Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research 44 (2017) 50–68

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Re-framing the characteristics of concepts and their relation to learning and cognition in artificial agents

Juan Sebastian Olier [a,b,*], Emilia Barakova [a], Carlo Regazzoni [b], Matthias Rauterberg [a]

[a] Department of Industrial Design, Eindhoven University of Technology, Eindhoven, The Netherlands
[b] Department of Electrical, Electronic, Telecommunications Engineering and Naval Architecture, University of Genoa, Genoa, Italy

Received 10 February 2017; accepted 20 March 2017
Available online 30 March 2017

Abstract

In this work, the problems of knowledge acquisition and information processing are explored in relation to the definitions of concepts and conceptual processing, and their implications for artificial agents.

The discussion focuses on views of cognition as a dynamic property in which the world is actively represented in grounded mental states which only have meaning in the action context. Reasoning is understood as an emerging property consequence of actions-environment couplings achieved through experience, and concepts as situated and dynamic phenomena enabling behaviours.

Re-framing the characteristics of concepts is considered crucial to overcoming settled beliefs and reinterpreting new understandings in artificial systems.

The first part presents a review of concepts from cognitive sciences. Support is found for views on grounded and embodied cognition, describing concepts as dynamic, flexible, context-dependent, and distributedly coded.

That is argued to contrast with many technical implementations assuming concepts as categories, whilst explains limitations when grounding amodal symbols, or in unifying learning, perception and reasoning.

The characteristics of concepts are linked to methods of active inference, self-organization, and deep learning to address challenges posed and to reinterpret emerging techniques.

In a second part, an architecture based on deep generative models is presented to illustrate arguments elaborated. It is evaluated in a navigation task, showing that sufficient representations are created regarding situated behaviours with no semantics imposed on data. Moreover, adequate behaviours are achieved through a dynamic integration of perception and action in a single representational domain and process.

Keywords: Concepts; Conceptual representations; Cognition; Artificial intelligence; Robotics; Machine Learning

AbbreviationsAI, Artificial Intelligence; CS, Cognitive Sciences; DL, Deep Learning; EC, Embodied Cognition; FC, Fully Connected Layer; GC, Grounded Cognition; GM, Generative Model; RNN, Recurrent Neural Network; SGP, Symbol Grounding Problem; VA, Variational Auto-encoder; VRNN, Variational Recurrent Neural Network

* Corresponding author at: Department of Industrial Design, Eindhoven University of Technology, Eindhoven, The Netherlands.
E-mail address: J.S.Olier.Jauregui@tue.nl (J.S. Olier).

http://dx.doi.org/10.1016/j.cogsys.2017.03.005
1389-0417/

1. Introduction

To interpret the world and appropriately act on it, an agent needs to solve a dynamic and situated problem that implies structuring and representing the environment. In particular, that requires specifications about how representations relate to the sensory data, and how they should be processed. To that end, it is necessary to define mechanisms of knowledge acquisition and information processing.

In this work, those mechanisms are directly related to concepts and the connection between their definition and the understanding of intelligence, learning, reasoning and action generation in general. Such relationships arise from the fact that concepts are generally seen as fundamental structures for representing and processing information.

Information processing is intrinsic to the agent and its capabilities to interact with the environment. Thus, all knowledge acquisition should be based on that, and not on external sources or definitions. However, many approaches to design intelligent artificial agents assume specific and, under the described notions, arbitrary semantics or symbolic sets for learning and processing representations. A driver of that is, for examples, viewing reasoning in terms of symbolic manipulations and as an isolated process in between independent perception and action capabilities, which is a common view in the most traditional version of the so-called action-perception cycle.

It could be argued that such ideologies about reasoning and intelligence arise from computation and the metaphorical evolution of the idea of brains functioning as a computational model. Nonetheless, the development of neurosciences and brain studies shows a shift in the understanding of cognition that draws the attention to the need for changing views on these topics.

In particular, fundamental changes have emerged under the postulates of grounded and embodied cognition. Such ideas follow empirical evidence suggesting definitions of conceptual representations as dynamic, flexible and context dependent. That challenges more traditional views that may consider concepts as related to the classification of elements in the environment from constant and context-neutral features.

Embodied cognition (EC) research is particularly relevant to the problem of concepts as it challenges the idea of behaviour as a consequence of internal algorithms performed on specific symbolic representations. Instead, EC sees cognition as an active coupling between body and environment, centred on action (Wilson & Golonka, 2013). As exposed by Anderson (2003), EC postulates go against the Cartesian claims distinguishing mind and body, and beyond the manipulation of abstract representations. Rather, cognition is understood as laying on the interaction with the environment. In that sense, cognition uses the world as its own model and creates structures to simplify cognitive tasks, where representations are grounded in the sensorimotor system and oriented towards the agent's needs and actions.

In that sense, to properly behave in a given environment an agent needs to represent its surrounding in accordance to performed tasks. Then, representations will have meaning for a particular task, and so should be constructed by the agent through interaction and observation, and not from external impositions. That is so since from a raw stream of sensory inputs one could not assume that any particular semantics exists as intrinsic to the data itself. Any segmentation of the data could be possible depending on the context, and the agent's constraints and goals. That, unless a given ontology is imposed on the sensors, which not necessarily is sufficient nor optimal. Thus, the only relevant representations are those built by the agent through experience, and to achieve useful and adequate behaviours.

Based on that, in the present work, it will be argued that regardless of the mentioned shifts and contributions in different branches of the cognitive sciences (CS), a persistence of former but settled views on concepts and cognition can be evidenced in the attitude towards intelligence. That applies in particular to many approaches to artificial intelligence (AI) and robotics.

Nonetheless, the adoption of different views into actual technical applications in such fields is not trivial, nor always seen as necessary. Particularly, new notions of concepts are still to be translated into theories and practice in intelligent systems beyond current efforts. To that end, some framing work might be essential, which is a primary goal of the present contribution. Moreover, it looks for providing a framework from which the idea of intelligence and the evolution of emerging methods in AI can be reinterpreted and connected to other also evolving theories in different fields.

Thus, the proposed discussion on concepts can be seen as an opportunity to step forward in the development of artificial intelligence and the understanding of cognition, especially when seen from the contribution of evidence-based studies in different branches of the CS.

To relate the more abstract ideas here exposed to more concrete applications, as well as to illustrate main points, an architecture based on deep generative models is proposed. Particularly, it is shown that with the proposed method there is no need for explicit symbolic representations in the definition of learning or processing. Instead, behaviour can emerge from a dynamic process that actively interprets the world with action as a fundamental and necessary part of the perceptual process. That is tested in a navigation scenario, where the behaviour emerges dynamically and not from predefined processes as obstacle detection. The architecture is based on ideas about active inference, generative models and predictive coding.

The framework here proposed tries to summarise main points and highlight the differences in the views of conceptualisation by suggesting a shift in perspective to current attempts and proposals of learning. Moreover, the technical implementation is used to exemplify a possible path based on the ideas of dynamic and enacted processing. Under that framework, concepts can be interpreted as

internalised relations between action and perception, which are expressed actively as accurate predictions about the world. In that way, a concept is not seen as an isolated category but as a phenomenon that has to emerge in a given situation and concerning action.

The present work first, in Section 2, reviews theoretical and evidence-based works about the definition of concepts and its evolution. Then, in Section 3, based on the mentioned review, a set of delimiting characteristics of concepts and conceptual processing to be met by artificial systems is proposed. Then, such ideas are used to explore and analyse recent attempts of building learning systems, in particular, those more explicitly related to conceptualisation and embodiment in Section 4. From there, in Section 5, an architecture is proposed to illustrate some main points based on deep dynamic generative models. Finally, discussion and conclusions are presented in the last Section 6.

2. Concepts

Concepts can implicitly or explicitly be linked to the processing of information, and thus are a fundamental part of the understanding of intelligence. Here it is argued that currently, the discussion on concepts seems to be evolving along with the transition from more computational perspectives on cognition to embodied and grounded ones. On one side, the most computational views can be linked to symbolic and static representations, whereas in more dynamic perspectives, distributed activations and flexible structures appear to be more accepted as a way of conceptual representations.

A definition of concepts is relative to the perspective and discipline from which they are studied. One widespread understanding has associated them with the meaning of objects, events or abstract ideas. In particular, in some lines of thought language and meaning are seen as the core for concepts, which are viewed as systematically linked to words, and being stored in networks of symbol-like representations (Levelt, Roelofs, & Meyer, 1999; Quillan, 1966). Such notions imply ideas focused on modularity and symbolic representations, which in fact have significantly influenced diverse subfields of AI. In particular, such prevailing views link intelligence and reasoning to isolated processes that manipulate crisp symbols imposed on perception by the problem's definition. Regardless of their inherent limitations, such approaches are widespread, settled, and useful in very delimited applications.

Those ideas are implicitly related to the rise of a particular discussion known as the symbol grounding problem (SGP). Roughly, the SGP inquiries on how the meaning of symbols can be anchored to the perceptual inputs, mainly with no initial symbolic structure, meaning or semantics. That is a difficult problem broadly explored in robotics and intelligent systems as described by Coradeschi, Loutfi, and Wrede (2013). To some authors, it is still an open issue (Müller, 2015), though to others it is not relevant if one is interested in intelligent behaviour under constrained scenarios (Cubek, Ertel, & Palm, 2015). On the contrary, other authors state that the SGP cannot be solved; particularly Fields (Fields, 2014) shows an equivalence of the SGP definition to quantum system identification problems known to be unsolvable.

Independently of the SGP, concepts have repeatedly been defined in relation to categorisation processes as forming parts of information bodies in long-term memory. To this idea, there is a group of proposals very similar amongst them, in which categories are practically equated to concepts. These ideas can be organised into three main groups: The rule-based theories (Bruner & Austin, 1986), where a concept is described within the boundaries of a set of rules for given characteristics. The Prototype based views (Rosch, 1973), which state that a sample belongs to a class depending on its comparison to stored prototypes calculated from previously seen instances. Similarly, in the exemplar-based views (Nosofsky, Kruschke, & McKinley, 1992), membership to a category is determined by the similarity to stored examples of a category previously seen.

Those ideas, have evolved throughout time but nonetheless remain substantially stable in their principles. They are sometimes considered as fundamental (Gabora, Rosch, & Aerts, 2008; Machery, 2010), or stated as the main mechanism of cognition (Harnad, 2005). Particularly in technical implementations the ideas of prototypes are prevalent, where objects or actions are classified based on prespecified features. For example, in Kalkan, Dag, Yürüten, Borghi, and Şahin (2014) the prototypes forming a concept are defined as relations between objects, actions and effects. Such idea tries to address the embodiment and grounding on concepts, but depends on a classification process for which an external labelling is needed.

Another approach on concepts aiming at a third way between symbolic and connectionist perspectives, are the conceptual spaces presented by Gärdenfors (2014, 2004). He proposes them as a basis for concepts and meaning, building them from geometrically constructed ranges in different features spaces of specific modalities. This approach presents a more dynamic view, though the idea resembles some principles of prototype accounts translated into geometrical spaces, which in turn relates to the definition of categories.

In general, the symbolic manipulation and categorisation ideas can be linked to each other since categories are seen as stable symbols to which meaning could be attached. As such, those ideas have been extensively adopted in AI and robotics, but, as mentioned by Machery (2010), they fail to explain all the known phenomena in the different branches of CS. In consequence, the need for crisp symbols as representatives of categories is currently a challenged idea. In particular, Barsalou (2008) states that such views should be reevaluated or replaced under a framework focused on simulation, situated action, and bodily states.

2.1. Understanding concepts form behavioural and neuroimaging studies

Kiefer and Pulvermüller present a relevant work regarding empirical evidence about concepts (Kiefer & Pulvermüller, 2012). They make a systematic review of behavioural and neuroimaging studies focusing on the definition of concepts from the structural and neural basis of conceptual memory. For their analysis they frame the discussion under four dimensions:

> (i) amodal versus modality-specific, (ii) localist versus distributed, (iii) innate versus experience-dependent, and (iv) stable versus flexible.

These dimensions are briefly summarised as follows: In amodal theories, concepts are detached from sensorimotor systems, whereas in modality-specific approaches concepts are indeed grounded in such systems. In the second dimension, in localist approaches concept are coded by a single representational unit, whilst in distributed theories each concept is coded by multiple units usually activated in parallel. For the third dimension, the innate approaches assume concepts formation as based on a priori specialisations in the brain, whilst experience-dependent approaches assume that concepts are formed through experience. Finally, flexibility refers to concepts responding dynamically to contextual constraints, whilst stable representations consider concepts as situational-invariant entities.

Kiefer and Pulvermüller (2012) conclude that concepts might be flexible, experience-dependent, modality-specific, and distributed across sensory and motor systems. In particular, for them flexibility implies a relevant factor in adapting to diverse situations. They argue that action information contributes to conceptual processing depending on learning experiences since sensory and motor interactions shape conceptual memories. Accordingly, concepts should be understood as flexible mental entities whose features are activated depending on situational constraints.

Along those lines, Gur (2015) argues in favour of flexible and distributed representations by focusing on the visual cortex. He argues against a hierarchical and convergence in expert cells since, to him, such approach ignores not only the perceptual reality but also parallelism as a brain's chief characteristic from which efficiency and relevant abilities arise.

Moreover, Fernandino et al. (2015) defend a parallel representation of concepts encoded in activation patterns of different modalities. They suggest that all modality-specific attributes contribute to the representation of concrete concepts with relative relevance, whilst some sensorimotor reenactment occurs depending on the demands of a given task. Similarly, Ghio, Vaghi, Perani, and Tettamanti (2016) evidence that specific representational contents may be associated with different categories with consistent responses sparsely distributed across the cortex.

In turn, Desai, Herter, Riccardi, Rorden, and Fridriksson (2015) show evidence for a close and causal relationship between sensorimotor and conceptual systems. Their results suggest that action concepts are dynamically grounded through motor simulations that become more detailed for more explicitly semantic tasks.

An important idea also accounting for distributed and dynamic representations are the notions of functional networks in the brain. Such ideas are central to current neuroimaging studies and aim at mapping functionalities to network structures instead of just isolated functional regions (Pessoa, 2014). The assumption is that processing tasks are carried out by diverse regions simultaneously. Then functionalities emerge dynamically from the activities of the whole network rather than by isolated and sequential computations.

2.2. Embodied and grounded cognition in relation to concepts

The shift from computational to embodiment and grounding views has had important implications for the understanding of concepts and intelligence. The goal of this section is to describe recent works regarding the definition of embodied and grounded cognition and the impacts of such approaches.

Wilson and Golonka (2013) review the definition and consequences of EC. They highlight the differences between EC and behaviourist views, which are a psychological explanation of behaviours as reflections of internal algorithms produced on demand. On the other hand, EC replaces internal control and explicit representations of behaviour or knowledge with carefully built bodies perceptually coupled to particular environments. So, they argue that in general abstract conceptual representations of objects are not needed since can be replaced by the dynamics of the actions related to them under a specific situation.

Situated information is in fact an idea that can relate to the definition of context exposed by Dourish (2004), who describes it as an emerging and dynamic property dependent on interactions. Such definition could be extended to concepts and their flexibility since, as in the case of context, concepts are enacted during an interaction.

Regarding phenomena usually associated with symbolic processes as is the case of language, Lebois, Wilson-Mendenhall, and Barsalou (2015) review evidence of words not having conceptual cores, stating that context-dependent activation of semantic information is the norm, and that even main word features are task dependent. That is also supported by van Dam, van Dijk, Bekkering, and Rueschemeyer (2012), who state that representations are flexible, characterised by relative perceptual activations and dependent on contextual constraints.

Such ideas are connected to the role of actions presented by Engel, Maye, Kurthen, and König (2013), who argue that cognitive processes should be studied primarily with respect to their role in action generation as a capacity for creating structure. That is, a cognitive agent is to be immersed in his/her task domain so that the system states acquire meaning by their role in the context of action.

In fact, Engel et al. (2013) suggest that attention and perceptual decision making may be described as biases in sensory processing imposed by action contexts. In that sense, activity patterns cannot be taken as encoding action-invariant descriptions of objects and scenes, but as supporting the capability of structuring contexts. Equally, regarding object concepts, Engel et al. stated that.

> knowing what an object is does not mean to possess internal descriptions of this object, but to master sets of sensorimotor skills and possible actions that can be chosen to explore or utilise it.

All of this together implies that there may not be such a thing as a context-neutral description of object features, as would be claimed by more traditional views on concepts.

In a comprehensive study, Gentsch, Weber, Synofzik, Vosgerau, and Schütz-Bosbach (2016) provide a systematic comparison of current models and prospective theories on embodied and grounded cognition. They created three groups of classification to help in understanding how diverse approaches have been laid down in literature.

In the first one, common coding accounts, separated representational codes are assumed as mediators between action and perception, that is, both are linked through a common representational code. Then, acting and perceiving interchange information but are functionally separated. These approaches thus are not in line with the general postulates of EC.

Internal model theories take the grounding relation one step further by assuming that predictions interpret perceived actions. The predictions are grounded in motor commands and related to context and prior knowledge.

In the final group, simulation theory accounts, the emphasis is on motor control mechanisms understood as wholly constitutive for some perceptual processes, where simulation is seen as a reenactment of sensorimotor modalities.

Based on that classification Gentsch et al. (2016) propose two minimum criteria for an EC theory. The first one on acquisition and constitution, stating that motion is to be correlated with the acquisition of concepts but not necessarily to their later usage. Equally, a criterion on partial constitution says that grounded action cognition should be understood and specified regarding a partial constitutive relations between perception and action. That is, a specific ability is to be partially constituted by others, thus not entirely dependent nor equivalent to the presence of any of its forming parts.

2.2.1. Simulation and grounding

Here the notion of simulation of sensorimotor information is further developed to highlight its relevance for conceptual acquisition and processing.

Barsalou elaborates on the role of simulation (Barsalou, 2008) arguing that EC is somehow limited by bodily states. He argues that even though they are necessary for learning, cognition often proceeds independently of the body. He refers to simulation as the reenactment of sensorimotor modalities. That notion is considered as the fundamental part of knowledge representation and conceptual processing in what is known as grounded cognition (GC).

From the GC views (Barsalou, 2008) concepts are situated, meaning that the situation is imagined or simulated during conceptual processing. Thus, building a concept implies representing situational and contextual information as a fundamental part of it. In other words, a conceptual representation cannot be context-agnostic by default. Wilson had highlighted this in relation to EC before (Wilson, 2002). She stresses the relevance of replacing notions about cognition previously seen as abstract, with theories that explain evidence showing off-line EC as a widespread phenomenon in the brain.

A systemic approach linkable to simulation is the Convergence-Divergence Zones proposed by Damasio (1989). That model presents a view of how information might be integrated in the brain, which can relate to the formation of concepts, and for which recent studies (Meyer & Damasio, 2009; Man, Kaplan, Damasio, & Damasio, 2013) have attempted to find empirical evidence. The convergence zones consist of units that bind modality-specific information in a distributed and hierarchical way, building more complex representations in higher levels. The divergence refers in turn to activations reflected down to more modality-specific areas. This yields to the capability of reenacting previous experiences during processing in a simulation-like way.

Along the lines of simulation, metaphors have been seen as ways to build different levels of representational abstractions. Gallese and Lakoff (2005) argue about the embodiment of conceptual knowledge through a mapping within the sensorimotor system, and under the principle that imagination and action use a shared neural substrate. In their view, understanding is a context-dependent imagination phenomenon. They focus particularly on language, stating that if it is the expression of concepts, then it uses the same brain structures used in perception and action. In other words, language exploits the pre-existing multimodal character of the sensorimotor system, and therefore, a disembodied and symbolic account of concepts would imply an implausible duplicate of premotor circuits in different parts of the brain.

Such notions may equally imply that concepts are built out of cognitive primitives that are themselves sensorimotor in nature, something also stated by Wilson (2002). Lakoff (2012) takes that further describing abstract concepts as formed by more primitive and grounded ones. He describes two basic forming kinds for such concepts: cognitive primitives and primary conceptual metaphors. The first are used in the semantics of natural language and are described as cognitive structures which shape visual perception, motor action, and mental images. The second kind, metaphors, work by mapping across conceptual domains and are embodied either via cognitive

primitives or primary metaphors that ground more complex ones in embodied experiences.

Those ideas suggest that language is grounded, and all processing is done through metaphors and simulation in the sensorimotor systems. Thus, a view of language as essential in reasoning might be partial given that it only resides on top of simulations as a form of expression. Moreover, it is implicit that all conceptual processing and reasoning not necessarily are based on language-like symbols and rules. Consequently, systems based on the idea of reasoning as expressible through language might be trying to mimic easily measurable consequences of an underlying processing. That idea relates to why some accounts of the SGP emerge, and why approaches focused on language-like symbols would not be addressing the subjacent mechanism of cognitive processing itself, but just its expressible consequences.

2.2.2. Active inference

The notions of basing cognition on simulation can be related to a line of thought elaborated by Friston (2010), Friston, Daunizeau, Kilner, and Kiebel (2010), and Friston, Mattout, and Kilner (2011) in which it is argued that the main function of the brain is to minimise free energy or suppress prediction error. In particular, Friston suggests that motor control has a role in reducing uncertainty and can be understood as fulfilling prior proprioceptive expectations.

These ideas, known as the free energy principle or active inference, are particularly inspired by the predictive coding views (Rao & Ballard, 1999), in which top-down information carries predictions about lower levels, whilst feedforward connections carry residual errors. A major reason why biological systems may process prediction error is that, as it is minimised, the energy and processing needed also decreases.

Active inference tries to explain how agents achieve stability by restricting themselves to a limited number of internal states. Such stability is to be achieved by minimising a free energy function of internal states representing beliefs about hidden causes in the environment they interact with.

Under these views, both the internal states and the prediction signals are considered as probabilistic; thus the prediction signal can be seen as a probability distribution over possible states, and the prediction error as calculated in relation to such distribution. Equally, the internal states represent beliefs over probable causes of the perceived data. In principle, a cognitive agent tries to maintain the number of states as minimum as possible as well as the variations amongst them.

It can be argued that in many scenarios, approaches based on predictive activation as a way of simulation respond to most, if not all, of the characteristics of concepts described so far. In architectures based on such notions representations emerge dynamically from an active interaction of all the elements in a network. The process is not static but depends on the evolution of the situation and what is perceived. Thus, the forming elements of conceptual representations are flexible, context-dependent and ultimately related to actions. Equally, all sensory information is intrinsically grounded by an ongoing integration of different modalities based on simulations or predictions.

In general, these ideas are of great relevance given that they unify empirical results with a plausible neural mechanism. Equally, they account for a way to go beyond the idea of concepts as based on a categorisation process given the dynamic nature of the approach.

Under this view, the whole process of acting in relation to perception is merged into a single system. That is, concepts are related to the acquisition of action-perception relations as part of a single process from which the representations of the world emerge as the result of an active interaction with it. Such capabilities are achieved by actively minimising uncertainty. In consequence, the whole process is dynamic and adapts to the situation, whilst staying completely grounded on the sensorimotor system.

3. Delimiting the definition of the concepts and their nature

Crucial remarks about concepts can be made based on the above, which in turn are central to discuss knowledge creation, information processing and reasoning. No clear cut definition of concepts can be easily stated, however, it can be delimited based on a broader notion focused on their central role in creating and using information to act in the world, and the implicit fact that interaction is chief for their acquisition and processing. In particular, learning should be seen more as concerned with grounding the possibilities of action in a given situation; whilst reasoning or conceptual processing, should be seen in relation to the emergence and enactment of such possibilities in context.

3.1. Characteristics of concepts and conceptual processing

Here a framework is proposed consisting of a set of framing characteristics about concepts and conceptual processing. They are elaborated to delimit the way in which information is to be used in a learning system. That implicitly delineates as well what a concept cannot be reduced to.

The framework proposed consists of three main groups of characteristics delimiting the capability of a system to form and process concepts.

3.1.1. Format and flexibility

- A concept should not be limited to a particular situation-invariant unit but represented by diverse adapting units distributed throughout the system and dynamically arranged depending on context.

- The forming elements of a concept have variable contributions depending on context. They simultaneously include features with characteristics ranging from more stable and dominant, up to features represented by variable activations modulated by situations.
- The relevance of a forming elements of a concept and their influence on processing are to be dynamic and dependent on the evolution of an active interaction with the environment.

3.1.2. Formation and elicitation

- Concepts arise from context-specific ignitions of dynamically recruited feature arrays, where activations are primed by situations and driven by dynamic guidelines that evolve with experience.
- Concepts cannot be stated as pre-specified or stable entities but instead as occasioned by a given activity and situation inside which they are actively produced, maintain and enacted.
- Forming of a concept should be achieved through experience and interaction, and only constrained by the agent's body and information coming from: its internal states, its own motion, or sensory data captured from the environment.
- Formation implies multi-modal integration beyond the association of independent learning results in different modalities.

3.1.3. Grounding

- Concepts are to emerge from modality-specific situated representational features, therefore including information about context, situations and actions.
- Concepts should be partially constituted by perception (from external and internal cues or simulations) and action, where abilities in both modalities are essential during formation but not necessarily during processing or elicitation.
- Recall implies simulations of situated modality specific information.

4. Concepts and learning in artificial systems

In this section, implicit notions and underlying assumptions on cognition, learning and conceptualisation in AI and robotics are analysed from the ideas presented so far. In particular, computational modelling and implementations that relate to the ideas of concepts, grounding, or embodiment, are reviewed. It is important to stress that the focus here is on the underlying ideas concerning concepts, cognition, and intelligence, and not on contributions of the works to particular problems, functionalities or other research agendas.

4.1. Categorisation

In the reviewed papers the idea of building concepts as an equivalent to creating categories or symbolic representations is widespread.

A prominent contribution along the lines of cognitive modelling comes from Ron Sun and colleagues. They have developed CLARION, a hybrid cognitive architecture aiming at psychologically plausible models of knowledge representation and processing. For it, the discussion over localist and distributed representations is considered, and qualities of both stressed. In particular, CLARION presents different processing and learning schema for the two kinds of representations, but the symbolic ones are understood as concepts. Sun argues on the relevance of such dual view in Sun (2013) by stating that symbolic representations can be shown to be psychologically and computationally significant. Thus, they claim that both kinds of representations might be needed to explain the full range of cognition.

For example, based on CLARION (Licato, Sun, & Bringsjord, 2014) a specific separation of the representational spaces is used to achieve structured knowledge for deductive and analogical reasoning. In that approach, there is a clear separation between localist and distributed accounts based on two different processing mechanisms. Thus, the two functionalities are not integrated into a unified system, but pattern matching is used to go from one to the other. That exemplifies a transformation from distributed representations to localist ones due to the inherent need that arises from understanding reasoning and conceptual processing as isolated from the sensorimotor system.

The proposals based on the CLARION architectures have been shown to explain many psychological phenomena, but do not account for a distributed integration of perception and reasoning of the characteristics of concepts exposed in Section 3. Notably, CLARION is substantially focused on language-based and logic-centred measures and capabilities, and thus the system is forced to respond with explicit representations to ensure transparency. Regarding interaction, Sun (2013) points out its relevance in CLARION for the acquisition of symbolic representations, but a general connection to the embodiment and grounding of concepts is missing. In particular, an exhaustive relation to perception, as elaborated for the simulation accounts for processing, is still not clear.

Minai and colleagues have proposed similar approaches based on categorisation (Iyer & Minai, 2011). They present a neurodynamical system for the formation of categories based on the extraction of contextually relevant features. The idea of concepts in that work is related to sets of features shared by the members of a class. Such features can be re-utilised independently of context but learned contextually through their occurrence on different categories. That idea is concerned with contextual learning. Nonetheless, it is clearly focused on the creation of categories for isolated processing independent of context or interaction

with the environment; which is contrary to what has been stated concerning EC, GC and the dynamism of concepts.

The previous are examples mostly focused on the categorisation approaches to learning, but without much emphasis on embodiment and grounding. Other methods have focused more on those notions, but the way they are understood and implemented varies broadly. Approaches range from systems with clear separation between control and reasoning (for example ACT-R/E Trafton et al., 2013), to more advanced ones where motion is an essential part of learning. Some approaches hold similarities in the way they address learning as they try to link amodal symbolic representations or particular states to the consequences of actions (Kalkan et al., 2014; Krüger et al., 2011), or to patterns of sensorimotor information (Mohan, Morasso, Sandini, & Kasderidis, 2013; Stramandinoli, Marocco, & Cangelosi, 2012). Also, intrinsic motivations and interaction with humans have been used as a central part of learning and exploration (Ivaldi et al., 2014).

When analysing those works, it is clear that there is a predominant symbolic understanding of concepts. In general, a classification or categorisation is central, though in some approaches it is more subtly implicit in the design of experiments, or even in the training data. For example, in the case of motion information, patterns may be considered as pre-specified and restricted to symbolic categories in the training samples. Consequently, the movement would be used not as a part of the concept creation, but as an isolated pattern to be classified; which is not precisely accounting for embodiment or grounding, nor for a dynamic learning process.

Another way in which tacit knowledge about what is being learned can be imposed, is by using sensors where the meaning or ontology of the perceived data is given before learning. That is, there is only one possible meaning the data can have independently of the task or situation in which the agent might be. For instances, a sensor that detects predefined relevant objects, or a distance sensor.

Particularly, in Mohan et al. (2013) it can be found that movement samples within particular classes constrain learning to a set of human instructions, and thus forces the result towards symbols representing such patterns. That is more explicit in Ivaldi et al. (2014), where sets of basic modular actions and rules are classified for the communication between otherwise isolated modules.

With similar underlying principles, the proposals in Zibner, Tekulve, and Schoner (2015), Sandamirskaya, Zibner, Schneegans, and Schöner (2013), and Barakova and Chonnaparamutt (2009) use dynamic field theory to achieve behaviours that are indeed dynamically adapting to the situation and are, in principle, grounded in the sensorimotor systems. They reach these results by acting on targets sensed as activations in the neural fields which represent objects or proto-objects. In other words, the dynamics are performed on top of implicit assumptions about the environment as the way objects are differentiated from the background, or how such information is translated to a position. Consequently, the actual sensory processing is not learned through interaction, nor enacted depending on context, and thus there is no effective multi-modal integration to structure the environment dynamically. On the contrary, there is processing on activations with a predefined symbolic meaning or ontology.

These examples show how closely related the ideas of concepts and learning are, but more importantly, what the implications of such relation might be. A particular interpretation of learning and processing might lead to the development of approaches that implicitly generate unovercomeable constraints as a consequence of the assumptions on which they are built. Here it is argued that the need for symbols leads to constraining learning by the pre-definition of categories in the input data. One reason why this is problematic is that one cannot assume that real sensory data contains specific semantics in itself. Thus, the only way to solve a problem under such assumptions is by manipulating the sensory data in accordance to the pre-specified symbols or desired ontology. However, that does not mean that the imposed definition is to be sufficient or optimal. Therefore, representations should emerge from the interaction of the agent with the environment, and be grounded as part of its active development.

Along those lines Mirolli (2012) argues on the need for representations in embodied agents. He elaborates the idea by designing a task that cannot be solved without some kind of internal representation. His conclusions aim at a reconciling point in line with many of the postulates here. He states that EC is not in contrast with all representations but only with the symbolic ones.

Furthermore, representations should also be grounded in the sensorimotor system via top-down influence on sensory processing accounting for active simulation. Given that, proposals like the ones found in Mohan et al., 2013 use the notions of simulation as part of the information processing in learning about objects, actions and relations between them.

More prominently in the literature, Stephen Grossberg and colleagues have proposed and developed the Adaptive resonance theory (ART) (Grossberg, 2013, 2017). A primary focus in that architecture is the so-called stability plasticity dilemma, which deals with the problem of how to actively update knowledge whilst not losing important memories. To face that, they propose the "match learning". This type of learning only occurs if a sufficient match is found between bottom-up information and top-down expectations. In the case of a match, the category is refined to include the new information. When a memory search does not match any known category, a process is activated to learn a new one.

ART is in line with the ideas of simulation and stresses on the relevance of integrating bottom-up and top-down flows of information for an active perception. Equally, it has the advantages of representing actively as in the case of active inference approaches. Nonetheless, the ART architecture does not explicitly present the benefits of

probabilistic inference, and focuses on the idea of categorisation as a primary goal of perception, which can be possible without interaction.

In sum, the proposals based on ART do not address interaction as necessary for learning and representing, meaning that processing is not grounded in action, nor necessarily situated or contextual. In that sense, representations will only respond as categories based on feature prototypes, and will not necessarily depend on their relation to the action or the intrinsic goals of the agent.

4.2. Generative models and sensorimotor integration

As mentioned before, simulation of sensory information has a central role in cognitive processing. Thus, a primary problem to be addressed is how to link such simulations to active representations that emerge with the agent's actions whilst immersed in a particular situation or task. It has been mentioned that Friston's ideas on active inference (Friston, 2010; Friston et al., 2010; Friston et al., 2011) explain plausible mechanisms for empirical results concerning the role of simulation and active representations on cognitive processes. As described in Section 2.2, active inference proposes an explanation of how agents achieve stability by minimising a free energy function that represents beliefs about unknown causes in the environment.

A major problem is how a model based on such notions can be learned through experience and interaction. In many proposals these ideas have been interpreted in terms of probabilistic, and particularly Bayesian, methods and generative models (GMs).

For example, the hierarchical prediction machines as examined by Clark (2013), which are closely related to the active inference approaches, are argued to offer the best option for unifying mind and action.

Some advantages of probabilistic inference include rapid learning, contextual reactions and causality (Chater, Tenenbaum, & Yuille, 2006). Implementations exemplifying these models can be found in cognitive systems like the ones elaborated by Mazzu, Morerio, Marcenaro, and Regazzoni (2016) and Biresaw, Cavallaro, and Regazzoni (2015). They have worked on using Bayesian approaches in developing cognitive systems that learn temporal relations to make active inferences. With these proposals, great improvements in accuracy and performance in signal processing tasks have been achieved, with particular importance for video analysis.

Such kind of solutions focuses on structuring knowledge and building inferences through GMs. However, the connection to the actual emergence of the representations on which such inferences are made is performed through separated learning stages. Therefore, a central challenge is the unification of such processes.

Another related problem is the need for integrating different sensory modalities when all sensory data is acquired without any particular meaning attached to it implicitly. Such problem can also be extended to the integration of motion and perception since, as suggested by Friston, Adams, Perrinet, and Breakspear (2012) and Friston et al. (2010), the role of motor control in reducing uncertainty can be understood as fulfilling prior expectations about proprioceptive sensations. In that sense, the motion is not a consequence of perception but an active part of it. Along those lines, Wilson (2002) had developed views on automatic behaviour as emerging form internal representations of a situation. In her opinion, processing is more feasible and efficient if rather than entirely encoding the world around, epistemic actions are used to alter the environment in order to reduce the cognitive work required.

Therefore, it is crucial to address the mentioned integration in the inference process as an active and emergent relation from which behaviour and reasoning can arise. Proposals as the ones in Friston et al. (2012) or more recently (Pio-Lopez, Nizard, Friston, & Pezzulo, 2016) take that into account, but assume that the GM is given. To overcome the need for fully pre-specified GMs, these should be learned through interaction and as a way of self-organization.

Along those lines, Tani (2014) has presented a set of works that build higher-order cognition in direct connection to the sensorimotor systems. The results link motion of a robot to the fulfilment of expectations and prediction error minimisation. Such expectations and the constitutionality of sensorimotor relations in internal states are learned from the continuous flow of information through the sensorimotor system, and concerning a given task.

In Park and Tani (2015) and Murata et al. (2015) those ideas are extended to dynamic interactions employing recurrent networks and the inference of probabilistic states. The network is trained by reducing prediction error, though internal updates are not based on it. The network is trained instead based on the construction of various abstraction levels with different time scales.

Those implementations present a limitation in that some preprocessing in the inputs is used, imposing an ontology or a semantics on the input data. For example, the position of a tracked object is assumed as an input to the system, implying that the agent relies on a symbolic representation that is not related to its learning process but on top of which it performs all its reasoning.

Part of such limitation is addressed in Hwang, Jung, Kim, and Tani (2016), where Hwang and colleges have included convolutional neural networks in the architecture. That change allows the system to process raw sensory data, overcoming the mentioned impositions.

Tani, together with Friston and Haykin, had argued in Tani, Friston, and Haykin (2014) that the core ideas in those works are in line with the active inference postulates, and show a path along which many challenges related to what is pointed out here could be addressed. Some of the ways in which Tani's works may differ from the postulates by Friston, however, is that the internal states generated are not treated in probabilistic terms and that the prediction error is used during training but not as the only input

to the network. Equally, predictions about future states are done deterministically both in training and testing, which implies that the system does not model variability in data, nor considers probabilistic relations and biases between internal states and actions.

4.2.1. Deep generative models

A set of approaches focused on learning generative models capable of creating realistic samples come from the deep learning (DL) community. That responds to a change from discriminative to generative models as a way of overcoming the need for labelled data in supervised learning, or of addressing goals such as building more biologically plausible DL methods (Bengio, Lee, Bornschein, & Lin, 2015). Equally, some attempts look for bringing together the inference capabilities of probabilistic models and the generalisation advantages of neural networks.

Deep GMs have been developed for years as a tool for unsupervised learning. These include from works on deep belief networks and restricted Boltzmann machines as reviewed in Salakhutdinov (2015), to training methods based on generative adversarial networks (Goodfellow et al., 2014). Lately, efforts have been focused on training probabilistic GMs based on back-propagation as an alternative to maximum likelihood methods and processes like Gibb's sampling. For example, in Bengio, Thibodeau-Laufer, Alain and Yosinski (2014) it is proposed a method based on different layers of latent variables which learn the transition operator of a Markov chain to estimate distributions of data.

Along these lines, Kingma and Welling (2014) and Rezende, Mohamed, and Wierstra (2014), have merged advantages of DL and Bayesian inference into one architecture and a single learning process. They introduced general-purpose methods capable of generating realistic samples and inferences about input information through continuous latent variables. With those methods, generative and recognition models are trained simultaneously through back-propagation. In Gregor, Danihelka, Graves, Rezende, and Wierstra (2015) and Rezende, Mohamed, Danihelka, Gregor, and Wierstra (2016) those proposals are extended to mimic foveation and attentional processes for inference and generation.

In Chung et al. (2015) and Fabius and van Amersfoort (2014) the authors have explored the inclusion of latent variables over the hidden states of a recurrent network with the aim of extracting temporal dependencies of data sequences for inference and generation.

Though relevant connections can be found between the GMs based on DL and the ideas here exposed, further steps have to be taken to achieve self-organising capabilities that can actively process information, and which can deal with multi-sensory integration and in particular with active emergence of behaviour.

5. Conceptual processing as a dynamic and emergent phenomena

Primary challenges to address have been stated from the perspective of concepts and their relation to learning and knowledge representation. In this section, a particular architecture is proposed and tested in a sample scenario to have a more technical specification towards those challenges.

Out of the main problems, the focus here lays on the integration of different modalities and action generation. The later is also approached as a form of multi-modal integration as it causes proprioceptive data, which might is a predictable sensory modality. Then, actions are performed as reflexes activated by expectations to be fulfilled.

One main point to address with an architecture along these lines is the capability of learning through and for interaction. As highlighted by Sun (2013), meaning is not in the world neither in the internal dynamics of the agent alone, but in their interaction.

Then, learning is looked at from the perspective of interaction and can be viewed under a phenomenological framework, as representations and behaviour emerge dynamically during the interaction.

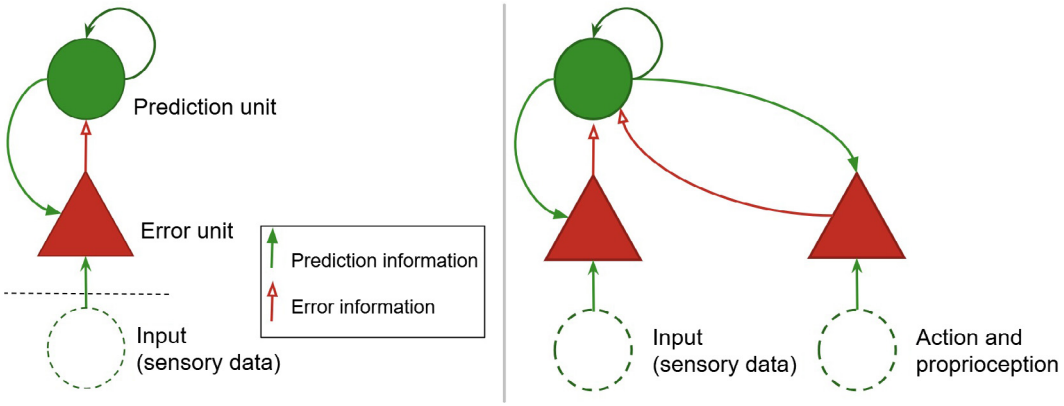

Fig. 1. Left: A schematic idea of an architecture based on error propagation and active predictive coding. Right: Schematic of active inference where both action, proprioception and sensory data are predicted. The internal state is updated based on the propagation of error from both predictions.

5.1. A model for conceptual learning and processing

A scenario is set where no assumption about any specific semantics in the sensory data is made; only a particular task is considered out of which the agent is to dynamically construct useful representations to enact an internalised coupling between its actions and the environment. As mentioned by Engel et al. (2013), here cognitive processes are analysed with respect to their role in action generation as a capacity for actively creating structure; thus all the agent learns makes sense in its particular task domain, where it acquires meaning in context.

Moreover, the emphasis is on control mechanisms as constitutive of perceptual processes, where simulation and particularly prediction, are seen as a reenactment of sensorimotor modalities, through which all processing and representations are grounded.

The model here proposed is in line with the probabilistic interpretation of active inference, but the update based on prediction error is made explicit, in contrast to an update of prior beliefs directly from observations. Equally, the transition, observation, and generation models are to be learned altogether from sample data in an end-to-end fashion.

To illustrate the core idea behind, in Fig. 1 some schematic descriptions are presented. The internal states in the prediction unit are used to construct estimations about the environment. Such predictive signals are then compared to information coming up from the sensory inputs. The prediction error, together with the current internal states, are used to update the representation of the environment. In general, both, the internal states and the predictive signals are probability distributions.

The ideas developed below assume an agent with sensory inputs, including proprioceptive information, and some defined motor capabilities.

The primary goal is to implement an active integration of the modalities, including proprioception, through a single set of latent variables as internal states updated with ascending prediction errors. Important elements in the environment are to be actively interpreted, and adequate behaviours generated dynamically for the specific interaction on which the system is trained.

The system is to actively represent the world through action as the core element of the perceptive process, allowing the agent to produce adequate behaviours. Thus, it should build a sufficient world model that is not only dependent on observations, but that actively updates itself with actions taken as a fundamental and necessary part of the world being represented or cognitively constructed.

The model proposed holds some relation to the work of Tani et al. (2014), Park and Tani (2015), Murata et al. (2015), and Hwang et al. (2016), but the here proposed approach focuses on internal activation as probabilistic states, and on using the prediction error as input information to the network. Equally, actions are generated as the fulfilment of expectations about proprioception.

Given the probabilistic nature of the problem, and that the different elements involved in the same learning process, the model is based on variational deep GMs. As has been shown in Kingma and Welling (2014), Rezende et al. (2014), Gregor et al. (2015), Rezende et al. (2016), Chung et al. (2015) and Fabius and van Amersfoort (2014), continuous variational representations can be constructed directly from data with end-to-end systems optimised with stochastic gradient descent. In particular Johnson, Duvenaud, Wiltschko, Adams, and Datta (2016), Chung et al. (2015), and Fabius and van Amersfoort (2014) have shown ways of including time dependencies on the inference process so that the state of the latent variables at a given time is dependent on the previous ones.

Thus, to address the need for dynamic updates of internal states, the principles of variational recurrent neural networks (VRNN), as presented in Chung et al. (2015), are used as a core part of the central block of the model. VRNNs add temporal dependencies to the variational auto-encoders (VAE) introduced in Kingma and Welling (2014). For an illustration of the proposed architecture see Fig. 2.

In a VAE an observation X is used to infer latent variables Z, which are to capture the variations in X. In particular, an inference model is considered to approximate $P(Z|X)$ by means of a NN. Such estimation is the inference of causes of X in a latent space. $P(Z|X)$ is assumed to be Gaussian: $Z \sim \mathcal{N}\left(\mu_x, diag(\sigma_x^2)\right)$, with $[\mu_x, \sigma_x] = \varphi^X(X)$, and φ^X approximated via a NN. The prior $P(Z)$ is assumed as normal $P(Z) \sim \mathcal{N}(0, I)$. The conditional $P(X|Z)$ is also approximated via a NN, and assumed to be Gaussian. A sample z is calculated as $z = \mu_x + \sigma_x * \epsilon$ with $\epsilon \sim \mathcal{N}(0, I)$ To train such model a loss is calculated to minimise the Kullback-Leibler divergence between $P(Z|X)$ and $P(Z)$, and maximise $log(P(X|Z))$.

The VRNN model introduces dependency on previous states for inferring Z_t at time t. That is done by encoding the sequence through a RNN, particularly based on LSTM (Hochreiter & Schmidhuber, 1997). The prior $P(Z_t)$ is calculated based on the recurrent state h_{t-1}.

The VRNNs solve part of the problem at hand but do not address the relation of representations to motion generation directly, either there is a direct relation between prior construction, future states inference, and action information. Equally, updates are not based on prediction error. These two points change the kind or representations being learned, as well as the processing cycle.

5.1.1. Model description

5.1.1.1. Prediction of sensory inputs. In VRNN the generative model approximates a distribution for the input signal at each time step as: $P(X_t|Z_{\leqslant t}, X_{\leqslant t})$; however, given the predictive nature of the approach being implemented, the generative model should produce at time t:

$$P(X_{t+1}|Z_{\leqslant t}, X_{\leqslant t})$$

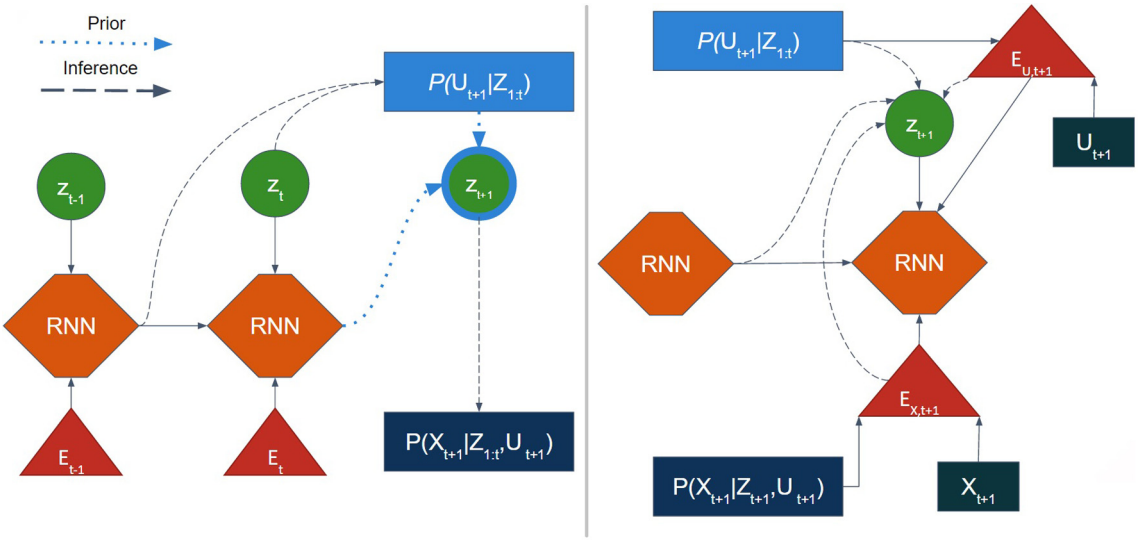

Fig. 2. Left: The calculation of U (control signal) and the prior estimation of Z (internal state). Right: Update of the posterior of Z given the prediction error, the previous states and the control.

5.1.1.2. Control and proprioception prediction. In general terms, control or actions, and proprioception can be seen simply as a part of the sensory input, thus could be predicted with the same generative model as the rest of the sensory inputs. However, given that the control taken can also change the priors on the next state, as detailed later, a separate generative model is considered to approximate:

$P(U_{t+1}|Z_{\leqslant t}, X_{\leqslant t})$.

The signal U is to be fulfilled by reflex arcs given the prediction, as elaborated in Friston et al. (2012). Nonetheless, the proprioceptive feedback due to such action, and the actual actions taken, may differ from expectations. In that sense U_{t+1} is to be composed by the control (C_{t+1}) and the expected proprioception in $t+1$ (R_{t+1}). C_{t+1} corresponds to the action taken at time t to arrive from the state Z_t to the next Z_{t+1}.

$U_t = [C_t, R_t]$

In particular, the update of internal states should depend on the prediction error of U and not only of sensory inputs.

Equally, as the generative model is to produce a distribution over motion possibilities, in order to actually take an specific action, then a random one is drawn from the distribution, in that way the network is to estimate the implicit variance in the control inside the sample data during training. During test the most likely action is always taken.

Moreover, this explicit separation between internal states representing causes, and the action generation is necessary to actually build concrete representations in probabilistic terms, as well as to manage the information flow in terms of predictions and error propagation.

$P(X_{t+1}|Z_{t+1})$ depends on the prior Z_{t+1}, which is to be calculated depending on previous states, and the control predicted.

Under this model both Z and the generative models approximate Gaussian posteriors. Particularly for $P(Z)$ the parameters μ and σ of the prior depend on the input sequences through a learned approximation that might change over time for a specific scenario.

5.1.1.3. Generative model. The generation process is given by

$$X_{t+1}|Z_{t+1} \sim \mathcal{N}\left(\mu_{X,t}, diag(\sigma^2_{X,t})\right), \text{ where } \left[\mu_{X,t}, \sigma_{X,t}\right] = \varphi^X(Z_{t+1})$$

In implementation it is really calculated using $\varphi^X(\mu_{o,t})$, where $\mu_{o,t}$ is the prior expectation of the next state. The prior distribution then has to match the posterior by minimising the divergence between the two, as explained later.

$$U_{t+1}|Z_{\leqslant t} \sim \mathcal{N}\left(\mu_{u,t}, diag(\sigma^2_{u,t})\right), \text{ where } \left[\mu_{u,t}, \sigma_{u,t}\right] = \varphi^U(Z_t, h_{t-1}) \text{ and } h_{t-1} \text{ the recurrent state of the RNN in the previous time step.}$$

5.1.1.4. Prior construction. In VRNNs the prior of Z is calculated from h_{t-1}, but given an action the transition may change and thus this prior will also depend on it such that:

$$Z_{t+1} \sim \mathcal{N}\left(\mu_{o,t}, diag(\sigma^2_{o,t})\right), \text{ where } \left[\mu_{o,t}, \sigma_{o,t}\right] = \varphi^{prior}(h_t, U_{t+1}).$$

Again, the prior is calculated using $\varphi^{prior}(h_t, u \sim U_{t+1})$ during training, and the expectation of U during test $\varphi^{prior}(h_t, \mu_{u,t})$.

It is written for Z_{t+1} to stress the fact that is used for prediction.

5.1.1.5. Error calculation. Also inspired by Friston et al. (2012) and as implemented in Murata et al. (2015), the error signal is proportional to the difference between predicted and actual input, and to the precision or inverse covariance. Given that the covariance matrix is assumed as

diagonal in the VRNN, the error is calculated for X at time t as:

$$E_{X,t} = \left(\mu_{X,t} - X_t\right)/\sigma_{X,t}$$

5.1.1.6. States update. At each time step both the variational representations Z and the recurrent states h_t are updated. First the posteriors of Z are inferred taking into account U, the recurrent state h and the prediction errors.

$$Z_t|X_t, U_t \sim \mathcal{N}\left(\mu_{z,t}, diag(\sigma^2_{z,t})\right), \text{ where } \left[\mu_{z,t}, \sigma_{z,t}\right] = \varphi^{enc}(\varphi^E(E_{X,t}, E_{U,t}), h_{t-1}, U_t).$$

Also sampling U_t only for training.

Then the RNN can be updated by the optimised function:

$$h_t = f_\theta\left(\varphi^E(E_{X,t}, E_{U,t}), \varphi^Z(Z_t), h_{t-1}\right)$$

All φ^X, φ^U, φ^{prior}, φ^{enc}, φ^E can be approximated by neural networks, in the presented implementation this is done with multilayer perceptrons, and using LSTMs (Hochreiter & Schmidhuber, 1997) as RNN. The NNs used are: for φ^X 4 fully connected layers (FC) with sizes (256, 256, 256, 600). For φ^{prior} 2 FC $(256, 4)$, φ^U 2 FC (256, 256), whilst for φ^{merge} and φ^Z 1 FC with sizes 512 and h_si respectively. φ^{enc} is divided in two, first a block of 2 FC with inputs h_{t-1} and φ^{merge} and size h_si both; the second part takes as input the output of the first and the control estimated in previous time, and is formed by 2 FC (256, 4). Outputs of φ^{enc}, φ^Z and φ^X are divided in two, one with linear activation, and the other softplus, corresponding to mean and standard deviation of the Gaussian distributions.

5.1.2. Training

The objective function to be optimised is formed by three parts. The first part corresponds to the minimisation of the Kullback-Leibler divergence between the approximated distribution and the prior:

$$KL(P(Z_t|X_t, U_t)||P(Z_t))$$

The two other parts correspond to the maximisation of the log likelihoods of $P(X_{t+1}|Z_{t+1}, U_{t+1})$ and $P(U_{t+1}|Z_{\leqslant t})$ Thus for a sequence of T time steps the target to maximise is given by:

$$\mathcal{L}_T = \frac{1}{T}\sum_{t=1}^{T}[-KL(P(Z_t|X_t, U_t)||P(Z_t)) + logP(X_{t+1}|Z_{t+1}, U_{t+1}) + logP(U_{t+1}|Z_{\leqslant t})]$$

5.1.3. Scenario

The scenario in which the architecture is tested is a navigation task in a simulated environment. The navigating agent has two sensory modalities to be integrated, the visual input from a mounted camera, and the proprioception, based on accelerometer measurements and wheels speed. The agent is to navigate by avoiding obstacles without any particular goal position. That implies that the objects and free paths are to be represented in the form of actions that emerge as an internalisation of the environment, and not with an imposed or predefined processing for it. In other words, a conceptual representation in terms of a coupling between the characteristics of the world and action is created for the particular task.

The training is performed on sequences of sensory and control data pairs from a simulated robot performing a programmed navigation task. The training data at time t corresponds to the expected prediction from the internal states of the agent. Such prediction is formed by an RGB image from the mounted camera at time $t+1$, and a matrix containing the control information for the transition from Z_t to Z_{t+1}, and the acceleration data sensed at time $t+1$. Then a set of T such data pairs is given as a single training sequence.

5.2. Results

To measure the success of the architecture the central notion to keep in mind is that action are reflexes fulfilling expectations about the environment. Since such predictions are generated from the same states from which sensory information is predicted, the results might imply that action and perception are part of the same representational process. Moreover, that indicates that the world is represented in terms of an environment action coupling.

The capability of the network to produce adequate expectations for control and proprioceptive signals is measured in terms of the prediction error with relation to the behaviour in the original sequences.

The architecture is trained over 5000 sequences of length $T = 20$, using mini-batches of size 40. A first testing scenario is considered with the same visual characteristics as the training set, but with different positions of the elements in the environment, with which it is tested that the system is not merely learning the sample sequences.

Moreover, the internal states of the agent make sense in a particular situation. However, the same task implies the same processing even when in a different environment. That is, the representations can be translated to scenarios where the action-environment coupling learned is used to perform the same task in a new environment. To test that, a second set, called test 2, is designed with different visual characteristics of the environment and different position of elements to the ones in the training set (see Fig. 3).

For both training and testing the data is normalised by subtracting the mean and dividing over the standard deviation of the training set.

In Fig. 4 are shown examples of prediction and expected action controls over sequences of 400 time steps of the training, test and test 2 sets.

The results over the whole sets are shown in Table 1, upper part for U.

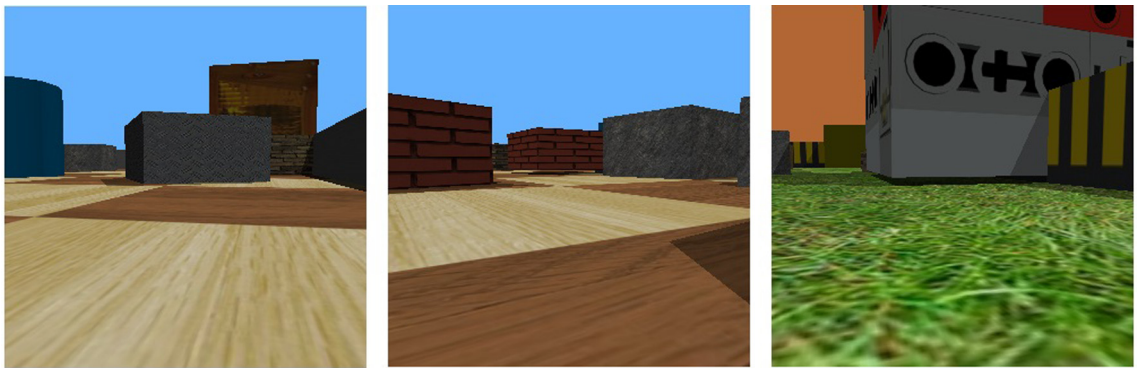

Fig. 3. Samples images from the mounted camera of the robot from left to right, the training, test and test 2 sets.

Fig. 4. Prediction (blue) and expected control (green) over a 400 time steps sequence. Training data in the left, test data in the centre, and test 2 in the right. The horizontal axis is time, and the vertical axis is the speed applied to the wheel from 0 to 1, being 1 maximum speed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Given that the probability distribution is optimised for the training set, the standard deviation is small for the characteristics of that set; thus, as is to be expected, the negative log likelihood for the other sets is higher. Nonetheless, when comparing the difference between the actions produced (predicted) and the expected behaviour regarding the root mean square error, it can be seen that the prediction error for the two test sets is not significantly different. That is so since the greatest part of the error comes from the situations where small movements are expected, some of which are noise.

On the other hand, in Table 1, lower part, similar results are shown for the visual modality. In that case, the differences are considerably higher between the two tests, though between the training and the first test the difference can be neglected. That happens since the generative model is tuned to produce different visual characteristics than the ones in test 2. However, taking into account these two results one can appreciate that, even if the visual characteristics are different, the system is generating internal representations that are general enough to reproduce the expected behaviour in great extent, also in unknown environments.

Moreover, the higher errors seen in the control signals for test 2 may arise from the larger prediction errors used to update the internal states caused by the different characteristics in that environments. That can be seen when the KL divergence is considered for the two tests, as shown in Table 2. In that case, it is clearer that due to the prediction error caused by the difference between predicted and real visual characteristics in test 2, there is a higher divergence between the prior and the posterior of internal states.

As expected, the biggest reconstruction error occurs when tested in a different environment, however, the magnitudes of the likelihoods for control generation does not change significantly.

Similarly, as has been stated, there is no single semantic segmentation of the data itself, but relevant representational states are constructed by the agent to achieve an adequate interaction with the environment under given conditions. That is, the representations do not respond to any external need or imposition beyond the ones strictly necessary for the agent to enact its internal states. Those states are actively created in terms of what is useful for the agent, its constraints, the environment and the situation; in this case, all given by the navigation task.

The representational states generated by the agent after training are dynamic and continuous but can be visualised and compared to an arbitrary semantics that makes sense for a human observer. With a simple example (see Fig. 5), it can be illustrated how internal states that are relevant for the task are generated in a predictable and even

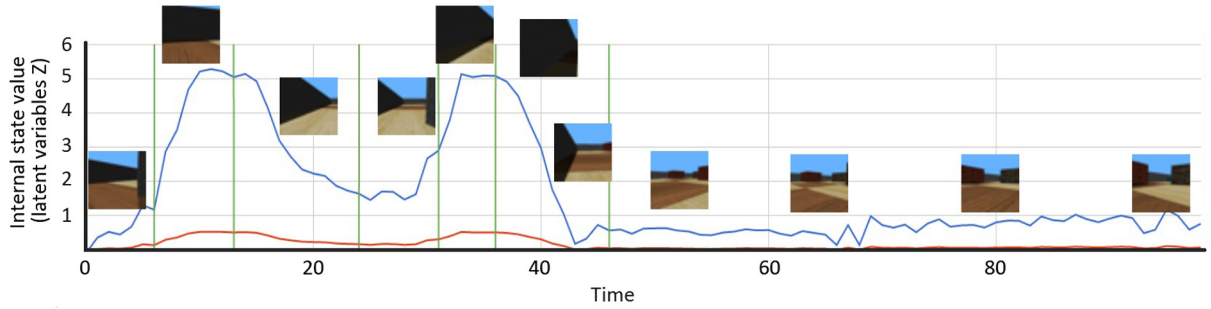

Fig. 5. Internal states of the network Z_t over 200 time steps, and the corresponding visual inputs in the separated periods. The periods, separated by vertical lines, are estimated manually by a human observer as corresponding to different moments in the navigation.

Table 1
Prediction accuracies in terms of negative log-likelihood and root mean square error.

Set	Mean $-\log P$	std $-\log P$	rms
Control			
Training	0.02	19.98	0.08
Test	10.77	87.37	0.17
Test 2	36.79	238.74	0.19
Visual input			
Training	0.83	5.52	0.65
Test	0.93	8.91	0.70
Test 2	4689	38,503	1.16

Table 2
KL divergence.

Set	mean KL	std KL
Training set	1.63	5.25
Test set	1.42	5.08
Test 2	99.27	131.94

descriptive way. This is so only because the task is known by the observer, and thus the internal states of the agent make sense in such context, not because the agent was thought to generate such states as meaningful for the observer. Moreover, that illustrates that matching states to represent specific symbols is not to be the learning goal. Internal states are significant for the behaviour to arise, it is only when the communication between agents is to be addressed that such states may acquire a symbolic meaning, only useful for the agents involved and in a given context.

There are many kinds of objects in the environment. Nonetheless, the agent builds a small set of representations given that the visual differences between the objects (obstacles) are irrelevant for the task.

There is an action-environment coupling that works as the representation of the perceived world, not only from one of the modalities but from the coupling between sensory data, action and the represented world.

6. Discussion

Ideas of concepts and conceptual processing have been explored from different views in an attempt to frame a relevant notion for the understanding of cognition and in particular the development of AI and robotics. A general relation between the ideas of concepts and the understanding of cognition and intelligence has been highlighted. In such relation categorisation based approaches to concepts match the computational views of cognition, whilst ideas of concepts as flexible, distributed and context-dependent account more for the dynamic aspects of embodied and grounded cognition. In particular, under the light of active inference and predictive coding, the capabilities of reproducing the environment and processing situations based on predictions have been argued to be fundamental to the more dynamic views on concepts.

Moreover, it has been elaborated that the ideas of categorisation seen as central to learning in diverse AI approaches have led to challenges such as the SGP, which in turn is a logical consequence of assuming reasoning as a process isolated from the sensorimotor systems. That separation in great comes from relating conceptual processing or reasoning to language, which is more naturally understood as the manipulation of symbols.

Instead, conceptual representations should be understood as an emerging and dynamic process that is flexible and capable of adapting to the situation, in an act that implies actively creating the world through interactions.

That leads to the simple yet relevant notion, that a precise semantics cannot be imposed on any particular data. On the contrary, appropriate and dynamic representations should emerge from experience, which should be sufficient to interact appropriately with the environment in a given situation and under the agent's constraints, needs and perspective.

That does not neglect symbols. However, it assumes they are only relevant for processes such as communication and should emerge only from and for that kind of tasks. Moreover, the arguments here do not pretend to state that building crisp and stable symbolic models of the world is not useful for particular tasks or application domains. Instead, they suggest that the underlying process of learning and building representations is a broader and more dynamic process, and thus should be seen as such also at a technical level.

In particular, a more dynamic view of concepts may boost efficiency since limitations emerge when trying to build learning and reasoning from the realisation of stable representations. Such restrictions are imminent as encapsulating all possible inputs in crisp symbols easily grows in complexity. Instead, if conceptual processing and the information on which it is performed are dynamically linked to the situation, then the possibility to adapt to always changing and new scenarios becomes more feasible. That is, it is more relevant to learn how to deal with new inputs from existing knowledge by producing actions in relation to sensory data, than trying to cluster each incoming instance in symbols for further processing.

From what has been exposed, it can be said that perception and action emerge dynamically from the interaction with the environment, but none of such capabilities can be constructed without the other. Moreover, in the presented approach no isolated process can be labelled as cognition; it is nothing but the emergent phenomena that arises from the interaction with the environment. Then, representations emerge as a way to couple the environment and the agent, and thus meaning appears contextually and concerning actions. In other words, a representation does not mean anything by itself but is the action that emerges in the coupling with the environment what develops the meaning dynamically.

In that way, an agent builds representations that may ignore much of the characteristics of the input data, which under other approaches may seem informative. That might seem suboptimal, for example in reconstructing the input data from encoded information, since many details are missed; however, it can be instead interpreted as a way of abstracting only the necessary information, leaving aside non-informative details, irrelevant for the interaction at hand.

That reinforces the point on why a semantics that is seen as a necessary imposition on data, or a given ontology in the sensor capabilities on which processing is to occur, may limit the possibilities of the agent and is always incomplete. That is so since such impositions do not necessarily match the needed information for the agent to perform the task. That is, for example, a detailed semantics of a problem may limit the capability of an agent to represent as equal the visual features of elements externally defined by distinct symbols, but which under a given situation may lead to the same action, and thus should be represented equally.

Moreover, the capability of representing the environment to properly act does not imply only a specific clustering of sensory data, but to dynamically adapt to the situation. That is why the dependency on previous states is crucial to have a dynamic interpretation of the environment in opposition to a static process. In the same way, the capability of dynamically updating internal believes is essential to such process, in that the evolution of an interaction depends on the internal state of the agent and not only on the environment.

In the model presented such capability is not only achieved by the memory in the LSTMs, but is extended by the updates from prediction error. In that way, there are fundamental differences concerning, for example, entirely feed-forward approaches, where the actions and predictions are made only from the current state of the environment. In those cases, the process does not get a situated decision including the past.

The results here exposed can also be linked to Kiefer and Pulvermüller (2012), in that action information contributes to conceptual processing depending on learning experiences given that sensory and motor interactions shape conceptual memories. Moreover, such ideas are also connected to Engel et al. (2013), who argue that cognition should be studied primarily with respect to action generation for creating structure, for which an agent is to be immersed in its task domain for the internal states to acquire meaning.

Moreover, also in this matter Engel et al. (2013) suggest that perceptual decision making may be described as biases in sensory processing imposed by action contexts. In that sense, activity patterns cannot be taken as encoding action-invariant descriptions of objects and scenes, but as supporting the capability of structuring contexts. Equally, from the GC view (Barsalou, 2008) concepts are situated, meaning that the situation is imagined or simulated during conceptual processing. That is, a conceptual category cannot be context-agnostic by default.

Those notions are clearly present in the architecture proposed, since all internal states are situated, and are meaningful only for generating actions in a given task, for which the interpretation of the sensory data is specific to the agent. Moreover, the ideas elaborated by Wilson and Golonka (2013) concerning EC, are directly related to the model presented, since for them cognitive phenomena thought to need symbolic or object specific representations can be replaced by the dynamics learned about entities and motion.

Another way in which the results can be interpreted is concerning the capability of learning through and for interaction. As highlighted by Sun in Sun (2013), meaning is not in the world or the internal dynamics of the agent alone, but in their interaction. However, such relationship is enacted by the activity of the agent and thus driven by its goals and on what is relevant to them.

In the scenario here presented the goals and motivations are implicit in the task; however this is clearly not a general case, and further exploration regarding internal motivations is needed since it lies in the middle of the discussion about how and what needs to be represented.

Along these lines, the relation between relevance and the development of intelligent agents is elaborated by Saariluoma and Rauterberg (2016), who present a set of arguments on why formal languages are limited as a tool for building thinking machines, since from such languages solely no relevance can emerge. In particular, they inquire about what can be expressed by formal concepts, and what

kind of languages are used to analyse human thinking. Amongst their arguments, they include the idea that human behaviour cannot be described by a fixed and finite set of rules since this cannot capture evolution and changes.

In particular, they focus on the fact that relevance cannot emerge from formal languages only, but all intelligent behaviour achievable by machines depends on a human definition of relevance. From a biological perspective, an agent develops the capabilities to define relevance from innate needs, which continually evolve as an irreversible process on an ever changing system. This problem is also linked to the SGP in that no symbol can be grounded from the formal language on which the machine runs, but such grounding finally depends on the connection defined by the programmer for a particular task.

One primary conclusion of Saariluoma and Rauterberg is that the problem of machine intelligence is rooted in the limited power of expression of formal languages, and suggest that a phenomenological approach to the problem could lead to a better understanding or modelling capabilities. Such notions can also be linked to the ideas here exposed in relation to the emergence of dynamic representational states contrary to the imposition of specific semantics on the sensory data. The model addresses the problem through a representational process grounded on interaction whit a clear phenomenological nature.

That reasoning starts to suggest a very relevant point for further development, which is how to involve inner needs and goals into the idea of learning and representing. In particular how to make the enactment and interaction based learning to be driven by internal needs.

Saariluoma and Rauterberg discuss that the problem of relevance, amongst others, relates to the intrinsic needs of the agent and how these intertwines with development and learning. Along similar lines, Stapleton (2013) states that a change in the attitude towards affect and emotion is missing from EC studies, for which an important start would be to focus on interoceptive, organismic basis of natural cognition as inherently entwined with affect. Equally, Martin (2015) describes results that account for the idea of socialemotional concepts being grounded as well, and thus, in general, even abstract concepts may be grounded on actions and perception, but also on the emotion systems. Ziemke and Lowe (2009) in turn discuss the relevance of emotion for EC and to which extend it would be needed in the development of artificial systems. Also, Kiefer and Pulvermüller state that along perception and action, abstract concepts may strongly depend on associations between emotional and introspective representations. Pezzulo et al. (2012) also elaborate this arguing that models should include perception, motor and affective systems.

Some recent approaches, based on the idea of deep reinforcement learning (Mnih et al., 2015) have shown the possibility of acquiring action policies directly from sensory data based on reward signals. Those ideas can be linked to learning based on internal needs, however, in such approaches, there is no constant feedback, nor simulation of the sensory input, whilst the processing is done in a feed-forward fashion based on the current states of the environment. Even though, the idea proves the feasibility of learning from unstructured sensory data based on given internal needs represented by reward. Then, the challenge to address in future works would be on how to embed such kind of reward or measure of internal needs into the learning and emergence of intelligent behaviour by means of dynamic interaction, whilst centred on grounding through simulation and predictive capabilities.

Acknowledgements

This work was supported in part by the Erasmus Mundus Joint Doctorate in Interactive and Cognitive Environments, which is funded by the EACEA Agency of the European Commission under EMJD ICE FPA n 2010-0012.

References

Anderson, M. L. (2003). Embodied cognition: A field guide. *Artificial Intelligence, 149*(1), 91–130.

Barakova, E. I., & Chonnaparamutt, W. (2009). Timing sensory integration. *IEEE Robotics & Automation Magazine, 16*(3), 51–58.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617–645.

Bengio, Y., Thibodeau-Laufer, E., Alain, G., & Yosinski, J. (2014). Deep generative stochastic networks trainable by backprop. In: Proceedings of the 31st International Conference on Machine Learning 226–234.

Bengio, Y., Lee, D.-H., Bornschein, J., & Lin, Z. (2015). Towards biologically plausible deep learning. arXiv preprint arXiv:1502.04156.

Biresaw, T. A., Cavallaro, A., & Regazzoni, C. S. (2015). Dynamic bayesian network modeling for self-and cross-correcting tracking. In *2015 12th IEEE international conference on advanced video and signal based surveillance (AVSS)* (pp. 1–6). IEEE.

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

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences, 10*(7), 287–291.

Chung, J., Kastner, K., Dinh, L., Goel, K., Courville, A. C., & Bengio, Y. (2015). A recurrent latent variable model for sequential data. In *Advances in neural information processing systems* (pp. 2980–2988).

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(03), 181–204.

Coradeschi, S., Loutfi, A., & Wrede, B. (2013). A short review of symbol grounding in robotic and intelligent systems. *KI - Knstliche Intelligenz, 27*(2), 129–136. http://dx.doi.org/10.1007/s13218-013-0247-2.

Cubek, R., Ertel, W., & Palm, G. (2015). A critical review on the symbol grounding problem as an issue of autonomous agents. In *KI 2015: Advances in Artificial Intelligence* (pp. 256–263). Springer.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition, 33*(1), 25–62.

Desai, R. H., Herter, T., Riccardi, N., Rorden, C., & Fridriksson, J. (2015). Concepts within reach: Action performance predicts action language processing in stroke. *Neuropsychologia, 71*, 217–224.

Dourish, P. (2004). What we talk about when we talk about context. *Personal and Ubiquitous Computing, 8*(1), 19–30.

Engel, A. K., Maye, A., Kurthen, M., & König, P. (2013). Where's the action? The pragmatic turn in cognitive science. *Trends in Cognitive Sciences, 17*(5), 202–209.

Fabius, O., & van Amersfoort, J. R. (2014). Variational recurrent auto-encoders. In *ICLR workshop track*.

Fernandino, L., Humphries, C. J., Seidenberg, M. S., Gross, W. L., Conant, L. L., & Binder, J. R. (2015). Predicting brain activation patterns associated with individual lexical concepts based on five sensory-motor attributes. *Neuropsychologia, 76*, 17–26.

Fields, C. (2014). Equivalence of the symbol grounding and quantum system identification problems. *Information, 5*(1), 172–189.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127–138.

Friston, K., Adams, R., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology, 3*, 151. http://dx.doi.org/10.3389/fpsyg.2012.00151.

Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: A free-energy formulation. *Biological Cybernetics, 102*(3), 227–260.

Friston, K., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics, 104*(1–2), 137–160.

Gabora, L., Rosch, E., & Aerts, D. (2008). Toward an ecological theory of concepts. *Ecological Psychology, 20*(1), 84–116.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology, 22*(3–4), 455–479.

Gärdenfors, P. (2004). *Conceptual spaces: The geometry of thought*. MIT Press.

Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. MIT Press.

Gentsch, A., Weber, A., Synofzik, M., Vosgerau, G., & Schütz-Bosbach, S. (2016). Towards a common framework of grounded action cognition: Relating motor control, perception and cognition. *Cognition, 146*, 81–89.

Ghio, M., Vaghi, M. M. S., Perani, D., & Tettamanti, M. (2016). Decoding the neural representation of fine-grained conceptual categories. *NeuroImage, 132*, 93–103.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672–2680).

Gregor, K., Danihelka, I., Graves, A., Rezende, D. J., & Wierstra, D., 2015. Draw: A recurrent neural network for image generation. In: Proceedings of the 32nd International Conference on Machine Learning (ICML-15).

Grossberg, S. (2013). Adaptive resonance theory. *Scholarpedia, 8*(5), 1569.

Grossberg, S. (2017). Towards solving the hard problem of consciousness: The varieties of brain resonances and the conscious experiences that they support. *Neural Networks, 87*, 38–95.

Gur, M. (2015). Space reconstruction by primary visual cortex activity: A parallel, non-computational mechanism of object representation. *Trends in Neurosciences, 38*(4), 207–216.

Harnad, S. (2005). To cognize is to categorize: Cognition is categorization. *Handbook of Categorization in Cognitive Science*, 20–45.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.

Hwang, J., Jung, M., Kim, J., & Tani, J. (2016). A deep learning approach for seamless integration of cognitive skills for humanoid robots. In *The sixth joint IEEE international conference developmental learning and epigenetic robotics (ICDL-EPIROB)*. IEEE.

Ivaldi, Serena, Nguyen, Sao Mai, Lyubova, Natalia, Droniou, Alain, Padois, Vincent, Filliat, David, Oudeyer, Pierre-Yves, & Sigaud, Olivier (2014). Object learning through active exploration. *IEEE Transactions on Autonomous Mental Development, 6*(1), 56–72.

Iyer, L. R., & Minai, A. A. (2011). A neurodynamical model of context-dependent category learning. In *The 2011 international joint conference on neural networks (IJCNN)* (pp. 2975–2982). IEEE.

Johnson, M., Duvenaud, D. K., Wiltschko, A., Adams, R. P., & Datta, S. R. (2016). Composing graphical models with neural networks for structured representations and fast inference. In *Advances in neural information processing systems* (pp. 2946–2954).

Kalkan, S., Dag, N., Yürüten, O., Borghi, A. M., & Şahin, E. (2014). Verb concepts from affordances. *Interaction Studies, 15*(1), 1–37.

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex, 48*(7), 805–825.

Kingma, D. P. & Welling, M. (2014). Auto-encoding variational bayes. In: Proceedings of the 31st International Conference on Machine Learning (ICML-14).

Krüger, N., Geib, C., Piater, J., Petrick, R., Steedman, M., Wörgötter, F., ... Omrčen, D. (2011). Object-action complexes: Grounded abstractions of sensory-motor processes. *Robotics and Autonomous Systems, 59*(10), 740–757.

Lakoff, G. (2012). Explaining embodied cognition results. *Topics in Cognitive Science, 4*(4), 773–785.

Lebois, L. A., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are automatic conceptual cores the gold standard of semantic processing? The context dependence of spatial meaning in grounded congruency effects. *Cognitive Science, 39*(8), 1764–1801.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(01), 1–38.

Licato, J., Sun, R., & Bringsjord, S. (2014). Structural representation and reasoning in a hybrid cognitive architecture. In *2014 International joint conference on neural networks (IJCNN)* (pp. 891–898). IEEE.

Machery, E. (2010). Precis of doing without concepts. *Behavioral and Brain Sciences, 33*(2–3), 195–206.

Man, K., Kaplan, J., Damasio, H., & Damasio, A. (2013). Neural convergence and divergence in the mammalian cerebral cortex: From experimental neuroanatomy to functional neuroimaging. *Journal of Comparative Neurology, 521*(18), 4097–4111.

Martin, A. (2015). Grapes grounding representations in action, perception, and emotion systems: How object properties and categories are represented in the human brain. *Psychonomic Bulletin & Review*, 1–12.

Mazzu, A., Morerio, P., Marcenaro, L., & Regazzoni, C. S. (2016). A cognitive control-inspired approach to object tracking. *IEEE Transactions on Image Processing, 25*(6), 2697–2711.

Meyer, K., & Damasio, A. (2009). Convergence and divergence in a neural architecture for recognition and memory. *Trends in Neurosciences, 32*(7), 376–382.

Mirolli, M. (2012). Representations in dynamical embodied agents: Re-analyzing a minimally cognitive model agent. *Cognitive Science, 36*(5), 870–895.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... Ostrovski, G. (2015). Human-level control through deep reinforcement learning. *Nature, 518*(7540), 529–533.

Mohan, V., Morasso, P., Sandini, G., & Kasderidis, S. (2013). Inference through embodied simulation in cognitive robots. *Cognitive Computation, 5*(3), 355–382.

Müller, V. C. (2015). Which symbol grounding problem should we try to solve? *Journal of Experimental & Theoretical Artificial Intelligence, 27*(1), 73–78.

Murata, S., Yamashita, Y., Arie, H., Ogata, T., Sugano, S., & Tani, J. (2015). Learning to perceive the world as probabilistic or deterministic via interaction with others: A neuro-robotics experiment. IEEE Transactions on Neural Networks and Learning Systems.

Nosofsky, R. M., Kruschke, J. K., & McKinley, S. C. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(2), 211.

Park, G., & Tani, J. (2015). Development of compositional and contextual communicable congruence in robots by using dynamic neural network models. *Neural Networks, 72*, 109–122.

Pessoa, L. (2014). Understanding brain networks and brain organization. *Physics of Life Reviews, 11*(3), 400–435.

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., & Spivey, M. J. (2013). Computational grounded cognition: A new

alliance between grounded cognition and computational modeling. *Frontiers in Psychology, 3*.

Pio-Lopez, L., Nizard, A., Friston, K., & Pezzulo, G. (2016). Active inference and robot control: A case study. *Journal of The Royal Society Interface, 13*(122), 20160616.

Quillan, M. R. (1966). Semantic memory, Tech. rep., DTIC Document.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*(1), 79–87.

Rezende, D. J., Mohamed, S., Danihelka, I., Gregor, K., & Wierstra, D. (2016). One-shot generalization in deep generative models. In: Proceedings of the International Conference on Machine Learning (ICML-16).

Rezende, D. J., Mohamed, S., & Wierstra, D. (2014). Stochastic backpropagation and approximate inference in deep generative models. In *Proceedings of the 31st international conference on machine learning (ICML-14)* (pp. 1278–1286).

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology, 4*(3), 328–350.

Saariluoma, P. & Rauterberg, M. (2016). Turings error-revised. *International Journal of Philosophy Study (IJPS) 4*, doi:http://dx.doi.org/10.14355/ijps.2016.04.004.

Salakhutdinov, R. (2015). Learning deep generative models. *Annual Review of Statistics and Its Application, 2*, 361–385.

Sandamirskaya, Y., Zibner, S. K., Schneegans, S., & Schöner, G. (2013). Using dynamic field theory to extend the embodiment stance toward higher cognition. *New Ideas in Psychology, 31*(3), 322–339.

Stapleton, M. (2013). Steps to a properly embodied cognitive science. *Cognitive Systems Research, 22*, 1–11.

Stramandinoli, F., Marocco, D., & Cangelosi, A. (2012). The grounding of higher order concepts in action and language: A cognitive robotics model. *Neural Networks, 32*, 165–173.

Sun, R. (2013). Autonomous generation of symbolic representations through subsymbolic activities. *Philosophical Psychology, 26*(6), 888–912.

Tani, J. (2014). Self-organization and compositionality in cognitive brains: A neurorobotics study. *Proceedings of the IEEE, 102*(4), 586–605.

Tani, J., Friston, K., & Haykin, S. (2014). Self-organization and compositionality in cognitive brains [further thoughts]. *Proceedings of the IEEE, 102*(4), 606–607.

Trafton, G., Hiatt, L., Harrison, A., Tamborello, F., Khemlani, S., & Schultz, A. (2013). Act-r/e: An embodied cognitive architecture for human-robot interaction. *Journal of Human-Robot Interaction, 2*(1), 30–55.

van Dam, W. O., van Dijk, M., Bekkering, H., & Rueschemeyer, S.-A. (2012). Flexibility in embodied lexical-semantic representations. *Human Brain Mapping, 33*(10), 2322–2333.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625–636.

Wilson, A. D., & Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology 4*.

Zibner, S. K., Tekulve, J., & Schoner, G. (2015). The neural dynamics of goal-directed arm movements: A developmental perspective. In *2015 Joint IEEE international conference on development and learning and epigenetic robotics (ICDL-EpiRob)* (pp. 154–161). IEEE.

Ziemke, T., & Lowe, R. (2009). On the role of emotion in embodied cognitive architectures: From organisms to robots. *Cognitive Computation, 1*(1), 104–117.