that goals (in this case buzzer locations) would be more easily attained if prior self-touch experiences had sensitized the particular locations stimulated, thus facilitating contingencybased learning. We did indeed confirm this prediction. However such a prediction would probably also be made by any reasonable model of sensorimotor learning. Our future work will hence involve finding more precise predictions of the model that would differentiate it from other possible models – in particular from models which do not make use of the notion of ‘goal’. We will then attempt to find a refinement of our experimental paradigm with buzzers in order to test such more critical predictions.

ACKNOWLEDGMENT

This research has received funding from the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement No. 713010, Project ‘Goal-Robots – Goal-based Open-ended Autonomous Learning Robots’. K. O’Regan also acknowledges partial support from ERC Advanced Grant ‘FEEL’, No. 323674.

REFERENCES

[1] G. M. Tarabulsy, R. Tessier, and A. Kappas, “Contingency detection and the contingent organization of behavior in interactions: Implications for socioemotional development in infancy.” *Psychological Bulletin*, vol. 120, no. 1, pp. 25–41, 1996.

[2] J. S. Watson, “Perception of contingency as a determinant of social responsiveness,” in *Origins of the infant’s social responsiveness*, 1979, vol. 1, pp. 33–64.

[3] J. Piaget, *The Origins of Intelligence in Children*. London: Routledge and Kegan Paul, 1953.

[4] J. J. Gibson, *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin, 1979.

[5] C. von Hofsten, “Action in development,” *Dev Sci*, vol. 10, no. 1, pp. 54–60, 2007.

[6] E. Thelen and L. B. Smith, *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: The MIT Press, 1996.

[7] A. Gopnik, C. Glymour, D. M. Sobel, L. E. Schulz, T. Kushnir, and D. Danks, “A theory of causal learning in children: causal maps and bayes nets.” *Psychological Review*, vol. 111, no. 1, p. 3, 2004.

[8] J. B. Tenenbaum, C. Kemp, T. L. Griffiths, and N. D. Goodman, “How to grow a mind: statistics, structure, and abstraction,” *Science*, vol. 331, no. 6022, pp. 1279–1285, 2011.

[9] P. Rochat and R. Morgan, “Spatial determinants in the perception of selfproduced leg movements in 3-to 5-month-old infants.” *Developmental Psychology*, vol. 31, no. 4, p. 626, 1995.

[10] C. K. Rovee-Collier and M. J. Gekoski, “The economics of infancy: A review of conjugate reinforcement,” in *Advances in child development and behavior*. Elsevier, 1979, vol. 13, pp. 195–255.

[11] A. J. DeCasper and M. J. Spence, “Prenatal maternal speech influences newborns’ perception of speech sounds,” *Infant behavior and Development*, vol. 9, no. 2, pp. 133–150, 1986.

[12] C. Kidd, S. T. Piantadosi, and R. N. Aslin, “The goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex,” *PLoS one*, vol. 7, no. 5, p. e36399, 2012.

[13] F. Taffoni, E. Tamilia, V. Focaroli, D. Formica, L. Ricci, G. Di Pino, G. Baldassarre, M. Mirolli, E. Guglielmelli, and F. Keller, “Development of goal-directed action selection guided by intrinsic motivations: an experiment with children,” *Experimental Brain Research*, vol. 232, no. 7, pp. 2167–2177, 2014.

[14] E. Polizzi di Sorrentino, G. Sabbatini, V. Truppa, A. Bordonali, F. Taffoni, D. Formica, G. Baldassarre, M. Mirolli, E. Guglielmelli, and E. Visalberghi, “Exploration and learning in capuchin monkeys (*sapajus spp.*): the role of action-outcome contingencies,” *Animal Cognition*, pp. e1–8, 2014.

[15] G. Baldassarre and M. Mirolli, Eds., *Intrinsically motivated learning in natural and artificial systems*. Berlin: Springer, 2013.

[16] P.-Y. Oudeyer and F. Kaplan, “What is intrinsic motivation? a typology of computational approaches.” *Frontiers in Neurobotics*, vol. 1, no. 6, pp. e1–14, 2007.

[17] G. Baldassarre, “What are intrinsic motivations? a biological perspective,” in *Proceedings of the International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob-2011)*, 2011, pp. e1–8, frankfurt am Main, Germany, 24-27 August.

[18] V. G. Santucci, G. Baldassarre, and M. Mirolli, “GRAIL: A goaldiscovering robotic architecture for intrinsically-motivated learning,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 8, no. 3, pp. 214–231, 2016.

[19] T. D. Kulkarni, K. R. Narasimhan, A. Saeedi, and J. B. Tenenbaum, “Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation,” arXiv preprint 1604.06057, 2016.

[20] V. Sperati and G. Baldassarre, “A bio-inspired model learning visual goals and attention skills through contingencies and intrinsic motivations,” *IEEE Transactions on Cognitive and Developmental Systems*, no. 99, 2017.

[21] A. N. Meltzoff and M. K. Moore, “Explaining facial imitation: A theoretical model,” *Infant and child development*, vol. 6, no. 3-4, pp. 179–192, 1997.

[22] D. Caligiore, T. Ferrauto, D. Parisi, N. Accornero, M. Capozza, and G. Baldassarre, “Using motor babbling and hebb rules for modeling the development of reaching with obstacles and grasping,” in *International Conference on Cognitive Systems (CogSys2008)*, 2008, pp. e1–8, Karlsruhe, Germany, 2-4 April 2008.

- [23] V. C. Meola, D. Caligiore, V. Sperati, L. Zollo, A. L. Ciancio, F. Taffoni, E. Guglielmelli, and G. Baldassarre, "Interplay of rhythmic and discrete manipulation movements during development: A policy-search reinforcement-learning robot model," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 8, no. 3, pp. 152–170, 2016.
- [24] C. Von Hofsten, "Eye–hand coordination in the newborn," *Developmental psychology*, vol. 18, no. 3, pp. 450–461, 1982.
- [25] M. Rolf, J. J. Steil, and M. Gienger, "Goal babbling permits direct learning of inverse kinematics," *IEEE Transactions on Autonomous Mental Development*, vol. 2, no. 3, pp. 216–229, 2010.
- [26] —, "Online goal babbling for rapid bootstrapping of inverse models in high dimensions," in *IEEE International Conference on Development and Learning (ICDL2011)*, vol. 2, 2011, pp. 1–8.
- [27] D. Pathak, P. Agrawal, A. A. Efros, and T. Darrell, "Curiosity-driven exploration by self-supervised prediction," 2017, arXiv:1705.05363.
- [28] F. Mannella, V. G. Santucci, S. Eszter, L. Jacquey, K. J. O'Regan, and G. Baldassarre, "Know your body through intrinsic goals," *Frontiers in Neurorobotics*, vol. 12, no. 30, pp. e1–17, 2018.
- [29] E. Cartoni and G. Baldassarre, "Autonomous discovery of the goal space to learn a parameterized skill," arXiv [cs.AI] 1805.07547v1, 2018.
- [30] T. Kohonen, *Self-organizing maps*, 3rd ed. Berlin: Springer, 2001.
- [31] G. Baldassarre, F. Mannella, V. G. Fiore, P. Redgrave, K. Gurney, and M. Mirolli, "Intrinsically motivated action-outcome learning and goal-based action recall: A system-level bio-constrained computational model." *Neural Networks*, vol. 41, pp. 168–187, 2013.
- [32] H. Jaeger, "The 'echo state' approach to analysing and training recurrent neural networks-with an erratum note," German National Research Center for Information Technology, Bonn, Germany, GMD Report 48, 2001.
- [33] V. G. Santucci, G. Baldassarre, and M. Mirolli, "Which is the best intrinsic motivation signal for learning multiple skills?" *Frontiers in Neurorobotics*, vol. 7, pp. e1–14, 2013.
- [34] J. Schmidhuber, "a possibility for implementing curiosity and boredom in model-building neural controllers," in *Proc. of the International Conference on Simulation of Adaptive Behavior: From Animals to Animats*, J. Meyer and S. Wilson, Eds. MIT Press, 1991.
- [35] V. G. Santucci, G. Baldassarre, and M. Mirolli, "Intrinsic motivation mechanisms for competence acquisition," in *Proceeding of the IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob 2012)*, 2012, pp. 1–6, San Diego - CA, USA, 7-9 November 2012.
- [36] R. S. Sutton and A. G. Barto, *Reinforcement learning: an introduction*. Cambridge, MA: The MIT Press, 1998.