Contents lists available at ScienceDirect

Cognitive Systems Research

journal homepage: www.elsevier.com/locate/cogsys

Conceptual development from the perspective of a brain-inspired robotic architecture

Ioanna Giorgi [1],*, Bruno Golosio [2], Massimo Esposito [3], Angelo Cangelosi [4], Giovanni Luca Masala [1]

[1] University of Kent, Canterbury, UK
[2] Department of Physics, University of Cagliari, and the National Institute for Nuclear Physics, Cagliari, Italy
[3] Institute for High-Performance Computing and Networking, National Research Council; Naples, Italy
[4] University of Manchester, Manchester, UK

ARTICLE INFO

Action editor: Felix Ramos

Keywords:
Brain-inspired model
Cognitive architecture
Robotic model
High-level cognition
Human language
Concepts
Categorisation
Conceptual development
Understanding

ABSTRACT

Concepts are central to reasoning and intelligent behaviour. Scientific evidence shows that conceptual development is fundamental for the emergence of high-cognitive phenomena. Here, we model such phenomena in a brain-inspired cognitive robotic model and examine how the robot can learn, categorise, and abstract concepts to voluntary control behaviour. The paper argues that such competence arises with sufficient conceptual content from physical and social experience. Hence, senses, motor abilities and language, all contribute to a robot's intelligent behaviour. To this aim, we devised a method for attaining concepts, which computationally reproduces the steps of the inductive thinking strategy of the Concept Attainment Model (CAM). Initially, the robot is tutor-guided through socio-centric cues to attain concepts and is then tested consistently to use these concepts to solve complex tasks. We demonstrate how the robot uses language to create new categories by abstraction in response to human language-directed instructions. Linguistic stimuli also change the representations of the robot's experiences and generate more complex representations for further concepts. Most notably, this work shows that this competence emerges by the robot's ability to *understand* the concepts similarly to human understanding. Such understanding was also maintained when concepts were expressed in multilingual lexicalisations showing that labels represent concepts that allowed the model to adapt to unfamiliar contingencies in which it did not have directly related experiences. The work concludes that language is an essential component of conceptual development, which scaffolds the cognitive continuum of a robot from low-to-high cognitive skills, including its skill to understand.

1. Introduction

The way humans acquire, represent and pass on knowledge is subject to cross-disciplinary debate. It is advocated that the body of knowledge we store in memory and use in high-cognitive activities is commonly linked to our ability to form concepts (Machery, 2009). Conceptual development grants human organisms plasticity to evolve and adapt to miscellaneous contingencies, using earlier-learned experiences to make sense of novel stimuli (Boyd et al., 2011). This conceptual development involves motor skills, perception, emotions, and language (Hargreaves & Pexman, 2012). Several attempts to model similar human mental and somatic skills in artificial artefacts, e.g., robots, have mostly focused on low-order cognitive phenomena, such as perception, manipulation, navigation and motor coordination. Despite their significance and their approximation to a hypothesised developmental paradigm of human cognitive functions, these attempts do not yet offer a clear sight of how their blueprints can explain or scale up to high-level cognitive competence (Mirolli & Parisi, 2011). For example, a fundamental form of high-level human cognition is grouping concepts pragmatically into coherent categories, the learning and utilisation of which allows us to draw non-trivial inferences on situations where we lack direct experience (Bruner & Austin, 1986). The act of categorisation deeply reflects problem-solving in child development (Lupyan & Bergen, 2016), ranging from perceptual clustering (e.g., object colour-grouping in prelinguistic

* Corresponding author.
E-mail addresses: i.giorgi@kent.ac.uk (I. Giorgi), golosio@unica.it (B. Golosio), massimo.esposito@icar.cnr.it (M. Esposito), angelo.cangelosi@manchester.ac.uk (A. Cangelosi), g.masala@kent.ac.uk (G.L. Masala).

https://doi.org/10.1016/j.cogsys.2023.101151
Received 22 December 2022; Received in revised form 25 April 2023; Accepted 7 July 2023
Available online 8 July 2023

infants) to nontrivial abstract thinking (Markman, 1989). Moreover, past and present-day research seem to converge towards the idea that conceptual abstract thinking is devoted to an essential part of human cognition: *language* (Borghi et al., 2011). An influential author to have attested in favour of language (overt or egocentric) is Vygotsky (Vygotsky, 1962; Vygotsky & Cole, 1978) who attributes a fundamental role to language and to the social foundation (e.g., tutoring by adults) for the conceptual development of infants. In his view, the emergence of some concepts, e.g., scientific concepts, requires formal-logical deductions and problem-solving methods that are entirely linguistic in nature. Recent hybrid models of sensorimotor and verbal behaviour seem to sustain the hypothesis that abstract concepts are grounded in sensorimotor, emotional, social, and linguistic experience (Andrews et al., 2014; Louwerse, 2011; Borghi et al., 1752). Thus, a desideratum for intelligent robots is to account for language when modelling their cognition.

First, let us clarify the meaning of a concept. In this work, concepts are viewed from a purely psychological angle used to describe a particular group of things of either materialistic (both animate and inanimate) or abstract existence that share a meaningful similarity with one another (Murphy & Medin, 1985). Humans draw concepts from pure observation (e.g., *cats, dogs, small, large*) or our inner perceptions, emotions, or beliefs (e.g., *likeable, beautiful*) and then use conceptual content to decipher novel contingencies by mapping them as closely as possible to the concepts we hold (Bruner, 1985). A concept is such if it possesses at least one or some *critical* attributes to the definition of the concept, while it may also include other *non-critical* attributes that are not essential to that definition. Concepts might emerge either through encounters with the environment first and lexicalised later (Markman, 1989), or from language first and then anchored in the environment/experience (Vygotsky, 1962). This is finely illustrated in the work of Sloutsky and Deng (2019) with the following example: infants are capable of categorising dogs after repeated encounters with dogs, even before learning the linguistic term *dog*. Instead, abstract concepts like *germ* do not originate directly in experience, but from language and are later grounded in some form of personal experience. Bottom-up concept learning (from experience) occurs without language, whereas top-down learning (from language) requires a fair amount of language. Our understanding of the human brain suggests that it uses the notion of sameness or equivalence to identify patterns that allow it to treat different entities as if they were similar in some way. When we establish an equivalence (category) and a label, we can use the label to equally mark entities. Ergo, the label ultimately becomes a concept.

Certain human intellectual processes, such as planning, thinking, reasoning, problem-solving, and decision-making rely on concepts (Sloutsky & Deng, 2019). For this, we must achieve an *understanding* of simple and complex concepts. By understanding concepts, humans can follow and give instructions and organise their knowledge to solve problems adaptively. The understanding of a concept refers to the ability to assess its *critical* over its *non-critical* attributes (Donahoe & Palmer, 1994). Thus, the first step to understanding is categorisation, which helps reduce the complexity of the environment and the necessity of constant learning.

This paper concerns the modelling of high-level cognition in robots beyond the extensively-studied phenomena of perception, manipulation, navigation and motor coordination, through appropriate *acquisition* and *understanding* of concepts by sufficient involvement of *language* in the robot's cognition. Specifically, it seeks to explore the following research question (**RQ**): ***Can cognitive robots understand the concepts they use? When can we assume such an understanding?***

Here, concept understanding is regarded from the view of learning sciences research (Sawyer, 2005), which consent that understanding is demonstrated if the learner can:

C1: *Identify examples of the concept that are subject to a high variation of non-defining attributes.*

C2: *Distinguish exemplars (an example of the concept) from close non-exemplars (something that is not an example) by assessing their significant attributes.*

C3*: Maintain these abilities in novel contingencies that were not presented when learning the concept.*

When the above criteria[1] are met and understanding is achieved, the organism (human or robot) should be able to demonstrate an ability to categorise, abstract and voluntary control behaviour for adaptive problem-solving. Thus, in this work, we designed an experimental protocol of concept attainment model to teach concepts to a cognitive robot and a series of behavioural experiments based on one-occasion learning to assess the emergence of such high-cognitive skills in the robot. The concepts are attained via a tutor-learning model, assuming the context in which a child discovers a hidden rule/concept invented by an experimenter (Bruner & Austin, 1986; Vygotsky, 1962). In each of the experiments, the robot meets the three criteria of understanding by attempting to learn, categorise, abstract concepts and control the task. The goal and available resources of each task are slightly altered using socio-centric linguistic stimuli. In response to the new stimuli (linguistic and not), the robot must change its previous categorisation decisions to accomplish the goal, if possible, while self-managing the decision of if and how to do this (e.g., resources are not appropriate for the goal). Moreover, to test if the concepts are appropriately mapped to meaningful experiences and understanding of those experiences, the third experiment involves multilingual stimuli. Here, the robot is expected to reproduce a certain earlier-learned sensorimotor behaviour in response to a new language when this experience and the concepts that surround it are not directly trained in that language. Hence, any decision-making or problem-solving of the robot is guided by its reasoning on the concepts, following the idea of Sloutsky and Deng (2019) that the label becomes the concept itself and, as such, it can map to the same mental representations associated with that concept that have originated from motor experience.

Our contribution can be summarised as follows: it is one of the few attempts in the literature on cognitive robotics, which addresses high-level human-specific cognitive skills in robots by combining the situated and interactive view of cognition with a proper understanding of human language. Moreover, to the best of our knowledge, this is the *first* work to explore the understanding of concepts by a robot, where concepts are treated from a purely psychological perspective that is closer to how humans acquire and make sense of their knowledge.

2. Related works

The scientific literature on computational modelling of concept development and the interaction of language with robot cognition is currently narrow. The natural attainment of concepts from the angle of learning sciences research, e.g., the concept attainment model, has little or not been addressed by present-day computational efforts. Among works that address concept learning, some have made significant contributions to the learning and generalisation of handwritten characters at a human level from one or few learning instances (Lake et al., 2015; Lázaro-Gredilla et al., 2019). Both models demonstrated outstanding results that closely resemble or are indistinguishable from the concepts produced by humans. However, they have considered concepts as simple programs instead of notions and, as such, they only involved perceptual categories (vision/action-based), which require little involvement of language and much less abstraction. Instead, we investigate concept attainment in a robotic model from a psychological angle. In our work, this attainment arises by combining situated and interactive cognition with language, which may lead to humanlike high-cognitive phenomena like categorisation, abstraction and voluntary control. For example, we

[1] C1, C2 and C3 refer to criteria of understanding.

aimed to show that lexicalising concepts can lead to generalising earlier-learned experiences to new endeavours with little learning. Moreover, these models use probabilistic methods. On the contrary, humanlike concept learning for complex behaviour requires multiple brain systems responsible for learning, memory, attention and reward (Zeithamova et al., 2019). This motivates why, in this work, we use a brain-inspired cognitive architecture that computationally represents all of the above.

Works that have studied high-level cognitive modelling in robotics include those on the role of language in category learning (Schyns, 1991; Cangelosi & Harnad, 2001) and categorisation (Lupyan, 2005; Mirolli & Parisi, 2005). Their findings suggest that learning requires proper exposure to linguistic stimuli from other organisms in the environment to categorise novel experiences. For example, Mirolli and Parisi (2005) showed that the internal representations of pre-linguistically learned objects in their "child brain" model underwent changes when the model learned to associate them with linguistic labels. The introduction of linguistic labels resulted in the objects falling more closely to one activation pattern corresponding to their categorisation and less so to other activation patterns, thus strengthening the accuracy of categorisation. Their findings suggested that organisms with language exhibit superior behaviour, and their computational model was the first to explicitly substantiate the Vygotskyan postulate. In this work, our robot also employs language to categorise its experiences during task-solving, albeit solving a different categorisation challenge, which involves categorising earlier-learned and new concepts or experiences in response to new stimuli from the environment and interactions leading to changed decisions while solving a task.

Another computational model that draws on the Vygotskyan idea of inner speech to empower cognitive functionalities such as categorisation and abstraction is introduced by Granato et al. (2020). The model was able to reproduce human-collected data on the Wisconsin Card Sorting Test (WCST) task while using inner speech as feedback during the categorisation of the cards, which suggested that self-directed language can support executive functions in many domain-specific tasks and that the categorisation process can generate more abstract patterns recursively. To illustrate, suppose the model was to initially sort the deck of cards by colour, it should abstract from their other attributes such as number or shape. Introducing inner-speech feedback may draw attention away from colour onto another attribute, say shape, therefore generating abstract concepts progressively like "colour", "shape", and "number". This study seems to also support the notion that internalised linguistic stimulation (inner speech) can shape/alter the model's problem-solving decisions, allowing it to adapt its behaviour more favourably when relevant stimuli are introduced. This voluntary control arises as a consequence of the skill to abstract, which prompts the model to modify its actions in response to the covert linguistic stimulus ("correct" or "not correct").

There have been multiple other attempts to introduce language to robot learning. A rich array of these works can be retrieved from Tellex et al. (2020). Nonetheless, much of this research has been highly focused on elementary lower-order cognitive phenomena (perception, manipulation, navigation, motor coordination) and there is no immediate indication of how this fits in the complex human cognitive continuum. Many of these advances rely on deep learning and reinforcement learning, for which the learning processes are generally poorly understood, and, in most cases, the underlying mechanisms of language acquisition are neglected. While DL/RL methods have propelled a broad range of domains, they come with unneglectable limitations, especially when extending them to robotics (Pierson & Gashler, 2017), for example, their hunger for large corpora and being designed around specific ad-hoc problems, for which goals and rewards might change only slightly. There remains much to explore on how active perception and adaptive low-level motor behaviour relate to high-level human-specific skills for abstract reasoning and complex decision-making.

Compared to the aforementioned studies, our work differs in several respects. Similarly (Granato et al., 2020), it draws on cognition-enabled models, specifically, the multicompartment working memory (WM) principles, by exploiting one of the few architectures that instantiate the theories of Baddeley (Baddeley & Hitch, 1974) and Cowan (Cowan, 1998). As such, it maps meaningfully onto human-like cognitive processes required for flexible concept learning and language development. Our study models computationally the procedural stages of concept attainment (CAM) through socio-centric tutor-based learning anticipating the emergence of high-level cognitive phenomena in a robotic architecture, i.e., such phenomena occur during "natural" impromptu interactions between the robot and a human tutor. Moreover, no studies in the current literature address such phenomena in intricate multilingual environments to demonstrate the theoretical insights of "labels being concepts". Finally, the most pertinent contribution of this work that we are aware of, is that it is the first to investigate the understanding of concepts by an artificial model (robot). Specifically, the works surveyed here, which have demonstrated high-cognitive skills like category learning, categorisation, abstraction and voluntary control to some extent, have not shown if such skills have emerged because their models could understand the concepts being used. Instead, here we show such understanding in the way we know humans understand as suggested by learning sciences research (Sawyer, 2005).

3. Methods

This work launches on an existing cognitive architecture. This section describes in brief the original model's organisation and computational details that are needed to fathom its plausibility for high-level cognition. The specific contribution of this work to the enhancement of this framework is also described, as is the final robotic system. Additionally, the section illustrates our novel experimental protocol of the Concept Attainment Model, which underpins the concept-like behavioural experiments to follow.

3.1. PART I. The cognitive framework

The cognitive model ANNABELL (Golosio et al., 2015) used here as our framework is consistent with the well-defined theoretical multicompartment Working Memory (WM) principles (Baddeley & Hitch, 1974; Cowan, 1998). The procedural implication and analogy with the WM support a type of cognitive learning in robots, which is closer to the mechanisms of information elaboration and reasoning in the human brain, compared to other methods (e.g., deep or reinforcement learning). The model adopts neuro-inspired functions that support high-level cognitive competence and allows using human language to build action-inference relationships.

ANNABELL (*Artificial Neural Network with Adaptive Behaviour Exploited for Language Learning*) is a large-scale and computationally efficient neural network designed to learn human language through a child-like developmental approach. It operates under the connectionist belief that linguistic skills are the behavioural manifestation of the internal representations in the brain that emerge in our interaction with the environment. Rather than relying on pre-existing knowledge from large corpora, ANNABELL learns language and behaviour through the *procedural neural mechanisms* involved in task-solving.

The neural components of the model include (Fig. 1**A**): 1) a ***long-term memory (LTM),*** to store and retrieve information as semantic or sensory memory, 2) a ***short-term memory (STM)*** that includes a *phonological loop*, a *goal stack* that contributes indirectly to decision-making by holding goal chunks when actions cannot be executed immediately, and a *comparison structure* that identifies information similarity in the STM and LTM, 3) a ***central executive (CE)*** that is a controller, which oversees manipulation between the former two subcomponents and manages all statistical decision-dependent processes inside the model, and 4) a ***reward structure (RW)***.

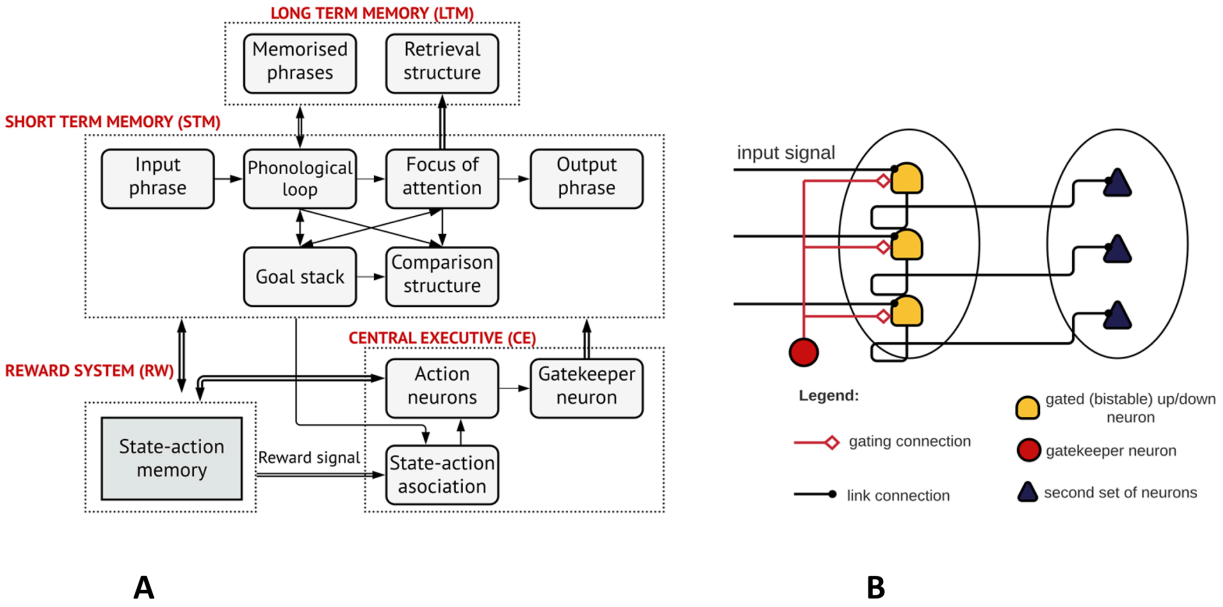


Fig. 1. (A) Schematic diagram of the ANNABELL learning framework. Each rectangle represents a subnetwork, composed of interconnected artificial neurons. Only the main high-level subnetworks are shown. The arrows that join the rectangles represent directional connections among neurons of different subnetworks and buffers. (B) The synaptic gating mechanism in the ANNABELL learning framework gates the information flow between action neurons and neuron sets of other SSMs.

3.1.1. Computational details of the original model

The ANNABEL architecture design is based on the concept of the sparse-signal map (SSM), consisting of several SSM subnetworks of interconnected artificial neurons (Golosio et al., 2015).

The neuron design prioritises computational efficiency over biological fidelity. Two types of connections are utilised: fixed-weight connections that transfer the pattern of neuron activity among SSMs through a gatekeeping mechanism, and variable-length (learnable) connections for learning and memory. The majority of the connections are learnable. The k-winner-take-all rule is combined with a discrete version of the Hebbian rule (DHL) to model inhibitory competition among neurons, which activates the *k* neurons with the highest activation state while turning off the remaining neurons. The Hebbian principle describes a theoretical learning mechanism based on synaptic plasticity, where neurons that fire together strengthen their synaptic junctions (Hebb, 2005). In ANNABELL, link weights saturate to a maximum value when the output states of the pre and postsynaptic neurons at the opposite ends of a connection are simultaneously above the threshold. While ANNABELL's DHL rule is relatively simple compared to models that focus on biological realism, it applies the same learning principle that underpins synaptic plasticity in biological neural networks. The model also applies a neural gating mechanism compatible with the synaptic gating theory in the cortex and other areas of the brain (Gisiger & Boukadoum, 2011) and is linked to the capacity of our working memory to filter out or retrieve information (McNab & Klingberg, 2008).

The flow of information across different SSMs is controlled by two types of neurons, the *action neurons,* and the *gatekeeper neurons*. Action neurons perform mental operations on sentences (*mental actions)*, producing an attentional selection of words and are connected to gatekeeper neurons. The latter neurons simulate the gating mechanism to allow or inhibit signal flow through sub-networks. Fixed neural connections ensure specific actions always activate the same state of the network (Fig. 1**B**). The gating mechanism is controlled by the state-action association (SAA) neural network that associates mental actions with the model's internal states. The SAA network receives a reward signal to permanently memorise valid associations, allowing the model to execute the same mental operations for similar inputs in the future.

3.1.2. Novel contributions to the cognitive model

This work contributes to the original model in several respects. First, the original model is purely verbal and does not include mechanisms for multimodal elaboration. To encompass it within a larger grounded robotic system, we designed a method that could accommodate stimuli, linguistic and not, without any organisational changes to the model, but only to its procedural and memory retrieval mechanisms. In the theoretical WM model, the structure responsible for binding information of several domains (auditory, visual, spatial) in chronological order is the *episodic buffer* (or *focus of attention* in Cowan's model). Here, we contribute to the model with a novel episodic buffer, which approximates it to Cowan's view in upholding that the WM involves abstract (phonological, semantic, spatial) and sensorial (visual, auditory, tactile) encoding all in one activated structure. We encode all inputs in the (robotic) model as lexicalised symbols, with the sole motivation that these are human-interpretable, which eases training and interaction with the model (Supplementary Materials). Moreover, this artifice introduces a further benefit for modelling conceptual content in the model. Encoding multiple types of stimuli as lexicalised symbols allows easier integration of manifold inputs involved in the representation of a concept. For example, the body of knowledge that represents the concept of a *dog* can be accessed by visually observing a dog, hearing a dog bark, by the label dog, touching a dog, and so on. Thus, the conceptual content of the model can be compartmentalised into various domains, with each domain representing and stimulating different perceptions (across multiple senses) or manipulations, hence radically transforming the procedural mechanisms and memory representation within the architecture. These domains also account for a diverse set of inputs that can influence the behaviour of the robot within its actual workspace.

The above contribution stretches further in this paper's original work on multilingualism in the conceptual content of the model. Here, each encoded non-linguistic sensorial stimulus is associated with two linguistic labels corresponding to a concept (e.g.,_dog dog,_dog cane - where cane is Italian for dog). The two linguistic labels trigger the same internal representation of the concept (e.g., motor representation of an action), causing a parallel rivalry between the labels. In the original model, words are represented as orthogonal vectors, thus making it unable to identify word meaning similarities. The joint representation of "dog" now brings together synonymous terms such as "dog" and "cane", without explicitly breaking the orthogonality of word vectors. This allows different labels to trigger the same mental action sequences, which relate to accumulated perceptual and motor experiences. The robotic model can activate these mental states to reproduce experiential knowledge without having to train new mental action sequences for that behaviour with semantics from a different language. This has a significant implication on the multilingual behavioural experiment with the robotic model, leading to high-cognitive skills and understanding, and emphasising how labels are themselves concepts. The technical details of this artifice are explained in Event 3.

3.1.3. *The proposed robotic architecture*

The enhanced cognitive model is coupled with the virtual PR2 robot (PR2 Robot, 2023) in a Webots simulator (Webots, 2023), which receives spoken utterances from a human tutor through a speech interface and generates robot motor behaviour in response to visual and auditory stimuli (Fig. 2). The PR2 robot received and produces multimodal data that are perfect sensory-independent samples. This is central to assessing the competence of the cognitive model alone and parting it from the end-to-end system's performance. Processing of the multimodal data (visual, motor, and linguistic) is performed inside the cognitive model. The visual data are extracted from PR2 in the Webots workspace and cover directly observable instances, such as objects, colours, shapes, etc. The motor data are primitive operations executed by the robot using elementary movements of the joints or body. We refer to *sensory data* - sensory stimuli - the visual and motor information of action primitives (***non-linguistic domain***). Encoded verbal utterances – auditory stimuli - comprise the *linguistic data* (***linguistic domain***).

3.2. *PART II. The experimental protocol*

The ***Concept Attainment Model (CAM)*** is an inductive thinking strategy stemming from the research efforts of Jerome Bruner (Bruner & Austin, 1986). It describes the process of constructing a meaningful definition of a concept by identifying those attributes that are salient to the concept (critical attributes) and disregarding those that are not (non-critical attributes). The task is learner-centred: the teacher who "knows the answer" guides the learner to attain a new concept that is initially beyond them (Wood et al., 1976). The CAM model follows these steps: 1) the teacher determines a new concept (concrete or abstract); 2) the teacher creates a list of exemplars of the concept (YES category) and another of non-exemplars (NO category); 3) the learners assess the attributes that all instances in the list of exemplars share among each other, which are not found in any of the instances of non-exemplars; 4) the learners determine the salient/critical attribute(s) that define the concept (*note:* the learners might make initial hypotheses on the concept, while the teacher may continue adding items in the lists of exemplars and non-exemplars to refine those hypotheses) 5) the teacher dictates the concept.

We illustrate the concept attainment model with the task in Fig. 3. This task is used to train our robot to produce the CAM as a "constituent" skill, pre-generalisation to the novel concepts used in our behavioural experiments.

The list of exemplars (YES category) includes three objects: a small blue cube, a small blue sphere, and a small blue cone (Fig. 3). All three attributes (*size, colour, scalar quantity*) are purposefully selected the same. When assessing which of these attributes is critical to the concept in question, the learner (our model) must determine if any of the objects in the list of non-exemplars (NO category) possesses that attribute (e.g., the square is blue and small, but no objects have volume). After having identified the critical attribute, the human interlocutor dictates the concept: a ***three-dimensional object.*** The model generates and permanently memorises the inference "*three-dimensional - volume*".

The CAM task can be used to attain new concepts following human-guided learning. The human tutor can construct categories of exemplars and non-exemplars with new objects and allow the model to autonomously make the necessary inferences on the (new) target concept.

4. Concept-like behavioural experiments

The central aim of our behavioural experiments is to *verify the three criteria of concept understanding*, by drawing on an artifice of *concept attainment*, i.e., using the discriminable features of an instance to anticipate its significant identity (Bruner & Austin, 1986) and *language-facilitated tutorial interaction* (Wood et al., 1976). Conceptual understanding is best achieved if concepts are situated in task-oriented events within a context. The context imposes specific constraints, which allows demonstrating the skills to categorise, abstract and voluntary control behaviour (Bruner & Austin, 1986).

We explored three events, each modelling a different task. The robotic model learns to manipulate the tools to reach the task goal using a ***snowballing*** artifice. The tutor guides the robot to build "higher-order" skills by orchestrating an appropriate combination of its "lower-order" skills to meet novel and more complex task requirements (in line with human theories (Bruner, 1973 Bruner, 1973). The model starts with a small but sufficient "lexicon" of basic skills (e.g., action primitives) that it combines preferentially in a certain order to achieve a particular end, by matching means with the expected outcomes. This is thoroughly explained in the Supplementary Materials submitted with this work.

The task in event 1 is used to emerge the intended concepts using the experimental protocol of the concept attainment model and a task-driven categorisation. The tasks in events 2 and 3, respectively, "scaffold" the task in event 1 to achieve higher level concept categorisation and abstraction in response to language (event 2) or multiple languages (event 3).

Event 1. Bottom-up concept attainment and categorisation

We modelled the task of *pouring liquid into a cup from a utensil*. This event imposes two constraints: 1) the utensil is a container, and 2) the colour of the requested liquid matches the anticipated identity of the

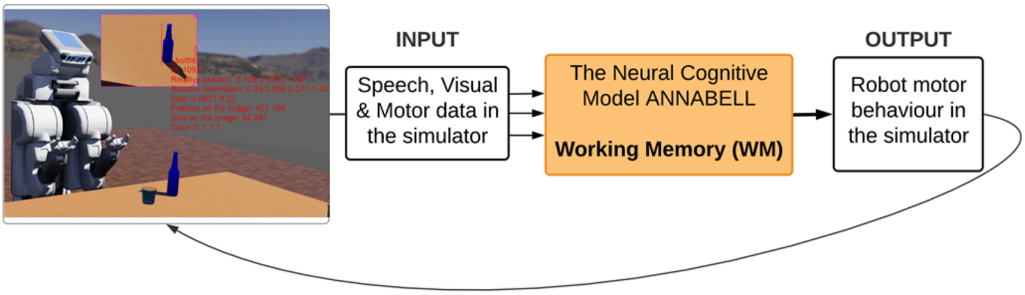

INPUT
Speech, Visual & Motor data in the simulator
The Neural Cognitive Model ANNABELL
Working Memory (WM)
OUTPUT
Robot motor behaviour in the simulator

Fig. 2. Interfacing ANNABELL with the virtual PR2 robot in Webots. The connection between the model and the robot is two-directional. The data extracted in the simulator are used as input to the model, which is responsible for decision-making during our concept-learning experiments. The output produced by the cognitive model from the stimuli received by the virtual PR2 robot is fed back to the simulator, to drive the robot's motor behaviour and complete the task(s).

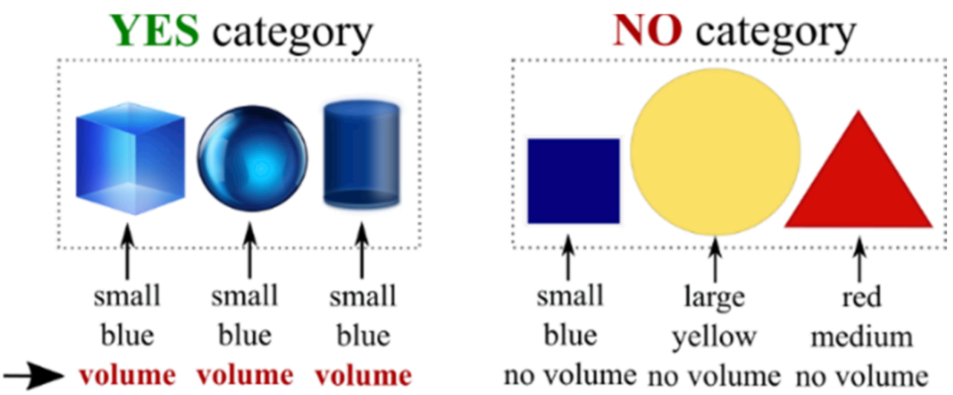


Fig. 3. Modelling the constituent skill of concept attainment model (CAM) task using geometrical shapes.

liquid. The robot is trained on one occasion to add water to the cup from a bottle. The aim of the event is to identify that, other utensils can generate the same response as the bottle, given their critical attribute to hold liquid, i.e., are containers (constraint 1). This constraint can only be met if the model has successfully attained the concept of [container] and can distinguish between exemplars and non-exemplars. Categorisation allows the model to use exemplars coequally, without learning the task anew on each (i.e., generalisation). This ability is humanlike: we can use available tools interchangeably, which possess similar functionalities or attributes that allow solving a problem, even with little learning or practice (Bruner & Austin, 1986). To attain the concept of [container], we modelled the following steps of the experimental verification.

Step 1: Concept not attained

To evaluate the model's behaviour before attaining the concept, we elicited it with the instruction "*take the container*". We trained the model once with the bottle, which was implicitly defined as a container in long-term memory ("*the bottle is a container*"). When processing the instruction, the model recalls this memory to execute the task. However, the model has no implicit knowledge, either dictated or self-learned, of any other object that can be considered a container. Hence, it cannot

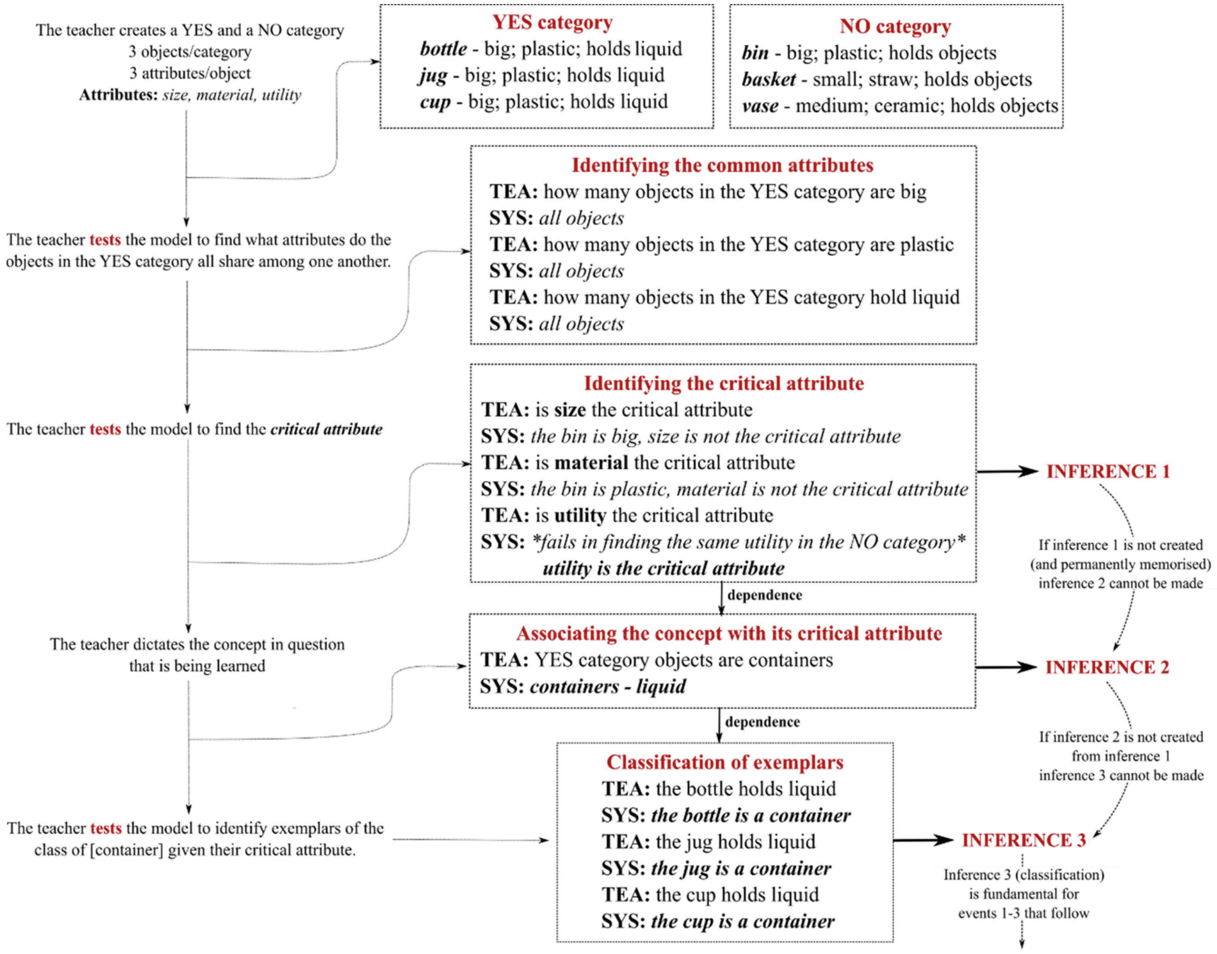

The teacher creates a YES and a NO category
3 objects/category
3 attributes/object
Attributes: size, material, utility
YES category
bottle - big; plastic; holds liquid
jug - big; plastic; holds liquid
cup - big; plastic; holds liquid
NO category
bin - big; plastic; holds objects
basket - small; straw; holds objects
vase - medium; ceramic; holds objects
The teacher tests the model to find what attributes do the objects in the YES category all share among one another.
Identifying the common attributes
TEA: how many objects in the YES category are big
SYS: all objects
TEA: how many objects in the YES category are plastic
SYS: all objects
TEA: how many objects in the YES category hold liquid
SYS: all objects
The teacher tests the model to find the critical attribute
Identifying the critical attribute
TEA: is size the critical attribute
SYS: the bin is big, size is not the critical attribute
TEA: is material the critical attribute
SYS: the bin is plastic, material is not the critical attribute
TEA: is utility the critical attribute
SYS: *fails in finding the same utility in the NO category*
utility is the critical attribute
INFERENCE 1
If inference 1 is not created
(and permanently memorised)
inference 2 cannot be made
dependence
The teacher dictates the concept in question
that is being learned
Associating the concept with its critical attribute
TEA: YES category objects are containers
SYS: containers - liquid
INFERENCE 2
If inference 2 is not created
from inference 1
inference 3 cannot be made
dependence
The teacher tests the model to identify exemplars of the class of [container] given their critical attribute.
Classification of exemplars
TEA: the bottle holds liquid
SYS: the bottle is a container
TEA: the jug holds liquid
SYS: the jug is a container
TEA: the cup holds liquid
SYS: the cup is a container
INFERENCE 3
Inference 3 (classification)
is fundamental for
events 1-3 that follow

Fig. 4. Generalisation of the concept attainment task (CAM) to attain the concept of [container] used in our experiments. Inferences 1, 2 and 3 that are made by the model itself are co-dependant and fundamental for the classification of exemplars and non-exemplars of the class of containers. This task is tutored-guided at runtime (i.e., in the test).

distinguish between examples or non-examples of the container class, leading to task failure (see Results).

Step 2: Attaining the concept (CAM)

We ran the CAM task at runtime to attain the concept of [container]. Two categories were created:

1. The exemplars (YES) category: *big plastic bottle, big plastic jug, big plastic cup*
2. The non-exemplars (NO) category: *big plastic bin, medium ceramic vase, small straw basket*

where the selected attributes are ***size*** (small, medium, big), ***material*** (plastic, ceramic, straw) and ***utility*** (holds liquid, holds objects).

The generalisation of the CAM (test) is shown in Fig. 4. The self-inferences 1–3 made by the model are fundamental to the continuation of CAM and for our next behavioural experiments. Should inference 1 fail, inference 2 is not obtained. Similarly, inference 2 constrains inference 3. The latter marks the attainment of the concept, which allows the model to classify items into exemplars or non-exemplars of the class. In making inference 1, the model self-determines the critical attribute that leads to the concept in question, whereas inference 2 associates the concept with its defining critical attribute.

Step 3: Discriminating exemplars against non-exemplars

Should the CAM succeed, the model will self-generate and permanently store in long-term memory the following phrases: "*the jug is a container*", "*the cup is a container*" (inference 3). Then, the model is tested again on the task of taking the container, without training the task from scratch.

The classification skill of the model is assessed as its ability to use the self-generated inferences impromptu to perform the task correctly using examples that belong to the class and cease its execution with those instances that do not belong to the class (see Results).

Step 4: Significance of the [concept – attribute] association

Fig. 4 illustrates that inference 3 is reliant on inference 2. When the model self-determines that the concept of [container] is closely linked to its ability to hold liquid (inference 2), it can autonomously classify a utensil as a container granted that it possesses that critical attribute (inference 3).

Given this logical construction, if one were to remove the fundamental inference 2, thus reducing the significance of the critical attribute associated with the concept, the attainment process will fail (inference 3 not obtained), which hinders the classification of exemplars and non-exemplars.

As constraint 2, we selected colour to discriminate liquids and exemplars with clear-cut boundaries (i.e., of one and a distinct colour) to reduce the cognitive load of classification. Differently from the concept of [container], the colour concept is assumed to pre-exist in long-term memory, e.g., *the water is transparent*, which can be hypothesised as semantic memory.

Given the slight limitations of the virtual simulator, the colour of the liquid matches that of the container. We do not consider other parameters e.g., whether the utensil is full or empty. The robot's vision module extracts the colour label of the utensil, and the cognitive model decides if it consistent with the verbal information known about the liquid stored in semantic LTM.

Event 2. Language in the attainment of abstract concepts

This event aims to explore how linguistically-based social learning can shape the robot's categorisation, abstraction, and voluntary control decisions to perform a new more complex task. Specifically, it assesses how language leads to the creation of new categories encompassing lower-level categories learned previously, i.e., in event 1.

The nature of the task assumes a tutorial process (Wood et al., 1976): let us consider that a child has learned to pour different liquids by colour using a range of containers (event 1). The adult seeks to teach the child to perform a new task that is initially beyond the child's skill, e.g., *preparing tea*. The constraint of the task is to use water. Thus, the child must form the new category of water container that now includes only certain (but not all) instances learned in event 1 as examples (those with water). Ultimately, two new sub-categories originate from the original category of [container]: a) water containers as exemplars and b) non-water containers as non-exemplars. When solving the new task, the child must therefore change the initial categorisation decisions, by abstracting from the differences and similarities of the larger category of containers to fit the current constraint, and decide whether to execute the task. The adult can guide the child's learning verbally in either way:

1. Explicitly telling the child that water is needed for tea.
2. Follow the concept attainment model (CAM) to teach the child to associate the concept of tea with the critical attribute of water.

In either case, whether by learning the procedures that emerge the skill or by relying on more direct social cues such as language, there is some form of tutoring required to acquire that skill (Wood et al., 1976; Vygotsky & Cole, 1978). In event 2, we thus explore both methods so as to determine if language alone (method 1) can elicit the same response to the robot's decision-making as the self-assessment of the concept through CAM (method 2). Direct intervention by the tutor, e.g., demonstrating the task, is not considered in this experiment.

A. Language-directed (method 1)

The model was trained on one occasion to make tea using a water bottle and given the factual affirmation "*to make tea you need water*" as a permanent semantic memory. When tasked to make tea, the model is trained to recall this information from memory and to verify if the colour of the referent (*water*) that it observes in the workspace is transparent. Hence, the verbal stimulus retrieved from memory is sufficient to allow the robot to abstract from the colour differences of the instances and, thus classify instances pragmatically into categories of examples and non-examples of the concept of [water container]. The new categorisation decision of the model must allow it to voluntarily control the task, by choosing to execute or cease the task based on the inferences it makes on the available concepts. This ability is humanlike: we produce a considerable number of categories in response to language over those produced in the absence of language and much of this categorisation is achieved through abstraction promoted by language (Mirolli & Parisi, 2006; Mirolli & Parisi, 2011). This also anticipates how verbal dictation during runtime interaction can continuously and progressively change the robot's decisions in solving different tasks, impromptu and with little learning. The model's generalisation skill is demonstrated by repeating the task consistently with all combinations of instances to assess their classification.

B. Self-assessment (method 2)

The model was trained to identify the significance of [water] attribute in attaining the abstract concept of making [tea] by generalising the CAM task. The instances were divided into a YES and a NO category as in Fig. 5. Only two attributes were considered: ***type*** (container) and ***utility*** (holds [liquid]). As in Fig. 4, the model is probed to self-generate the inferences that lead to the intended concept, with the aim of associating the concept with its critical attribute, i.e., inference 2, which is held in memory. This is the equivalent of the verbal cue "*to make tea you need water*", but instead of being built in semantic memory, it is created by the model itself from experience. The model's ability to classify is tested during task-solving in event 2, where the robotic model must decide preferentially to execute or cease the task case by case.

Note that, the model does not attain the abstract concept of [tea] *during* the task of making tea. Instead, it must attain the concept and its strong association with the critical attribute [water] *before* the task, and the high-cognitive skills involved in decision making are validated while executing the task.

In event 2, the model builds upon the earlier-learned experience of manipulating containers to fill a cup (event 1) as a natural learning continuum. Earlier-learned experiences are memorised in memory and recalled as an acquired skill to solve a larger problem, without retraining the motor experience. This continuum ensures that the outcomes of

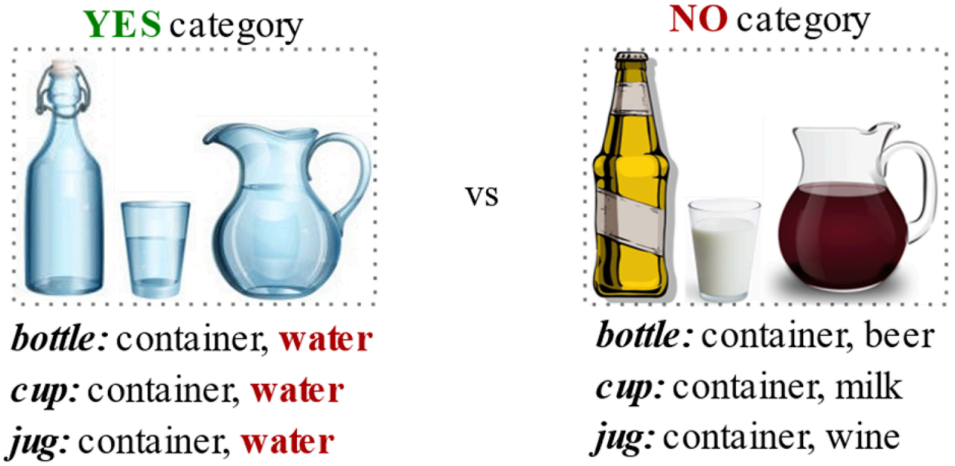


Fig. 5. The CAM task to obtain the abstract concept of [tea] and its association with the critical attribute of [water].

earlier tasks are directly applied to future and more complex events, thus "scaffolding" the robot's motor and cognitive skills.

Event 3. Multiple languages in the manipulation of internal conceptual representations

This event aims to delve deeper into whether the high-cognitive skills of the robot root in its conceptual foundation. In particular, by far, the motor-perceptual experiences of the robot and language semantics were trained together. Here, we aimed to investigate what would happen if those experiences were to be triggered using labels of a different language. Specifically, by disentangling semantics from the experience trained jointly, we expected to determine if labels themselves can carry enough conceptual content for problem-solving, i.e., if labels can act as concepts themselves. The ability for multilingual cognition is humanlike and devoted to our conceptual development (Boyd et al., 2011): we are able to understand concepts expressed in distinct verbal expressions (Cook, 1992); for example, *I like milk* and *mi piace il latte* (Italian) are distinct utterances, whose respective predicates are different formulations of the same concept. The central idea of this event is not concept attainment *per se* as in events 1 and 2; rather after the model has appropriately developed the concepts in event 1, to identify those concepts each time it encounters them in a new event and then relate them to its earlier-learned experiences to adapt its decisions. For this, the robot should first intuit that labels of distinct languages refer to the same concept, hence to the same representations mapped to that concept, which are required for problem-solving. This would reinforce the conclusion that the robot is understanding, based primarily on criterion C3.

We modelled the task of a robot waiter *serving a beverage of preference to a person.* The task "scaffolds" that in event 1, with the constraint that, here, the liquid exemplar is only the one that matches the preferred beverage of the referent actor. The remaining liquids are non-exemplars. Hence, as in event 2, two new sub-categories will be created from the larger category of containers: a) those that refer to a preferred beverage for each person, and b) those that do not. Hence, the robot is challenged with a new categorisation decision: to abstract from the liquids based on if their identity matches a person's preferences. To solve this task, the robot must identify the goal and constraints, and then take the necessary steps to perform or cease the task. The robot is verbally informed what each "actor's" preference is, which coheres in a realistic social context when, for example, ordering a drink in a bar. To introduce multilingual stimuli, these verbal affirmations are dictated in *Italian.* The following assumptions are made: a) the robot has acquired the conceptual experience of pouring liquids from different containers in an English-guided workspace (event 1), and b) the robot has a basic lexicon of Italian words and their mapping to non-lexical representations, but no such mapping to any experience with the concepts in the Italian-guided workspace. The task can be viewed as a multidomain multi-knowledge source task (Fig. 6).

The semantic sources are independent of one another, i.e., no translation between English and Italian semantics is given. Hence, the robot should find a way to determine when these semantics are equivalent (i.e., refer to the same concept) and why. For this, we utilised the non-linguistic representation of the relevant concepts as a bridge (pivot) between their distinct linguistic representations (labels) (Fig. 7).

When a label is attached to the concept, it becomes part of the conceptual representation itself (Vygotsky, 1962; Sloutsky & Deng, 2019). Children may learn category lexicalisations (linguistic representation) at the same time as the perceptual category (non-linguistic representation). This occurs when the child is verbally dictated the label of a referent explicitly while showing or pointing at the referent. Bilingual infants learn two lexicalisations simultaneously for the same perceptual category, even before they formally learn the languages (Cook, 1992; Crinion et al., 2006). Hence, our proposed solution not only fits the child development theories, it also offers a new method for modelling multilingual cognition, without translation. It allows the robot to utilise different language sources by mapping their respective labels to the same perceptual categories and related experiences they refer to, i.e., making labels part of the concept.

We trained the model to solve the task when introducing a new stimulus in Italian. Additionally, we trained the model to solve the same task, but when asked to do so in Italian. Each training was performed only once per language. Training to solve the task in Italian with no previous learning in Italian not only enables to transfer and adapt the robot's earlier-learned experiences to new contexts, but also promotes faster learning of the semantics of the new language, directly while solving a task (learning-by-doing). From a computational angle, this also challenges the model's capabilities on a high hardware resource demanding task allowing assessing its robustness.

5. Results

The results report the ability of our robotic architecture to meet the criteria of concept understanding (C1–C3) in task-driven events 1–3. The model's generalisation competence is *tested by consistency* (Bruner & Austin, 1986), i.e., habitually finding the concept that solves the task, using 3-fold cross-validation (CV) in each event and displaying high-level cognitive skills. In each round of the CV, one of the containers was randomly selected to train the task on one occasion and the remaining two were used to test. The datasets are illustrated in Tables 1–3 (*events 1–3,* respectively) in the Supplementary Materials. They represent only the round of the cross-validation that assumes the bottle as the learning sample.

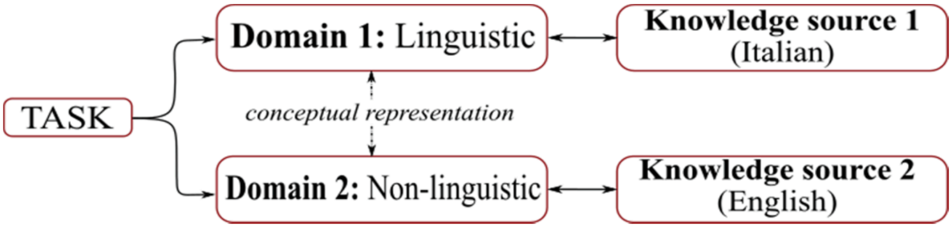

TASK
Domain 1: Linguistic
Domain 2: Non-linguistic
conceptual representation
Knowledge source 1 (Italian)
Knowledge source 2 (English)

Fig. 6. Multidomain multiple-sources task-solving. The robotic model with a certain non-linguistic experience (domain 2) accumulated in an English-guided environment (knowledge course 2) receives language-directed instructions (domain 1) in the Italian language (knowledge source 2). No direct link between the sources indicates no cross-lingual translation. The necessary cross-domain relations are appropriately orchestrated by the internal representations of concepts.

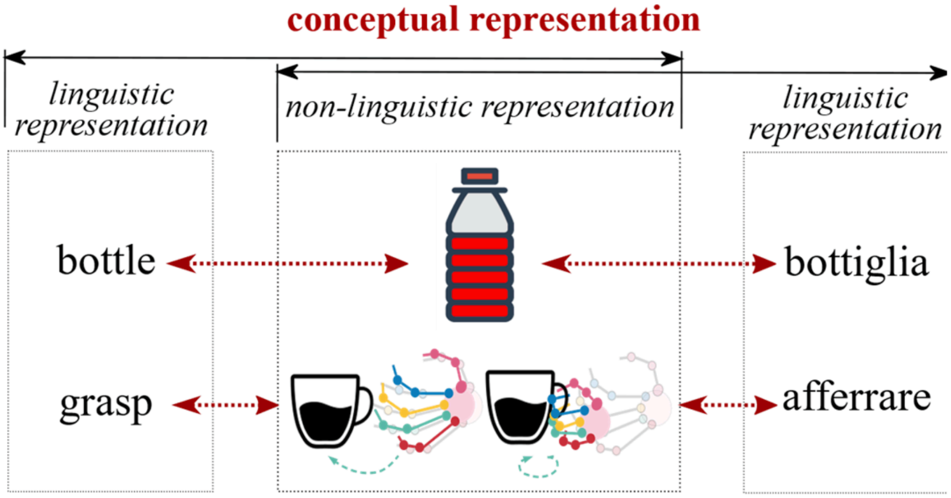

conceptual representation
linguistic representation
non-linguistic representation
linguistic representation
bottle
bottiglia
grasp
afferrare

Fig. 7. The (simplified) representation of a concept includes linguistic and non-linguistic parts. The non-linguistic representation for concrete concepts is somewhat fixed, while the linguistic representation can vary between multiple languages. Hence, the former can be used as common ground to map (pivot) the multilinguistic terms associated with the concept. The linguistic labels can directly trigger the non-verbal representation of the concept or the experience that surrounds it and, indirectly, the label of the other language. Thus, language can support the manipulation of internal representations.

Event 1. Bottom-up concept attainment and categorisation

In event 1, the robot was consistently tested using exemplar (A, D) and non-exemplar (B, E) objects not introduced when learning the task (Fig. 8). We probed the model on a total of 74 test samples, among which 14 cases met the event constraints (groups A and D) and 60 cases that did not (group E). Instead, non-exemplars of group B were tested using the concept attainment model (CAM) (Fig. 4).

Criterion 1: *Identify examples of the concept that are subject to a variation of non-defining attributes.*

In the CAM task, the model demonstrated the ability to identify the critical attribute associated with the concept (*ability to hold liquid*) against the non-defining attributes of the concept of container (*size*, *colour*). Thus, it could successfully build the self-inferences 1–2 with **100 %** accuracy (Fig. 4). Identifying the critical attribute was fundamental to classifying exemplars from non-exemplars of the concept of [container] (see criterion 2).

Criterion 2: *Distinguish exemplars (an example of the concept) from close non-exemplars (something that is not an example) by assessing their significant attributes.*

Through self-inferences 1 and 2 (CAM), the model could successfully build inference 3 (Fig. 4) to classify exemplars of containers (**100 %** accuracy) and performed ***100 %*** in the task "*take the container*".

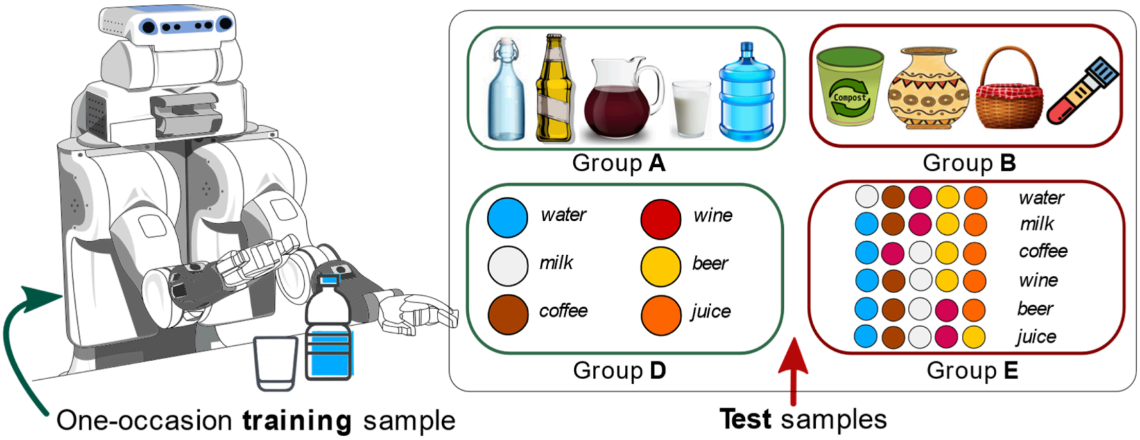


Fig. 8. Event 1: Bottom-up concept attainment and categorisation. The objects in group A belong to the class of containers, whereas objects in group B do not. The model was tasked with classifying containers (A) and non-containers (B) following the steps of CAM. To solve the task in event 1, the model must also separate liquids based on their discriminable colour. An example of this concept refers to the queried liquid matching its observed colour identity in the scene (group D); otherwise, it refers to non-examples (group E).

The robotic model distinguished the examples of containers from non-examples with close attributes (*size, colour, ability to hold objects*), even for novel instances that were not presented in the YES/NO categories. For example, when the model was given a new object (mug) and was verbally told that "*the mug holds liquid*", the following output was produced:

TEA: the mug holds liquid

SYS: the mug is a container

The model's classification of containers is shown in Movie 0 of the Supplementary Materials.

Criterion 3: *Maintain these abilities in contingencies not presented when learning the concept.*

The model verified criterion 3 in consistently reproducing the task "*add liquid to the cup*" with examples of containers not introduced when learning the task. For instance, the model had never learned to use a jug when the task was trained owever, after being able to classify the jug as a container (criterion 2), the model was able to use the jug autonomously to reproduce the task. This ability was maintained for the other candidate members of the class of [container] *as if they were the same thing* (Bruner & Austin, 1986). Moreover, the model could discriminate the liquids consistently by their colour attribute. Notice that the model was trained on one occasion to verify the colour of the water. The total performance of the robotic model in event 1 (add liquid to the cup) is illustrated in Fig. 11**A**. The results of **94.14 %** showed that the model was able to categorise instances successfully, generate similar behaviour with all other liquids and preferentially control task-solving: proceed with the task when all constraints were met and cease the task otherwise (i.e., voluntary control). These results are also supported by Movie 1 of the Supplementary Materials.

In conclusion, with a generalisation capability of "1 vs 74", the robotic model displayed the ability to learn categories, categorise instances, and voluntarily control the task, showing a degree of understanding of the task by meeting the suggested criteria (C1–C3).

Event 2. Language in the attainment of abstract concepts

In each round of the cross-validation, we used a total of 34 test samples to measure the model's competence in preparing tea with any other water container (group A, Fig. 9) and cease the task for non-exemplars of the class (group B, Fig. 9).

The same protocol illustrated in Fig. 9 was used when the concept was self-assessed by the model (CAM) – 17 cases or the constraint was dictated by the human (language-directed) – 17 cases. First, we report the results obtained using the self-assesment method (CAM) and, next, we compare the results with the language-directed method.

A. Self-assessment (method 2)

In the CAM task, the model was probed to find the critical attribute in the YES category that objects in the NO category did not have (Fig. 5), leading to inference 1:

TEA: is utility the critical attribute

SYS: *does not find the attribute of holding water in the NO category*

utility is the critical attribute

Inference 1 supports the competence of the model in meeting criterion 1 of concept understanding (critical vs. non-critical attributes). Next, the model was led to self-build inference 2, as follows:

TEA: YES category objects make tea

SYS: tea – water

The model generated inferences 1 and 2 with **100 %** accuracy. These were fundamental for the task of event 2. The task-solving performance when the model self-assessed the concept of tea in association to its critical attribute of water was **98.04 %** in the cross-validation as illustrated in Fig. 11**B**.

B. Language-directed (method 1)

The model was verbally instructed by the human on the significance of the water attribute to the concept of [tea]. The performance in response to the verbal stimulus is illustrated in Fig. 11**B**. It can be seen that there is no significant difference in task-solving compared to the self-attainment of the concept (**96.08 %** for language-directed). Using only verbal cues, the robotic model could preferentially control the task by suitably categorising exemplars of water containers from non-exemplars of the class (Fig. 11**B**). This supports criterion 2. The ability of the model to generalise on unlearned instances supports criterion 3, i.e., the robot reproduced the task with instances of water containers not introduced during learning. Event 2 is demonstrated in Movie 2 of the Supplementary Materials.

In conclusion, with a generalisation capability of "2 vs 34", the robotic model demonstrated that it could generate *higher-order categories* from a previous categorisation, in response to a linguistic cue, which caused some instances to fall closer together in a category than other instances previously grouped with them. This required *abstraction*: the model had to intuit that within the larger category of containers, instances of water containers share a greater equivalence among one another than non-water containers, in a task of making tea. As such, the robot was able to voluntarily control the task, self-deciding when to execute it and when not (the latter not trained specifically – the model is trained only once to successfully solve the task). When displaying such competencies, the robot was consistently meeting the suggested criteria of understanding the task.

Event 3. Multiple languages in the manipulation of internal conceptual representations

In the test, the robot was tasked with serving a drink to four people

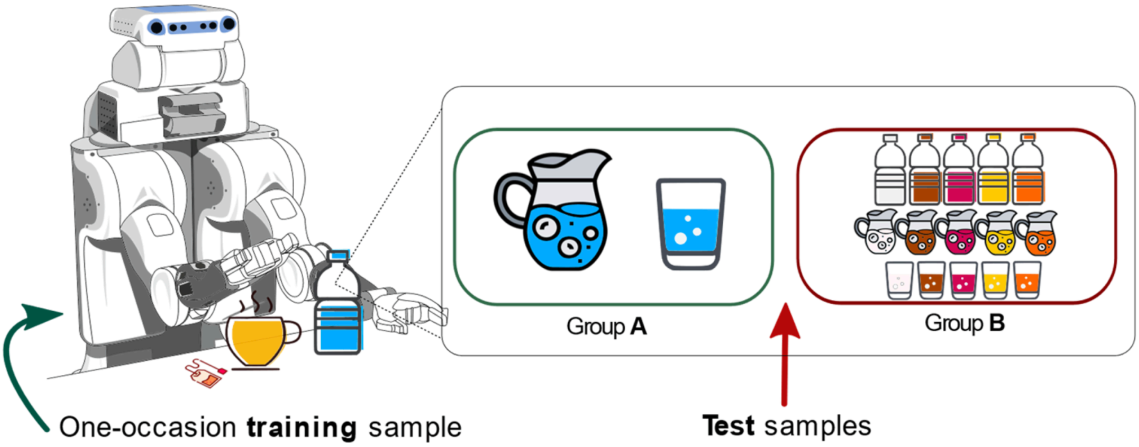


Fig. 9. Event 2: Language for the attainment of abstract concepts. The model is trained to make tea using a bottle of water. Learning to use the involved tools (cup, teabag) is achieved via our novel snowballing artifice (Supplementary Materials). Group A includes instances that can be used to prepare tea, whereas group B involves instances that cannot.

(including Mary) using exemplars of the preferred drink (group A) and non-exemplars of the class (group B) (Fig. 10).

We investigated a total of 47 cases for each language, with 11 cases meeting the constraints and 36 cases that did not. The instruction was repeated in each language. The results are reported in Fig. 11**C** and illustrate the performance comparison between languages. The model could solve the task satisfactorily with a total average of **93.97 %** in both languages (94.33 % in English, 93.62 % in Italian), requiring that it categorised instances into examples and non-examples of the class of preferred drink in response to the language-directed cues. This competence verifies criterion 2. The model changed its initial categorisation decisions in response to language, and this ability was maintained for all members of the class not introduced during learning (criterion 3). Most importantly, the criteria of concept understanding were maintained in response to multiple languages, demonstrating a capability to map language to inner conceptual representations to favourably manipulate those representations during task-solving (see also Movie 3 submitted with the Supplementary Materials).

In conclusion, with a generalisation capability of "2 vs 94" in the multilingual task, the robot displayed the skill to identify concepts when they were expressed in different lexicalisations, and map its internal representations of experience respectively. Not only were labels used to shape categorisation, abstraction, and voluntary control of the robot, but also to produce adaptable conceptual behaviour in contingencies where this behaviour was not directly trained, thus reinforcing criterion C3.

The results obtained in the cross-validation in each event are summarised in Fig. 11**.** Instead, Fig. 12 reports the theoretical performance of the robotic model. While the results in Fig. 11 illustrate the measured performance of the model during the cross-validation, these results assumed that event 1 is achieved at 100 % performance (only accounting the successfully solved cases, which produced experiences that were recalled in later events). Events 2 and 3 are, however, reliant on the outcomes of event 1. Given that the model error is random and averaged across the three rounds of the cross-validation, the theoretical performance accounts for the effect of the (average) accuracy of event 1 in event 2 and event 3, respectively.

6. Discussion

At present, many attempts have been done to model low-level skills in robotics, with the literature being much narrower in the direction of complex cognition. Human learning theories and contemporary views in cognitive robotics research suggest that *language* could explain and help model this humanlike cognitive continuum. Driven by this motivation, this work aims to delve deeper into using language for the emergence of high-level cognitive phenomena in robots. Given that our capacity for intelligent thinking is heavily linked to our ability to form concepts, this work sustains that the skills of a robot to learn, categorise, abstract and control should start from teaching it the right concepts. This paper presents a method for achieving concept attainment within a robotic cognitive architecture that adheres to the principles of the human working memory. The WM-analogous mechanisms of the model support its ability to attain concepts, which is believed to be the first step in the human cognition process (concepts first emerge in the working memory before being transferred to long-term memory (Cowan, 2014).

The main aim of the work was to determine if the model was displaying high-cognitive skills during task solving because it could understand the concepts used in each task. Understanding was based on three well-defined criteria in learning sciences research. Our theory-driven analysis revealed that these criteria were met above a minimum of 93.97 % in experiments drawn upon theories of humanlike behaviour. An important highlight of our results is that this competence was achieved by training the model on *one occasion* only, in each event, and generalising consistently to ample instances of the test set. The robot could create more categories in response to language cues than those learned initially. This result is similar to that obtained with the "child brain" model of Mirolli and Parisi (2005), which explicitly supports the Vygotskyan postulate. In our work, the model is not only concerned with the categorisation of words, but of their relation to experiences for task-solving. An important aspect of our method, is that the earlier-experiences related to the attained concepts could be directly retrieved to solve more complex problems in a hierarchical continuum without retraining those experiences in each event, which may have a real impact in the training of robots.

We noticed that in some cases the error of the model during the 3-fold cross-validation would depend on the learning sample, i.e., some samples resulted more difficult than others. For example, in event 2, using *cup* to train the task produced slightly better accuracy than the other two containers. This was because the model had learned two ways to manipulate the cup for this task: to place the (tea) cup on the table and to use the (container) cup to add water, which may have affected the ability of the central executive (CE) of the cognitive model to resolve the

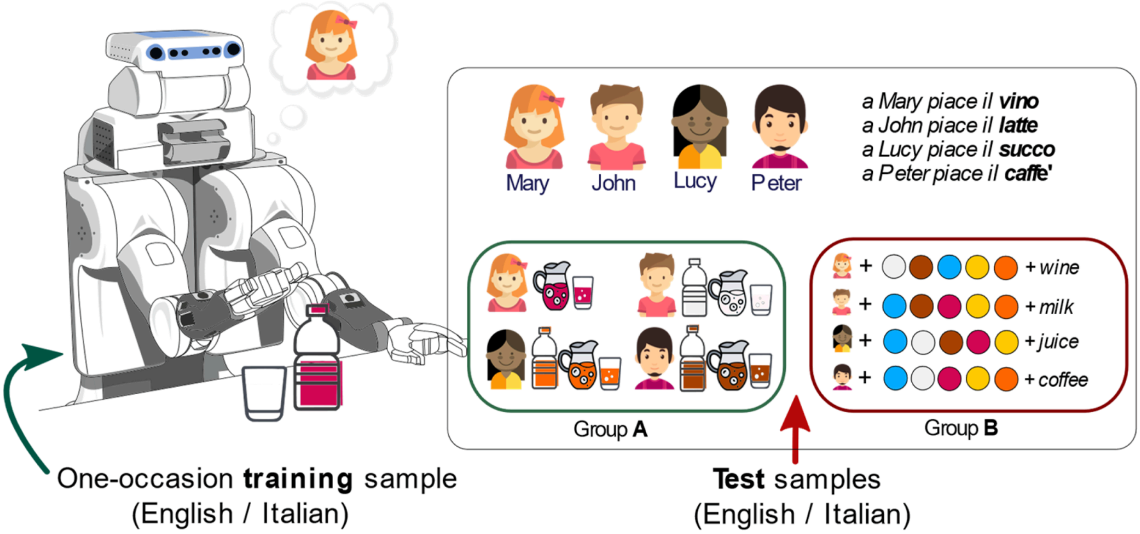


Fig. 10. Event 3: Multiple languages in the manipulation of internal conceptual representations. The model is tasked with serving a preferred drink to a person, dictated in simple Italian phrases. Group A includes examples of the concept of preferred drink, in which the event-imposed constraints are met, whereas group B includes cases in which the observed drink does not match the anticipated identity of the actor's preferred beverage. The task is repeated two times: (a) model queried in English and (b) model queried in Italian.

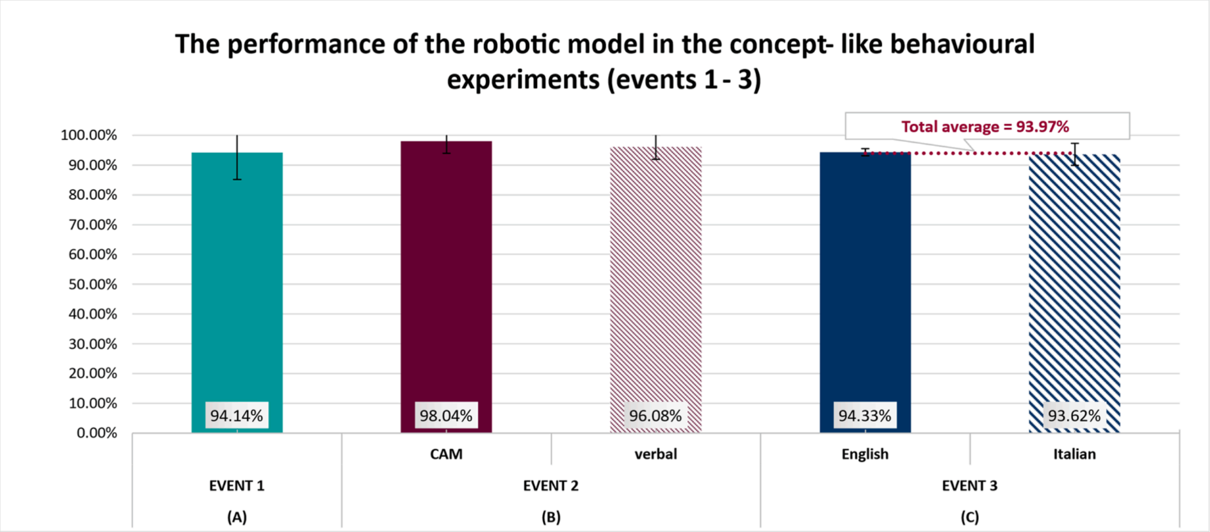
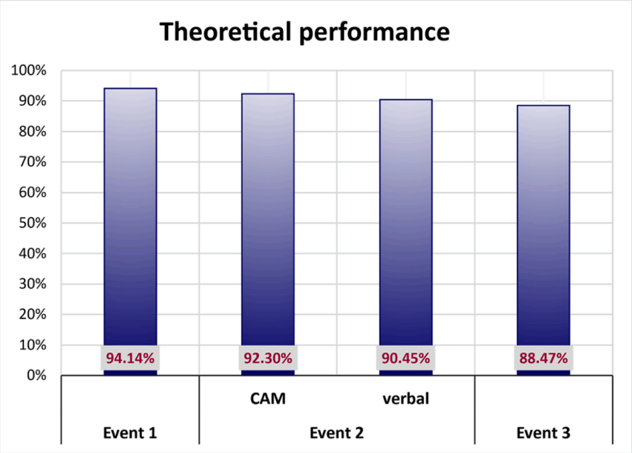


Fig. 11. The performance of the model measured through 3-fold cross-validation in event 1 (A), event 2 (B), and event 3 (C), respectively.



	Theoretical performance
Event 1	4.14%
Effect of event 1 in event 2 (CAM)	94.14% *x* 98.04% = 92.30%
Effect of event 1 in event 2 (verbal)	94.14% *x* 96.08% = 90.45%
Effect of event 1 in event 3 (total)	94.14% *x* 93.97% = 88.47%

Fig. 12. The theoretical performance (i.e., accuracy) of the robotic model, which considers the effect of the outcomes of event 1 in the proceeding events 2 and 3 as calculated on the right.

competition. When trained specifically on the *cup*¸this competition was resolved by learning the task; however, when generalising (i.e., other containers as the learning sample), there was a slight confusion between the two cup instances that were produced in ouput. However, our observations concluded that in all the events (including event 2), the error performed during the cross-validation was random. In turn, we observed that when a failed instruction was repeated a second time, the model could resolve it successfully. Nevertheless, we did not include repeated instructions in the success rate of the model and only considered the first attempt to solve the task in each round.

We designed our robot experiments as a tutoring problem given two rationales. First, our cognitive model assumes the stages of child development (4+ years old); hence, the type of learning performed here approximates that of Bruner's scaffolding theory (Bruner, 1985) that considers infants somewhat dependent on the knowledge and competencies of their tutors. Second, we attribute a key role to socio-centric language in the development of higher more complex skills in robots. The results obtained in event 2 releaved that language-directed cues to guide robot learning in novel problems are no different from the self-assessment of the task (Fig. 11**B**). We believe this may significantly impact robot learning directly from humans in a rather natural way to develop task competencies that are initially beyond the robot's skill and outstrip its unassisted efforts. For example, we performed a simple illustrative case in a similar task, e.g., making hot chocolate, by verbally telling to the robotic model at **runtime** that "*to make hot chocolate you need milk*". Runtime, here, means that the initial trained state of the model was not changed, and the task is not trained *de novo.* This showed that the robot unaware of the end goal of the task could achieve a new outcome, where the elements of the task were "controlled" preferentially by the human through language alone (see Movie 1).

Finally, the verification of concept understanding across multiple languages revealed two important outcomes for robot learning, which highlight the significance of a well-developed conceptual content:

1. the robot could comprehend a verbal cue formulated in a different language because linguistic labels become part of the conceptual representation. Hence, a robot with suitably developed conceptual content can produce fine outcomes by analysing the (unlearned) events in terms of their surrounding concepts, having language drawing attention to those concepts. For example, the robot solved the task using only cues from the Italian language (Fig. 11**C**, English).
2. Conceptual representations enable concept-related experiences to be readily accessed in new contingencies. For example, when the robot had earlier-learned motor experience around the concept, its linguistic label triggered the inner representation of that concept and the mental states of the motor experience. Thus, the label of a

different language allowed the activated motor experience to be directly recalled without learning it explicitly in that language. This is demonstrated given the comparable performances in English and Italian, for which (the latter) the robot lacked direct experience with the attained concepts (Fig. 11**C**, Italian).

7. Conclusions

This work aimed to address the emergence of high-level human-specific cognitive skills in robotics, such as the ability to learn, categorise, abstract and voluntary control. Achieving this requires building a cognitive robotic model with sufficient conceptual content through its physical and social interaction with the environment. This concerns not only the senses and motor abilities of the robot but also its use and understanding of human language, which is essential for social and cognitive development. To initially attain enough concepts, we modelled computationally the well-designed strategy of the Concept Attainment Model and designed three one-occasion learning experiments where those concepts would be used in related situated experiences. We examined how the initial decisions of the robot changed when it was exposed to social (i.e., linguistic) stimuli from a human tutor to solve different and more complex tasks. We demonstrated how the robot was using language to generate new categories by abstracting the concepts' attributes and self-mapping its inner representations of motor experiences to human's instructions, as such eventually leading in the creation of high-level concepts, e.g., of tea making or preferred beverage. Notably, the robot was able to recognise concepts by their labels and retrieve related experiences, and it identified equivalent lexicalisations expressed in multiple languages. In essence, the robot could map a word in a new language with the concept it represented and produce a behaviour with the concept which was not trained directly in that language. Hence, the robot could adapt to intricate contingencies with minimal training through its established conceptual content. The main result of this work is that we demonstrate that the robot *understands* the concepts it uses, based on three well-defined criteria from the learning sciences research, which is similar to how humans demonstrate understanding. In summary, this study demonstrated that the conceptual content of a robot is crucial for it to emerge high-cognitive skills, including the skill to understand, and language is a fundamental component of this conceptual development.

Future work: Understanding concepts and language in rich contexts is a long-term goal. In the near sight, however, our results seem to validate the promising potential of robotic models in emerging high-level cognitive processes such as categorisation, abstraction, and voluntary control. For example, the foregoing example of making hot chocolate given a simple verbal cue must be explored exhaustively to verify if robots can indeed abstract in workspaces directly from human language to change their decisions or produce novel goals. Supported by the ability to learn these skills at runtime, i.e., without re-training, such models can significantly develop and adapt in real environments. Consider our results of classification of novel instances of containers: when the model was shown a mug and told that "*the mug holds liquid*", it could readily classify the mug and use it directly in the task as the other members of the class, without learning anew.

Moreover, conceptual understanding in multilingual context requires significant attention in future works, for instance if inner conceptual representations offer a finer and more natural way to achieve cross-lingual retrieval than simple translation. A profound investigation could reveal if this might lead to a reduced necessity to re-train robots in each language. In turn, it might overwhelm the problem of some languages being more disadvantaged in learning resources. The skill to transfer earlier-learned experiences as well as to exploit every piece of information available, including eclectic linguistic scenes that are sources rich in semantics and interactions, can be a powerful tool to advance robotics.

CRediT authorship contribution statement

Ioanna Giorgi: Writing – review & editing, Writing – original draft, Investigation, Visualization, Conceptualization, Methodology, Validation. **Bruno Golosio:** Writing – review & editing, Resources, Conceptualization, Methodology, Validation. **Massimo Esposito:** Writing – review & editing, Conceptualization, Methodology, Validation. **Angelo Cangelosi:** Writing – review & editing, Supervision, Resources, Conceptualization, Methodology, Validation. **Giovanni Luca Masala:** Writing – review & editing, Supervision, Resources, Conceptualization, Methodology, Validation, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cogsys.2023.101151.

References

Andrews, M., Frank, S., & Vigliocco, G. (2014). Reconciling embodied and distributional accounts of meaning in language. *Topics in Cognitive Science, 6*, 359–370.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47–89). Academic Press.

Borghi, A. M., Barca, L., Binkofski, F., & Tummolini, L. (1752). Varieties of abstract concepts: Development, use and representation in the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*, Article 20170121.

Borghi, A. M., Flumini, A., Cimatti, F., Marocco, D., & Scorolli, C. (2011). Manipulating objects and telling words: A study on concrete and abstract words acquisition. *Frontiers in Psychology, 2*, 15.

Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences, 108* (Suppl. 2), 10918–10925.

Bruner, J. S. (1973). Organization of early skilled action. *Child Development, 44*(1), 1–11.

Bruner, J. S. (1973). *Beyond the information given: Studies in the psychology of knowing*. W. W. Norton.

Bruner, J. (1985). Child's talk: Learning to use language. *Child Language Teaching and Therapy, 1*(1), 111–114.

Bruner, J. S., & Austin, G. A. (1986). *A study of thinking*. Transaction Publishers.

Cangelosi, A., & Harnad, S. (2001). The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories. *Evolution of communication, 4*(1), 117–142.

Cook, V. J. (1992). Evidence for multicompetence. *Language Learning, 42*(4), 557–591.

Cowan, N. (1998). *Attention and memory: An integrated framework*. Oxford University Press.

Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review, 26*(2), 197–223.

Crinion, J., Turner, R., Grogan, A., Hanakawa, T., Noppeney, U., Devlin, J. T., … Usui, K. (2006). Language control in the bilingual brain. *Science, 312*(5779), 1537–1540.

Donahoe, J. W., & Palmer, D. C. (1994). *Learning and complex behavior*. Allyn & Bacon.

Gisiger, T., & Boukadoum, M. (2011). Mechanisms gating the flow of information in the cortex: What they might look like and what their uses may be. *Frontiers in Computational Neuroscience, 5*, 1.

Golosio, B., Cangelosi, A., Gamotina, O., & Masala, G. L. (2015). A cognitive neural architecture able to learn and communicate through natural language. *PLoS ONE, 10* (11), Article e0140866.

Granato, G., Borghi, A. M., & Baldassarre, G. (2020). A computational model of language functions in flexible goal-directed behaviour. *Scientific Reports, 10*(1), 1–13.

Hargreaves, I. S., & Pexman, P. M. (2012). Does richness lose its luster? Effects of extensive practice on semantic richness in visual word recognition. *Frontier in Human Neuroscience*, Article 234.

Hebb, D. O. (2005). *The organization of behavior: A neuropsychological theory*. Psychology Press.

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science, 350*(6266), 1332–1338.

Lázaro-Gredilla, M., Lin, D., Guntupalli, J. S., & George, D. (2019). Beyond imitation: Zero-shot task transfer on robots by learning concepts as cognitive programs. *Science Robotics, 4*(26).

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science, 3*(2), 273–302.

Lupyan, G., & Bergen, B. (2016). How language programs the mind. *Topics in Cognitive Science, 8*(2), 408–424.

Lupyan, G. (2005). Carving nature at its joints and carving joints into nature: How labels augment category representations. In *Modeling language, cognition and action* (pp. 87–96).

Machery, E. (2009). *Doing without concepts*. Oxford University Press.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. MIT Press.

McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience, 11*(1), 103–107.

Mirolli, M., & Parisi, D. (2005). Language as an aid to categorization: A neural network model of early language acquisition. In *Modeling language, cognition and action* (pp. 97–106).

Mirolli, M., & Parisi, D. (2006). Talking to oneself as a selective pressure for the emergence of language. In *The evolution of language* (pp. 214–221).

Mirolli, M., & Parisi, D. (2011). Towards a Vygotskyan cognitive robotics: The role of language as a cognitive tool. *New Ideas in Psychology, 29*(3), 298–311.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*(3), Article 289.

Pierson, H. A., & Gashler, M. S. (2017). Deep learning in robotics: A review of recent research. *Advanced Robotics, 31*(16), 821–835.

PR2 Robot - Wevolver, Willow Garage [online], available at https://www.wevolver.com/wevolver.staff/pr2, last accessed April 2023.

Sawyer, R. K. (Ed.). (2005). *The Cambridge handbook of the learning sciences*. Cambridge University Press.

Schyns, P. G. (1991). A modular neural network model of concept acquisition. *Cognitive Science, 15*(4), 461–508.

Sloutsky, V. M., & Deng, W. (2019). Categories, concepts, and conceptual development. *Language, Cognition and Neuroscience, 34*(10), 1284–1297.

Tellex, S., Gopalan, N., Kress-Gazit, H., & Matuszek, C. (2020). Robots that use language. *Annual Review of Control, Robotics, and Autonomous Systems, 3*, 25–55.

Vygotsky, L. S. (1962). *Thought and language*. MIT Press.

Vygotsky, L. S. (1962). An experimental study of concept formation. In L. Vygotsky, E. Hanfmann, & G. Vakar (Eds.), *Thought and language* (pp. 52–81). MIT Press.

Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.

Webots: Open-Source Robot Simulator – Cyberbotics [online], available at https://cyberbotics.com/, last accessed April 2023.

Wood, D., Bruner, J. S., & Ross, G. (1976). *The role of tutoring in problem solving*. Child Psychology & Psychiatry & Allied Disciplines.

Zeithamova, D., Mack, M. L., Braunlich, K., Davis, T., Seger, C. A., Van Kesteren, M. T., & Wutz, A. (2019). Brain mechanisms of concept learning. *Journal of Neuroscience, 39* (42), 8259–8266.