The Project IM-CLeVeR – Intrinsically Motivated Cumulative Learning Versatile Robots: A Tool-box for Research on Intrinsic Motivations and Cumulative Learning

Project website:www.im-clever.eu

Authors:
Gianluca Baldassarre, Vieri Santucci, Marco Mirolli, Elisabetta Visalberghi, Eugenio Guglielmelli, Flavio Keller, Peter Redgrave, Kevin Gurney, Martin McGuinnity, Jochen Triesch, Mark Lee, Juergen Schmidhuber, Andrew Barto

2013

Contact: Gianluca Baldassarre (Coordinator), Consiglio Nazionale delle Ricerche, Istituto di Scienze e Tecnologie della Cognizione (CNR-ISTC) Phone: +39 06 44595231 E-mail: gianluca.baldassarre@istc.cnr.it

Table of Contents

Introduction

The goal of this paper is to furnish a *tool-box for research on intrinsic motivations and cumulative learning* based on the main ideas produced within the Integrated Project "IM-CLeVeR – Intrinsically Motivated Cumulative Learning Versatile Robots".

IM-CLeVeR is a project funded by the European Commission under the 7th Framework Programme (FP7/2007-2013), "Challenge 2 - Cognitive Systems, Interaction, Robotics", grant agreement No. ICT-IP-231722. Much information on the project can be found in its web-site: *www.im-clever.eu*.

We thank the Project Officer Cécile Huet, and the Project Evaluators, Luc Berthouze, Ben Kuipers, and Yasuo Kuniyoshi, for their guidance and steering that have importantly contributed to the achievements illustrated this document.

The IM-CLeVeR project

The following web-page of the project web-site reports an introduction to the project concept and objectives
http://www.im-clever.eu/project/project-description
This web-page also contains the links to download a number of presentations that illustrate the project goals and also the work carried out by partners during the four years of the projects.

Key publications external to the project

This section introduces the key papers on intrinsic motivations and hierarchical actions that represented important background knowledge for the IM-CLeVeR research and can serve as an entry point for those new to this research field. Several publications listed below are classic works.

Empirical experiments

Berlyne, D. (1960), Conflict, Arousal and Curiosity (McGraw Hill, New York)
A key seminal work on the role of intrinsic motivation in behaviour.

Harlow, H. F., Harlow, M. K., & Meyer, D. R. (1950). Learning motivated by a manipulation drive. Journal of Experimental Psychology, 40, 228-234.
This seminal work provides a clear example of how animals may learn how to solve tasks even in the absence of extrinsic reinforcement. Action acquisition is discussed in terms of intrinsically-motivated manipulation drive.

Hughes, R. (1997) Intrinsic exploration in animals: motives and measurement. Behavioural Processes 41, 213–226.
In this paper the author introduces the concept of intrisic motivation in animals, discusses the various attempts to measure this phenomenon in laboratory settings and the tests that may provide reasonably valid measures of intrinsic exploration.

Redgrave, P., Gurney, K (2006). The short-latency dopamine signal: a role in discovering novel actions? Nat Rev Neurosci, 7(12):967–975.
This paper argues that the sensory processing providing the input to generate phasic dopamine signals seems unsuited to reinforce the maximisation of future reward. It suggests, on the basis of anatomical projections and signal timing, that phasic dopamine could be used to reinforce the discovery of agency and the development of novel actions rather than extrinsic rewards.

Ryan, R. M. and Deci, E. L. (2000), Intrinsic and extrinsic motivations: Classic definitions and new directions., Contemporary Educational Psychology, 25, 1, 54–67
The key reference for psychological definitions of intrinsic and extrinsic motivations.

White, R. (1959), Motivation reconsidered: the concept of competence., Psychological Review, 66, 297–333
One of the first descriptions of intrinsically motivated behaviours in animals with a stress on action rather than of particular features of stimuli. Very important for the concept of competence based intrinsic motivations.

Models of behaviour and brain

Doll, BB, Simon, DA, Daw, ND (2012). The ubiquity of model-based reinforcement learning. Curr Opin Neurobiol 22:1075-1081.
A recent paper that discusses the conceptual distinction between model-based and model-free reinforcement

learning and how these modes of control may be realised in the brain.

Doya, K (2000). Complementary roles of basal ganglia and cerebellum in learning and motor control. Curr Opin Neurobiol, 10(6):732–739.
A paper proposing models of different brain systems relevant for the hierarchical organisation of actions in real brains.

Glimcher, PW (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. Proc Natl Acad Sci U S A 108:15647-15654.
A clear exposition of the reward prediction error hypothesis of phasic dopamine signalling which is coming under increasing scrutiny.

Vernon, Hofsten, Fadiga (2010). A Roadmap for Cognitive Development in Humanoid Robots. Springer
This book looks at applying our knowledge of development in human infants to humanoid robots. It develops a roadmap of forty-three guidelines for the design of a cognitive architecture and includes a case study on the iCub robot.

Machine learning and robotic systems

Asada, Hosoda, Kuniyoshi, Ishiguro, Inui, Yoshikawa, Ogino, Yoshida (2009). Cognitive Developmental Robotics: A Survey. IEEE Transactions on Autonomous Mental Development, 12--34)
This paper studies a range of cognitive developmental robot (CDR) projects and proposes a model of cognitive development all the way from prenatal sensory-motor mapping through to social behaviour.

Barto, AG, Singh, S, Chentanez, N (2004). Intrinsically motivated learning of hierarchical collections of skills. International Conference on Developmental Learning (ICDL), ed. by Triesch, Jochen and Jebara, Toni, pp. 112-119, LaJolla, CA, UCSD Institute for Neural Computation (ISBN: 0-615-12704-5).
A pioneering work on competence-based intrinsic reinforcement learning, grounded on the option framework.

Kaelbling, L.P., Littman, M.L., Moore, A.W. (1996). Reinforcement learning: A survey. J. Artificial Intelligence Res., 4:237–285.
Good introduction to reinforcement learning, also pointing to several key works.

Metta, G, Sandini, G, Vernon, D, Natale, L, and Nori, F (2008). The iCub humanoid robot: an open platform for research in embodied cognition In: Performance Metrics for Intelligent Systems Workshop (PerMIS 2008), ed. by Madhavan, R. and Messina, E., ACM (ISBN: 978-1-60558-293-1).
The iCub robotic platform, very important for studying intrinsic motivations.

Nehmzow, U, Iglesias, R, Kyriacou, T, Billings, S (2006). Robot learning through task identification. Robotics and Autonomous Systems, 54(9):766-778.
On system identification techniques for autonomous robotics.

Oudeyer, P.-Y., Kaplan, F., Hafner, V.V., (2007) "Intrinsic Motivation Systems for Autonomous Mental Development," Evolutionary Computation, IEEE Transactions on , vol.11, no.2, pp.265,286.
Exploratory activities seem to be intrinsically rewarding for children and crucial for their cognitive development. After discussing related research coming from developmental psychology, neuroscience, developmental robotics, and active learning, this paper presents the mechanism of Intelligent Adaptive Curiosity, an intrinsic motivation system which pushes a robot towards situations in which it maximizes its learning progress.

Oudeyer P.-Y, Kaplan, F. (2008). "How can we define intrinsic motivation?", in "proceedings of the 8th international conference on epigenetic robotics : modeling cognitive development in robotic systems"

This paper presents a unified definition of intrinsic motivation, based on the theory of Daniel Berlyne. Based on this definition, a landscape of types of computational approaches is proposed, making it possible to position existing and future models relative to each other, and to show that important approaches are still to be explored.

Schembri, M, Mirolli, M, Baldassarre, G (2007). Evolving childhood's length and learning parameters in an intrinsically motivated reinforcement learning robot. Proceedings of the Seventh International Conference on Epigenetic Robotics (EpiRob2007), ed. by Berthouze, Luc and Dhristiopher, G.Prince and Littman, Michael and Kozima, Hideki and Balkenius, Christian, pp. 141-148.
A hierarchical RL system composed of experts, which develops in two learning phases based on actor-critic RL: childhood and adulthood. It exploits intrinsically motivated acquired skills (childhood) to obtain extrinsic rewards (adulthood) produced by a genetic algorithm. One of the first works on competence-based intrinsic motivations exploiting the TD RL signal.

Schmidhuber, J (1991). A possibility for implementing curiosity and boredom in model-building neural controllers. Proceedings of the International Conference on Simulation of Adaptive Behavior: From Animals to Animats, ed. by J. A. Meyer and S. W. Wilson, pp. 222-227, MIT Press/Bradford Books.
A pioneering work on curiosity-driven agents and intrinsic reinforcements based on the errors of a predictor.

Sutton, R.S., Barto, A.G. (1998). Reinforcement Learning: An Introduction. MIT Press.
The standard book on reinforcement learning, technically accessible but covering all the key aspects on the topic.

Key publications produced during the project

This section lists and briefly explains the main contribution of the key papers produced during the project.

Empirical experiments

Bednark JG, Reynolds JN, Stafford T, Redgrave P, Franz EA (2013) Creating a movement heuristic for voluntary action: electrophysiological correlates of movement-outcome learning. Cortex 49:771-780.
This paper is an example of how the joy-stick task can be modified to investigate the electrophysiological correlates of action outcome learning.

Manrique, HM, Sabbatini, G, Call, J, Visalberghi, E (2011). Tool choice on the basis of rigidity in capuchin monkeys. Animal Cognition 14:775-786
The paper shows that capuchin monkeys facing an out-of-reach reward efficiently used information previously gathered about tool affordances in the absence of an extrinsic reward.

Redgrave P, Rodriguez M, Smith Y, Rodriguez-Oroz MC, Lehericy S, Bergman H, Agid Y, DeLong MR, Obeso JA (2010) Goal-directed and habitual control in the basal ganglia: implications for Parkinson's disease. Nature Reviews Neuroscience 11:760-772.
This high profile review with world leading clinicians specialising in Parkinson's disease considers the implications of the territorial segregation of function within the basal ganglia for interpreting the symptoms of Parkinson's disease.

Redgrave P, Vautrelle N, Reynolds JN (2011) Functional properties of the basal ganglia's re-entrant loop architecture: selection and reinforcement. Neuroscience 198:138-151.
This is an update of the original Nature Reviews Neuroscience publication (2006) which implicates the basal ganglia architecture with an appreciation of agency and the development of novel actions.

F. Taffoni, D. Formica, G, Schiavone, M. Scrocia, A. Tomassetti, E. Polizzi di Sorrentino, G. Sabbatini, V. Truppa, F. Mannella, V. Fiore, M. Mirolli, G. Baldassarre, E. Visalberghi, F. Keller, E. Guglielmelli (2013) The

"Mechatronic Board": a tool to study intrisic motivation in humans, monkeys, and humanoid robots, in Baldassarre, G. ; Mirolli, M (Eds.) Intrinsically Motivated Learning in Natural and Artificial Systems, pp 411-432. Springer Berlin Heidelberg

The book chapter illustrates the mechatronic board and its use to test IM learning, providing examples with monkeys, children and robots.

Taffoni, F. et al (2013) The "Mechatronic Board": a tool to study intrinsic motivations in humans, monkeys, and humanoid robots, in Baldassarre, G.; Mirolli, M. (Eds.), Intrinsically Motivated Learning in Natural and Artificial Systems, pp. 411-432. Springer Berlin Heidelberg.

In this work is presented a multidisciplinary design approach of platform for behavioural analysis. The platform developed for empirical experiment in the IM-CLeVeR project is presented and input coming from neuroscientists, psychologists, primatologists, roboticists, and bioengineers, discussed.

Taffoni, F. et al. (2012) "A mechatronic Platform for behavioural analysis of Non Human Primates", Journal of Integrative Neuroscience, Vol. 11(1): 87-101.

In this work the main requirement of a mechatronic platform for behavioural studies of non-human primates is described and an example of its use with capuchin monkey without extrinsic reward presented.

Taffoni, F. et al (2012) , "A Mechatronic Platform for Behavioral Studies on Infants" , in procs of the 4th IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics (BIOROB 2012), pp. 1874-1878.

This article report preliminary results of pilot experiments carried out in 2010-2011 on 12 children aged between 24-68 months. This data will be used to refine the protocol adding a different control condition and reducing age range.

Taffoni, F., Vespignani M., Formica D., Cavallo G., Polizzi di Sorrentino E., Sabbatini G., Truppa V., Visalberghi E., Keller F., Guglielminelli E., (2012) A mechatronic platform for behavioral analysis of nonhuman primates. Journal of Integrative Neuroscience, 11: 87-101

The paper presents the mechatronic board, an innovative tool for behavioural analysis, providing examples of its use with non-human primates.

Truppa, V., Piano Mortari, E., Garofoli, D., Privitera S. & Visalberghi, E. (2010). Identity concept learning in matching-to-sample tasks by tufted capuchin monkeys (Cebus apella). Animal Cognition. 13:835–848

This study investigates cumulative learning processes in a concept learning task. The ability of capuchin monkeys to learn same and different concepts is investigated by testing whether subjects use them to solve a relational matching-to-sample task.

Visalberghi, E., & Fragaszy, D. (2013). The Etho-Cebus Project: Stone-tool use by wild capuchin monkeys. In Tool Use in Animals: Cognition and Ecology (eds Sanz, Call, Boesch), Cambridge University Press.

This book chapter summarizes the studies carried out by the Ethocebus team on tool-using skills of wild capuchin monkeys, and suggests that curiosity-driven exploration contributes to the acquisition of tool use.

Models of behaviour and brain

Baldassarre, G., Mirolli, M., eds. (2013), Intrinsically Motivated Learning in Natural and Artificial Systems (Springer-Verlag, Berlin)

This edited book collects key research articles on intrinsic motivations, both in biological agents and in artificial systems. Each chapter of the book offers a critical review of the main works of the authors of the chapter, and highlights open problems and future research directions.

Baldassare, G.; Mannella, F.; Fiore, V. G.; Redgrave, P.; Gurney, K. & Mirolli, M. (2012), 'Intrinsically motivated action-outcome learning and goal-based action recall: A system-level bio-constrained computational

model.', Neural Networks 41, 168-187

This paper presents the model resulted from the demonstrator CLEVER-B2. The model is a bio-constrained system-level model focused on three processes related to IMs and on the neural mechanisms underlying them: (a) the acquisition of action-outcome associations (internal models of the agent-environment interaction) driven by phasic dopamine signals caused by sudden, unexpected changes in the environment; (b) the transient focussing of visual gaze and actions on salient portions of the environment; (c) the subsequent recall of actions to pursue extrinsic rewards based on goal-directed reactivation of the representations of their outcomes.

Baldassarre G. (2011). What are intrinsic motivations? A biological perspective. In Proceedings of the IEEE Conference on Developmental Learning and Epigenetic Robotics (ICDL2011), e1-8. IEEE.

This theoretical paper aims to clarify what intrinsic motivations are from a biological perspective. It shows how intrinsic motivations can be defined contrasting them to extrinsic motivations from an evolutionary perspective. It proposes that extrinsic motivations generate learning signals on the basis of events involving body homeostatic regulations, whereas intrinsic motivations generate learning signals based on events taking place within the brain itself.

Caligiore, D., Parisi, D., Baldassarre, G. (submitted), 'Integrating Reinforcement Learning, Equilibrium Points and Minimum Variance to Understand the Development of Reaching: A Computational Model', Psychological Review.

This paper is important for the "hierarchical" aspects of IM-CLeVeR project from a biological perspective as it enlightens the functions and the structure of the cortical hierarchy within brain, focussing on its two main neural pathways, the ventral and the dorsal pathway.

Ciancio, A.L. et al (2013), "The role of learning and kinematic features in dexterous manipulation: a comparative study with two robotic hands"International Journal of Advanced Robotic Systems, 2012

The purpose of this study is to investigate the role of thumb opposition during cyclic manipulation tasks through the interaction with different objects and a bio-inspired control architecture based on reinforcement learning. The control architecture has been implemented in simulated environment on two robotic hands with different thumb features, i.e. the iCub hand and the DLR/HIT Hand II, interacting with objects of different sizes and shapes.

Ciancio, A.L. et al (2011) "Hierarchical reinforcement learning and central pattern generators for modeling the development of rhythmic manipulation skills," Development and Learning (ICDL), 2011 IEEE International Conference on , vol.2, no., pp.1,8, 24-27.

In this work a computational bio-inspired model to investigate the development of functional rhythmic hand skills from initially unstructured movements is presented. The model is based on a hierarchical reinforce-ment-learning actor-critic model that searches the parameters of a set of central pattern generators (CPGs) hav-ing different degrees of sophistication.

Mirolli, M., Santucci, V. G. , Baldassarre, G. (2013). 'Phasic dopamine as a prediction error of intrinsic and extrinsic reinforcement driving both action acquisition and reward maximization: A simulated robotic study', Neural Networks 39, 40-51.

This paper presents a unifying hypothesis on the role of phasic dopamine, supported by the implementation of a simulated robotic model. The idea is that phasic dopamine can signals *both* intrinsic and extrinsic motivation learning signals rather than only one of the two.

Stafford T, Thirkettle M, Walton T, Vautrelle N, Hetherington L, Port M, Gurney K, Redgrave P (2012) A novel task for the investigation of action acquisition. Plos One 7:e37749.

This paper introduces the joy-stick task as a new procedure for investigating the multidimensional aspects of novel action acquisition.

Thirkettle M, Walton T, Shah A, Gurney K, Redgrave P, Stafford T (2013) The path to learning: action acquisition is impaired when visual reinforcement signals must first access cortex. Behav Brain Res 243:267-272.

This paper shows that visual reinforcement signals are more effective when they engage both cortical and subcortical sensory processing than when they engage cortical processing alone.

Machine learning and robotic systems

Earland, Law, Shaw, Lee (Submitted) Overlapping Structures in Sensory-Motor Mappings. Plos One
This paper examines the biologically-inspired representation technique used as the learning substrate in most of our other papers. It focuses on the overlapping properties of receptive fields and how they affect accuracy and efficiency.

Frank, M. Leitner, J. Stollenga, M., Kaufmann, G., Harding, S., Förster, A., Schmidhuber, J. (2012). The Modular Behavioral Environment for Humanoids & other Robots (MoBeE). 9th International Conference on Informatics in Control, Automation and Robotics (ICINCO). Rome, Italy. July 2012.
To produce even the simplest human-like behaviors, a humanoid robot must be able to see, act, and react, within a tightly integrated behavioral control system. This paper presents MoBeE, a novel behavioral framework for humanoids and other complex robots, which integrates elements from vision, planning, and control, facilitating the synthesis of autonomous, adaptive behaviors.

Law, Lee, Hulse, Tomassetti (2011). The infant development timeline and its application to robot shaping. Adaptive Behavior, 335—358).
This paper introduces key issues, from a robotics perspective, in developmental learning and produces explicit timelines that display the relative ordering of emergent competencies and concomitant stages in behaviour.

Law, Shaw, Lee (2013). A biologically constrained architecture for developmental learning of eye–head gaze control on a humanoid robot. Autonomous Robots, 1—16.
This paper describes a developmental learning architecture for eye-head gaze control deployed on an iCub robot. The approach aims to acquire sensorimotor competence through growth processes modelled on data and theory from infant psychology.

Lee, Law, Shaw, Sheldon (2012). An Infant Inspired Model of Reaching for a Humanoid Robot. ICDL-EPIROB, 1—6.
This outlines the biological and psychological processes behind the learning of saccade control, gaze control, torso control, and visually elicited reaching and grasping in 3D space. It demonstrates the efficiency of the technique on an iCub, learning reaching behaviours in just 2.5 hours.

Leitner, J., Harding, S., Chandrashekhariah, P., Frank, M., Förster, A., Triesch, J., Schmidhuber, J. (2013). Learning Visual Object Detection and Localisation Using icVision. Biologically Inspired Cognitive Architectures.
The paper presents a framework combining computer vision and machine learning for the learning of object recognition in humanoid robots. A biologically inspired, bottom-up architecture is introduced to facilitate visual perception and cognitve robotics research. An object detection filter is trained by Cartesian Genetic Programming (CGP).

Luciw, M., Kompella*, V., Kazerounian, S., Schmidhuber, J. (under review). An Intrinsic Value System for Developing Multiple Invariant Representations with Incremental Slowness Learning. *Joint First Authors.*
Curiosity Driven Modular Incremental Slow Feature Analysis (CD-MISFA) is a model of intrinsically-motivated invariance learning that combines: unsupervised representation learning through the slowness principle; generation of an intrinsic reward signal through learning progress of the developing features; balancing of exploration and exploitation to maximize learning progress and quickly learn multiple feature sets for perceptual simplification.

Ngo, H., Luciw, M., Förster A., Schmidhuber, J. (2012). Learning skills from play: Artificial curiosity on a katana robot arm. Proceedings of the International Joint Conference on Neural Networks (IJCNN), Brisbane, 10

June 2012.
Artificial curiosity tries to maximize learning progress. We apply this concept to a physical system. The system is intrinsically motivated to explore its world. Our Katana robot arm curiously plays with wooden blocks, using vision, reaching, and grasping. It learns how to place blocks stably, and how to stack blocks without any extrinsic supervision or reward.

Schmidhuber, J. (2012). A Formal Theory of Creativity to Model the Creation of Art. In McCormack J., d'Inverno, M. (Eds.), Computers and Creativity. MIT Press.
A creative agent – one that never stops generating non-trivial, novel, and surprising behaviours and data – must have two learning components: a general reward optimiser or reinforcement learner, and an adaptive encoder of the agent's growing data history. The learning progress of the encoder is the intrinsic reward for the reward optimiser. To maximise expected reward (in the absence of external reward), the reward optimiser will create more and more-complex behaviours that yield temporarily surprising (but eventually boring) patterns that make the encoder quickly improve. This simple principle explains science, art, music and humour.

Sheldon, Lee (2011). PSchema: A developmental schema learning framework for embodied agents. IEEE, Development and Learning (ICDL), 2011 IEEE International Conference, 1—7.
This paper introduces PSchema, an implementation of framework for Piagetian schema learning which allows for the direct use of symbolic schema learning in a robotic environment. The paper demonstrates a unique generalisation technique that significantly increases the schema capabilities.

Stollenga, M., Pape, L., Frank, M., Leitner, J., Forster, A., Schmidhuber, J. (under review) Task-Relevant Roadmaps: A Framework for Humanoid Motion Planning.
Task-Relevant Roadmaps (TRMs) can be used to plan complex task-relevant motions on robots with many degrees of freedom. To this end we create a new sampling based inverse kinematics optimizer called Natural Gradient Inverse Kinematics (NGIK), based on the principled heuristic solver called natural evolu- tion strategies (NES). We show NGIK outperforms recent algorithms in this domain and show the effectiveness of our method on the iCub robot, using the 41 DOF of its full upper body, arms, hands, head, and eyes.

Tommasino, P., Caligiore, D., Mirolli, M., Baldassarre, G. (in preparation), 'Transfer expert reinforcement learning (TERL): a reinforcement learning architecture that transfers knowledge between skills', Neural Networks.
This work presents a bio-inspired reinforcement-learning (RL) modular system developed to enhance skill-to-skill transfer (the model is called TERL – Transfer Expert RL model). The functioning of TERL is shown based on reaching tasks a simulated planar dynamic arm and also a more sophisticated 3D 4DOFs simulated robotic arm.

A reading guide for the "Frontiers" special issue

This section illustrates they key paper published by the project Consortium in the special Research Topic organised by the project for *Frontiers in Cognitive Science* and *Frontiers in Neurorobotics*:
http://www.frontiersin.org/Cognitive_Science/researchtopics/Intrinsic_motivations_and_open/1326
http://www.frontiersin.org/neurorobotics/researchtopics/Intrinsic_motivations_and_open_1/1797
The papers are representative of some of the best works produced by the consortium.

Authors: Fiore, V.G., Sperati, V., Mannella, F., Mirolli, M., Gurney, K., Friston, K., Dolan, R.J., Baldassarre, G.
Title: Keep focussing: striatal dopamine multiple functions resolved in a single mechanism tested in a simulated humanoid robot.
Abstract: This paper illustrates a model derived from the Demonstrator CLEVER-B3 of the project IM-CLeVeR. The effects of striatal dopamine on behaviour have been widely investigated over the past decades, with "phasic" burst firings considered as the key expression of a reward prediction error responsible for reinforcement learning. Less well studied is tonic dopamine, where putative functions include the idea that it is a regulator of vigour, incentive salience, disposition to exert an effort and a modulator of approach strategies. We present a model combining tonic and phasic dopamine to show how different outflows triggered by either intrinsically or extrinsically motivating stimuli dynamically affect the basal ganglia by impacting on a selection process this system performs on its cortical input. The model, which has been tested on the simulated humanoid robot iCub in the interaction with a mechatronic board, shows the putative functions ascribed to dopamine emerging from the combination of a standard computational mechanism coupled to a differential sensitivity to the presence of dopamine across the striatum.

Authors: Frank, M., Leitner, J., Förster, A., Schmidhuber, J.
Title: Embodies, Modular, Intrinsically Motivated Reinforcement Learning for Humanoid Robots.
Abstract: Much of the previous work on Artificial Curiosity and Intrinsic Motivation focused on basic concepts and theory. Most applications were limited to simple scenarios, often in simulated rather than real worlds. We present an applied research project on embodying curiosity in the iCub humanoid robot. From the machine learning standpoint Artificial Curiosity can be implemented as an Intrinsically Motivated Reinforcement Learner (IMRL). For an reinforcement learner the state-action space of the iCub humanoid is far too large to be conquered by a single map for any kind of not trivial task. For this reason, it is essential to modularize our learning system in terms of the natural hierarchy apparent in the kinematics of the human or humanoid and learn the sub-systems independently. In this paper, we will investigate the following question: What happens when we define two separate IMRL Agents, and allow them to interact with one another by controlling different but related parts of the same robot.

Authors: Law, J., Shaw, P., Earland, K., Sheldon, M., Lee, M.
Title: A psychology based approach for longitudinal development in cognitive robotics
Abstract: A major challenge in robotics is the ability to learn, from novel experiences, new behaviour that is useful for achieving new goals and skills. For autonomous systems such learning must be entirely unsupervised, thus ruling out preprograming or extensive training. Developmental constraints that exist on the infant, both sensory and motor, are beneficial and can be exploited in the structuring of behavioural growth stages. We use sensorimotor mappings as abstractions of the neural substrate and find that the overlap between receptive fields to be significantly advantageous. For intrinsic motivation we found that simple novelty detection, based on coincident, unexpected, but repeatable sensorimotor events is a strong and general motivator. Motor babbling behaviour was found to be an important element because correlations can be tested by repeated evaluation and we used a simple novelty-driven excitation mechanism to modulate babbling and direct attention. For long term memory (for e.g. recognition of objects) we use a schema system that can learn and recall schema patterns representing patterns of sensorimotor experience.

Authors: Ngo, H., Luciw, M., Förster A., Schmidhuber, J.
Title: Confidence-Based Progress-Driven Self-Generated Goals for Skill Acquisition in Developmental Robots.
Abstract: Artificial curiosity tries to maximize learning progress. We apply this concept to a physical system. Our Katana robot arm curiously plays with wooden blocks, using vision, reaching, and grasping. It is intrinsically motivated to explore its world. As a by-product, it learns how to place blocks stably, and how to stack blocks.

Authors: Santucci, V.G., Baldassarre, G., Mirolli, M.
Title: Learning Multiple Skills Through Competence-Based Intrinsic Motivations: A Computational Embodied Model
Abstract: Humans and other biological agents are able to autonomously learn and cash different skills in the absence of any biological pressure or any assigned task. Intrinsic Motivations (i.e motivations not connected to reward-related stimuli) seems to play a significant role in these kind of processes. From a computational point of view, Intrinsic Motivations can be considered a useful tool to improve the implementation of more autonomous and more adaptive artificial agents. In this work, we provide an exhaustive analysis of a scarcely investigated problem: which kind of IM reinforcement signal is the most suitable for driving the learning of skills in the shortest time? To this purpose we implemented an artificial agent with a hierarchical architecture that allows to learn and cash different skills and we tested its performance in a setup with continuous states and actions, comparing the performance of the system when driven by different intrinsic motivation signals representative of related literature.

Authors: Stafford, T. and Thirkettle, M.
Title: Linking total movement history to action learning
Abstract: We have developed a new behavioural task which requires the agent to identify a target movement via exploratory movements. This conceptually simple, but versatile, paradigm allows the full history of movements made to be linked to learning of actions. Here we review recent results which show how the task can elucidate different aspects of the functional and anatomical basis of reinforcement learning in the human motor system. We also present a novel analysis which captures the relative influence of movement history on learnt actions, at each point in the path of that learnt action. From a reinforcement learning perspective, this analysis can be thought of as revealing the shape of the eligibility trace: the relative strength of credit assignment to the motor efference copy across time.

Authors: Thirkettle, M., Walton, T, Redgrave, P. Gurney, K. and Stafford, T.
Title: No learning where to go without first knowing where you're coming from: Action discovery is trajectory, not endpoint based.
Abstract: Intrinsic motivations drive an agent to explore, providing essential data for linking behaviours with novel outcomes and so laying the foundation for future flexible action. We introduce experiments using a new behavioural task which allows us to interrogate the connection between exploration and action learning. Human participants used a joystick to search repeatedly for a target location, only receiving feedback on successful discovery. Feedback delay was manipulated, as was the starting position. Experiment 1 employed stable starting positions, so the task could be learnt with respect to a target location or a target trajectory. Participants were able to learn the correct movement under all delay conditions. Experiment 2 used a variable starting location, so the correct movement could only be learnt in terms of target location. Participants displayed no learning in this experiment. These results suggest that movements on this scale are stored as trajectories rather than in terms of target location. Overall the experiments demonstrate the potential of this task for uncovering the native representational substrates of action learning.

The CLEVER-B architectures

At the end of each of the four years of its duration, the project presented a Robotic Demonstrator based on the iCub robot: CLEVER-B1, CLEVER-B2, CLEVER-B3, CLEVER-B4. The four Demonstrators aimed to capture, in system-level models, the key elements of brain putatively underlying intrinsic-motivation driven cumulative learning in animals. The demonstrators are all based on the Board Experiment run in the project with children and monkeys. The Demonstrators are documented with explanations, figures, and videos in the project web-site: http://www.im-clever.eu/documents/demonstrator

The CLEVER-K architectures

At the end of each of the four years of its duration, the project presented a Robotic Demonstrator based on the iCub robot: CLEVER-K1, CLEVER-K2, CLEVER-K3, CLEVER-K4. The four Demonstrators aimed to show the most advanced machine learning techniques developed within the project to implement intrinsically-motivated cumulative learning in robots. The Demonstrators are documented with explanations, figures, and videos in the project web-site:
http://www.im-clever.eu/documents/demonstrator

Software tools

The project produced a large amount of software. Part of this software is made available to the public at the project web-site:
http://www.im-clever.eu/resources/models

Videos of models produced by the project

The project produced several robotic models. The functioning of these models can bee seen in the collection of videos from the project available at the project web-site:
http://www.im-clever.eu/announcement/videos

Other material: summer schools, tutorials, deliverables and presentations

Other material on the project, and information on the community involved in the research of intrinsically motivated cumulative learning, can be found in the web-pages of the project related to its various activities concerning the summer schools, public deliverables, presentations, and university lectures:
http://www.im-clever.eu/documents/

A list of key research questions that still need to be fully investigated

Below we present a list of key questions that we isolated after the 4 years of research of the project on intrinsic motivations and cumulative learning. We think such questions might guide the future research in the field. We divide the questions in three themes: empirical experiments, models of behaviour and brain, machine learning and robotic systems.

Empirical experiments

- How do intrinsic motivations guide children exploration during development? Which is the role of intrinsic motivations in the acquisition of new skills especially in newborns? Does the acquisition of internal models affect the development of intrinsic motivated exploration the environment? This questions might be addressed with experiments that build on the project experiments run with children and monkeys with the mechatronic board.

- How do action-outcome contingencies affect intrinsic motivations to explore the environment in children and

primates and how do they lead to learn new skills? The experiments directed to face this problems may involve tasks where action produces different effects changing in terms of delay, intensity (energy) involved, sizes of the space involved, visual appearance, duration, etc.

- What is the role of intrinsic motivations in cumulative learning in primates? In particular, how can they drive the acquisition of a hierarchically organised set of skills? The experiments directed to face this problems may involve a set of tool-use tasks that need to be solved with sequences or hierarchies of actions.

- The experiments of the project with monkeys and children using yoked controls showed that these controls loose interest in relation to the interaction with the board (as it is no more novel? As their actions do not produce any effect?). How do personality and temperament traits affect the sensitivity to in this cases? Experiments could involve testing the correlation between personality measures on spontaneous exploratipon, curiosity, and perseverance with individuals' performance in the yoked control condition.

- How does animal ecology affect intrinsically-motivated learning? As generalists species need to respond rapidly to changing environments it pays for them to be highly exploratory. By contrast, species with very specific niches may be more selective in choosing what they work for since they have less to gain by acquiring general knowledge. Species with different ecological niches could be tested in a comparative fashion to investigate possible differences in their performance during intrinsic motivation learning tasks.

- Neuroscientific experiments on superior colliculus have revealed that this is at the basis of important instances of intrinsic motivations, in particular those related to phasic dopamine signals evoked by events (e.g., a light suddenly going on). However, there are still open issues to be solved, in particular related to the specific mechanisms that lead to the progressive inhibition of the signal when the event becomes progressively familiar: is this based on prediction or habituation? Does the predictor/habituator directly inhibit the superior colliculus or the downstream dopaminergic areas?

- The specific implementation of other intrinsic motivation mechanisms in brain should be investigated. For example, there is a large literature on the hippocampal system detecting novelty of stimuli. This has been related to memory formation, but it has not been casted within the larger framework of intrinsic motivations while this might enlighten new aspects of the phenomenon. Theoretical analysis and empirical research is needed to uncover these aspects.

Models of behaviour and brain

- Current reinforcement learning models rely on various forms of temporal difference learning and actor-critic architectures to maximise reward acquisition in the long term. Biological systems seem to have multiple reinforcement mechanisms that co-operate to maximise reward acquisition and minimise pain over the long term. For example, in biological systems different mechanisms are required to reinforce biological salient sensory stimuli (stimuli that predict reward or punishment both need to be positively reinforced), and action-selection (actions associated with reward needs to be positively reinforced while actions associated with punishment needs to be negatively reinforced – Thorndike's Law of Effect). Further investigation of biological systems is needed to identify these different mechanisms and the existence of multiple reinforcement mechanisms needs to be recognised and translated into different reinforcement learning algorithms.

- The nature of goal-directed and habitual control also needs more exploration in biological systems. For example, actions, ideas, memories, and emotions can all be subject to goal-directed and habitual control. Further computational exploration of these concepts is also required.

- The concept of 'action' needs to be more thoroughly investigated. In the project, it has become apparent that actions frequently have multiple facets (WHERE, WHAT, WHEN, HOW), which are likely to be acquired in different and specialised neural networks in brain. If so, modelling research is needed to specify the mechanisms that support such different aspects of action. Moreover, the specialisation of action components generates a

'binding problem' analogous to that of perceptual binding: how is this solved?

- Developmental models of early behaviour necessarily construct egocentric spatial representations based on the available sensory and motor data. After a few months infants start to display allocentric spatial awareness. It has been hypothesized that locomotion, object perception, and various other perceptual competencies are involved, but this important question is significantly under investigated. An interesting project would be to study this issue with robotic model, in particular by exploring the requirements and conditions necessary for allocentric representations to be supported by existing egocentric abilities, and examining emergent sensitive periods through robot experiments.

- Biology and theoretical analysis showed that there might be important differences between intrinsic motivations based on novelty and those based on surprise. What are the best computational models to definitely disentangle the two? Can we ground such models on what we know from biology?

- Bio-constrained models of animals learning in a cumulative fashion necessarily involve several brain components and processes (system-level models). This leads to create complex models having several free parameters. This poses difficult methodological challenges: How can these models be easily built, trained, communicated? How can they be best exploited to impact the research of the psychologist and the neuroscientist? How can they be rigorously validated against empirical data?

Machine learning and robotic systems

- There is much current research into motor and goal babbling as intrinsically motivated exploration strategies in developmental robotics, (e.g. A. Rolf et al. and P.-Y. Oudeyer et al.). It has also been hypothesized that play is a related strategy (Lee, Epirob 2011). A significant project would investigate different algorithms for play generation, evaluate their combinatorial properties, and relate the findings to the best models of motor and goal babbling behaviour. This will make a major contribution to knowledge and theory of self-generated, goal-based exploration.

- Staged development is now recognised as an effective strategy for learning in developmental robotics. Naturally occurring constraints influence the emergence of stages in behaviour and are under active research. However, human infants display "sensitive periods" during which skills must be learned or are lost for ever. A valuable project would be to search for evidence of sensitive periods in staged robotic development, build them into current theoretical models of development, and investigate their role and significance.

- Improve online-learning frameworks that bridge the gap between abstract, preprocessed sensor input, on one side, and a reinforcement intrinsically motivated learner, on the other side. This will enable robots to build better models of the environment and at the same time improve their interactive abilities.

- Develop benchmarks and define criteria and metrics to evaluate the efficiency of intrinsically motivated agents and to compare their performance. In contrast to reinforcement learners the goal of an intrinsically motivated learner is to acquire knowledge about the world through curious exploration without externally-posed tasks. The benchmarks should consider this fundamental difference.

- Study better practical adaptive data encoders to represent the learned world models of intrinsically motivated learners and investigate under which conditions learning progress can be measured both accurately and efficiently.

- Multiple curious agents can act in the same environment where they influence each other. The interaction might drive the agents to explore more interesting areas than by acting alone. How can we implement this idea in robots? Moreover, can we make more progress toward knowledge acquisition by developing hierarchies of curious agents?

- The research carried out in these four years showed that cumulative learning of different skills has to rely on

hierarchical architectures that support a faster learning of new skills on the basis of already-acquired skills. What are the best architectures to implement transfer of knowledge, for example for transfer reinforcement learning?

- The last two years of the project showed the paramount importance of pivoting on *goals* for implementing cumulative learning. An important challenge for intrinsic motivations is explaining how an autonomous agent can self-generate goals without external support.

- What are the key ingredients ultimately needed to produce autonomous cumulative learning versatile robots? The project highlighted that these ingredients might be intrinsic motivations (prediction-based, novelty-based, competence-based) and hierarchical architectures (for the acquisition of skills and goals, and hierachies of skills and goals). However, much research is needed to refine the mechanisms needed to implement them.