The Knowledge Level in Cognitive Architectures: Current Limitations and Possible Developments

Antonio Lieto[a,b], Christian Lebiere[c], Alessandro Oltramari[d]

[a] University of Turin, Department of Computer Science, Torino, Italy
[b] ICAR-CNR, Palermo, Italy
[c] Carnegie Mellon University, Department of Psychology, Pittsburgh, USA
[d] Bosch Research and Technology Center, Pittsburgh, USA

Preprint version. Final version in Cognitive Systems Research,
https://doi.org/10.1016/j.cogsys.2017.05.001

Abstract

In this paper we identify and characterize an analysis of two problematic aspects affecting the representational level of cognitive architectures (CAs), namely: the limited *size* and the homogeneous *typology* of the encoded and processed knowledge. We argue that such aspects may constitute not only a technological problem that, in our opinion, should be addressed in order to build artificial agents able to exhibit intelligent behaviours in general scenarios, but also an epistemological one, since they limit the plausibility of the comparison of the CAs' knowledge representation and processing mechanisms with those executed by humans in their everyday activities. In the final part of the paper further directions of research will be explored, trying to address current limitations and future challenges.

Keywords:
Knowledge Representation, Cognitive Architectures, Knowledge Heterogeneity, Knowledge Processing

1. Introduction

Handling a considerable amount of knowledge, and selectively retrieving it according to the needs emerging in different situational scenarios, is an important aspect of human intelligence. For this task, in fact, humans adopt a wide range of heuristics [1] due to their bounded rationality [2]. In this perspective, one of the requirements that should be considered for the design, the realization and the evaluation of intelligent cognitively-inspired systems should consist in their ability to heuristically identify, retrieve, and process, from the general knowledge stored in their artificial Long Term Memory (LTM), that one which is synthetically and contextually relevant. This requirement, however, is often neglected. Currently, artificial cognitive systems and architectures are not able, de facto, to deal with complex knowledge structures that can be even slightly comparable to the knowledge heuristically managed by humans. In this paper we will argue that this is not only a technological problem but also, in the light of the distinction between functionalist and structuralist models of cognition, an epistemological one. The rest of the paper is organised as follows: Section 2 introduces the two main problematic aspects concerning the knowledge level in cognitive architectures, namely the *size* and the *homogeneous typology* of the encoded knowledge. Section 3 provides a focused review of the Knowledge Level of four of the most well known and widely used cognitive architectures (namely SOAR, ACT-R, CLARION and Vector-LIDA) by pointing out the respective differences and, in the light of our axis of analysis, their problematic issues[1]. In doing so we will illustrate the main attempts that have been proposed to address such problems and we will highlight the current limitations of such proposals. In the final sections, we present an overview of three different alternative approaches that can provide a possible solution for dealing with, jointly, both the size and the knowledge homogeneity problems: namely the Semantic Pointer Perspective (section 4), the idea of Conceptual Space as intermediate level of representation connecting connectionist and symbolic approaches (section 5) and the novel versions of the Hybrid Neuro Symbolic Approaches currently developed in the field of CAs (section 6). Interestingly all such proposals converge in suggesting that the neural level of representation can be considered irrelevant for attacking the above mentioned problems, and suggest to address these

[1]In the present paper we will leave aside many other aspects (e.g. those related to the knowledge acquisition problems) which are related to, and also affect, the problems in focus.

issues by operating at more transparent and abstract levels of representation. Section 7, finally, considers the dual process hypothesis as a possible reference framework for the integration of different types of knowledge processing mechanisms assumed to cooperate in a CA adopting a heterogeneous representational perspective. As we will show, the advantages provided by the adoption of this approach are still not completely clear and deserve further investigations.

2. Open Issues: Knowledge Size and Knowledge Homogeneity

Current cognitive artificial systems and architectures are not equipped with knowledge bases comparable with the conceptual knowledge that humans possess and use in the everyday life. From an epistemological perspective this shortcoming represents a problem: in fact, endowing cognitive agents with more realistic knowledge bases, in terms of both the size and the type of information encoded, would allow, at least in principle, to test the artificial systems in situations closer to those encountered by humans in real life. This problem becomes more relevant if we take into account the knowledge level[2] of Cognitive Architectures [4], [5]. While cognitively-inspired systems, in fact, could be designed to deal with only domain-specific information (e.g. a computer simulation of a poker player), Cognitive Architectures (CA), on the other hand, have also the goal and the general objective of testing - computationally - the

[2]The description of the 'knowledge problems' affecting the current Cognitive Architectures (i.e. the knowledge size and the knowledge homogeneity, see below) is provided at the knowledge level in the sense intended by Newell (i.e. we point out that, given the current state of affairs, the rational behavior of a cognitive artificial agent adopting such architectures can be predicted as a limited one simply on the basis of the analysis of the content of its available representations, its limited knowledge of its goals etc.). On the other hand, the possible solutions proposed for dealing with these problems, sketched in the final part of the paper, are focused on what Newell calls the Symbol Level, since they concern the actual information-processing mechanisms that the system uses in order to reach its goal, given the knowledge that possesses. The close relations between these two levels is explained in [3]. According to Newell, this hierarchy of levels (that includes further levels involving the hardware implementation), characterizes the Physical Symbol System architecture [3].

general models of mind they implement. Therefore, if such architectures only process a simplistic amount (and a limited typology) of knowledge, the structural mechanisms that they implement concerning knowledge processing tasks (e.g., retrieval, learning, reasoning etc.) can be only loosely evaluated, and compared, w.r.t. those used by humans in similar knowledge-intensive situations. In other words, from an epistemological perspective, the explanatory power of their computational simulation is strongly affected (on these aspects see [6], [7], [8]). This aspect is problematic since this class of systems, designed according to the "cognition in the loop" approach, aims both at i) detecting novel and hidden aspects of the cognitive theories by building properly designed computational models of cognition and ii) at providing technological advancement in the area of Artificial Intelligence (AI) of cognitive inspiration. In this perspective, purely *functionalist* models [9], based on a weak equivalence (i.e. the equivalence in terms of functional organization) between cognitive processes and AI procedures, are not considered as having a good explanatory power w.r.t. the target cognitive system taken as source of inspiration. Conversely, the development of plausible "structural" models of our cognition (based on a more constrained equivalence between AI procedures and their corresponding cognitive processes) are assumed to be the way to follow in order to build artificial cognitive models able to play both an explanatory role about the theories they implement and to provide advancements in the field of the artificial intelligence research.

By following this line of argument, therefore, we claim that computational cognitive architectures aiming at providing a knowledge level based on the "structuralist" assumption should address, at their representational level, both the problems concerning the limited "size" and "homogeneity" of the encoded knowledge. Let us explore in more details the nature of such aspects: while the size problem is intuitively easy to understand (i.e. it concerns the dimension of the knowledge base available to the agents), the problem concerning the homogeneous typology of the encoded knowledge needs some additional clarification and context. In particular, this problem relies on the theoretical and experimental results coming from Cognitive Science. In this field, different theories about how humans organise, reason and retrieve conceptual information have been proposed. The oldest one, known as "classical" or Aristotelian theory, states that concepts - the building blocks of our knowledge infrastructure - can be simply represented in terms of sets of necessary and sufficient conditions (and this is completely true, for example, for mathematical concepts: e.g. an EQUILATERAL TRIANGLE can be classically defined as a regular polygon with 3 corners and 3 sides). In the mid 1970s, however, Rosch's experimental results demonstrated its inadequacy for ordinary –or *common sense*– concepts, which cannot be described in terms of necessary and sufficient traits [10]. In particular, Rosch's results indicate that conceptual knowledge is organized in our mind in terms of *prototypes*. Since then, different theories of concepts have been proposed to explain different representational and reasoning aspects concerning the *typicality* or, in other terms, the common-sense effects[3]. Usually, they are grouped in three main classes, namely: prototype views, exemplar views and theory-theories (see e.g., [11], [12]). All of them are assumed to account for (some aspects of) typicality effects in conceptualization.

According to the prototype view (introduced by Rosch), knowledge about categories is stored in terms of some representation of the best instances of the category. For example, the concept BIRD should coincide with a representation of a prototypical bird (e.g., a robin). In the simpler versions of this approach, prototypes are represented as (possibly weighted) lists of features.

According to the exemplar view, a given category is mentally represented as a set of specific exemplars explicitly stored in memory: the mental representation of the concept BIRD is the set of the representations of (some of) the birds we encountered during our lifetime.

Theory-theory approaches adopt some form of holistic point of view about concepts. According to some versions of the theory-theories, concepts are analogous to theoretical terms in a scientific theory. For example, the concept BIRD is individuated by the role it plays in our mental theory of zoology. In other ver-

[3] A review of all the typicality-theories can be found in [11] and in [12].

sion of the approach, concepts themselves are identified with micro-theories of some sort. For example, the concept BIRD should be identified with a mentally represented micro-theory about birds.

Although these approaches have been largely considered as competing ones, several results (starting from the work of [13]) suggested that human subjects may use, in different occasions, different representations to categorize concepts. Such experimental evidence led to the development of the so called "heterogeneous hypothesis" about the nature of concepts, hypothesizing that different types of conceptual representations exist (and may co-exist): prototypes, exemplars, theory-like or classical representations, and so on [12]. All such representations, in this view, constitute different *bodies of knowledge* and contain different types of information associated to the the same conceptual entity. Furthermore, each body of conceptual knowledge is distinguished by specific processes in which such representations are involved (e.g., in cognitive tasks like recognition, learning, categorization, *etc.*). In particular prototypes, exemplars and theory-like representations are associated with the possibility of dealing with typicality effects and non-monotonic strategies of reasoning and categorization[4], while the

[4]Let us assume that we have to categorize a stimulus with the following features: "it has fur, woofs and wags its tail" the result of a *prototype-based categorization* would be dog, since these cues are associated to the prototype of dog. Prototype-based reasoning, however, is not the only type of reasoning based on typicality. In fact, if an exemplar corresponding to the stimulus being categorized is available, too, it is acknowledged that humans use to classify it by evaluating its similarity w.r.t. the exemplar, rather than w.r.t. the prototype associated to the underlying concepts [12, 14]. For example, a penguin is rather dissimilar from the prototype of bird. However, if we already know an exemplar of penguin, and if we know that it is an instance of bird, it is easier to classify a new penguin as a bird w.r.t. a categorization process based on the similarity with the prototype of that category. This type of common sense categorization is known in literature as *exemplars-based categorization* (and in this case the exemplar is favoured w.r.t. the prototype because of the phenomenon known as *old-item effect*). Finally, an example of theory-like common sense reasoning is when we typically associate to a light switch the learned rule that if we turn it "on" then the light will be provided (this is a non-monotonic inference with a defeasible conclusion). All these representations, and the corresponding reasoning mechanisms, are assumed to be potentially

classical representations (i.e. those based on necessary and/or sufficient conditions) are associated with standard deductive mechanism of reasoning[5]. In the representational level of the current cognitive architectures the heterogeneity hypothesis, assuming the availability of different types of knowledge encoded in a conceptual structure, is almost neglected [6] (even if some differentiations between the architectures exist, as we will see in the next section)[7]. In general, despite some efforts have been done to implicitly address the presented problems they are, as we will show below, not completely satisfactory for solving, jointly, both the mentioned limitations.

3. The Knowledge Level in Cognitive Architectures

Cognitive architectures have been historically introduced i) to capture, at the computational level, the invariant mechanisms of human cognition, including those underlying the functions of control, learning, memory, adaptivity, perception and action [16] and ii) to reach human level general intelligence, by means of the realization of artificial artifacts built upon them[8]. During the last decades many cognitive architectures have been realized, - such as SOAR [18], ACT-R [19] etc. - and have been widely tested in several cognitive tasks involving learning, reasoning, selective attention, recognition etc. However, as previously mentioned, they are affected by the following problem: they are gen-

co-existing according to the heterogeneity approach.

[5]As before mentioned, an example of standard deductive reasoning is the categorization as *triangle* of a stimulus described by the features: "it is a polygon, it has three corners and three sides". Such cues, in fact, are necessary and sufficient for the definition of the concept of triangle.

[6]There are, however, some proposals going in this direction. See e.g. [15].

[7]The heterogeneity problem is a multifaceted one since, as mentioned, it not only assumes the existence of multiple representations but, for each of them, different kinds of categorization and reasoning mechanisms and processes are assumed to exist and need to be integrated in order to let intelligent behaviour emerge.

[8]There is an alternative perspective that sees CAs as the initial point of departure for the subsequent autonomous development of a cognitive system (see [17]).

eral structures without a general content. Thus, every evaluation of systems relying upon them is necessarily task-specific and do not involve even the minimum part of the full spectrum of processes involved in human cognition when the knowledge comes to play a role. In more practical terms this means that the knowledge embedded in such architectures, and processed by artificial agents, is usually built ad hoc, domain-specific, or based on the particular tasks they have to deal with. Such limitation, however, affects the advancement in the cognitive research concerning how humans heuristically select and deal with the huge and varied amount of knowledge they possess when they have to make decisions, reason about a given situation or, more generally, solve a particular cognitive task involving several dimensions of analysis. This problem, as a consequence, also limits the advancement of the research in the area of General Artificial Intelligence. In the following we provide a short overview of some of the most widely known and adopted CAs: SOAR [18], ACT-R [19], CLARION [20] and LIDA [21] (in its novel version known as Vector-LIDA [22]). The choice of these architectures has been based on the fact that they represent some of the most widely used systems (adopted in scenarios ranging from robotics to videogames) and their representational structures present some relevant differentiations that are interesting to investigate in the light of the issues raised in this paper. By analyzing, in brief, such architectures we will exclusively focus on the description of their representational frameworks since a more comprehensive review of their whole mechanisms is out of the scope of the present contribution (detailed reviews of their mechanisms are described in [23], [24], and [25]). We will show how they are all affected, at different levels of granularity, by both the size and the knowledge homogeneity problems [9].

[9] By analyzing the latter aspect we will not take into direct consideration the theory-like representations introduced in the previous section, since the corresponding theory-theory approach is, to a certain extent, more vaguely defined when compared to both prototypes and exemplar-based approaches. As a consequence, at present, its theoretical and computational treatment seems to be more problematic. In addition we can take for granted that all the currently available architectures are able to learn some forms of ad-hoc micro-theories according

3.1. SOAR

SOAR (see figure 1 below) is one of the most mature cognitive architectures and has been used by many researchers worldwide during the last 30 years. This system was considered by Newell a candidate for a Unified Theory of Cognition [5]. One of the main themes in SOAR is that all cognitive tasks can be represented by problem spaces that are searched by production rules grouped into operators. These production rules are fired in parallel to produce reasoning cycles. From a representational perspective, SOAR exploits symbolic representations of knowledge (called chunks) and use pattern matching to select relevant knowledge elements. Basically, where a production match the contents of declarative (working) memory the rule fires and then the content from the declarative memory (called Semantic Memory in SOAR) is retrieved. This system adheres strictly to Newell and Simon's physical symbol system hypothesis [26] which states that symbolic processing is a necessary and sufficient condition for intelligent behavior.

With respect to the size and the heterogeneity problems, the SOAR knowledge level is problematic for both aspects. SOAR agents, in fact, are not endowed with general knowledge and only process ad-hoc built (or task-specific learned) symbolic knowledge structures [10]. Such type of knowledge structures, in particular, are usually heavily used to perform standard logical reasoning and, as a consequence, are strongly biased towards a "classical" conceptualisation of knowledge in terms of necessary or sufficient conditions. In general, symbolic

to their interaction with the external environment. A general objection that can be raised to all of them is, however, that they are not architecturally equipped with mechanisms able to define the dynamics of the interaction between this kind of theory-like typical knowledge and the other common-sense knowledge components (e.g. prototypes or exemplars). In addition, such theories are local, they have no generality and the current CAs are not designed to provide any kind of interaction process able to couple different local and possibly contrasting micro-theories.

[10] Despite this problem being acknowledged in [18] there is no available literature attesting progress in this respect.

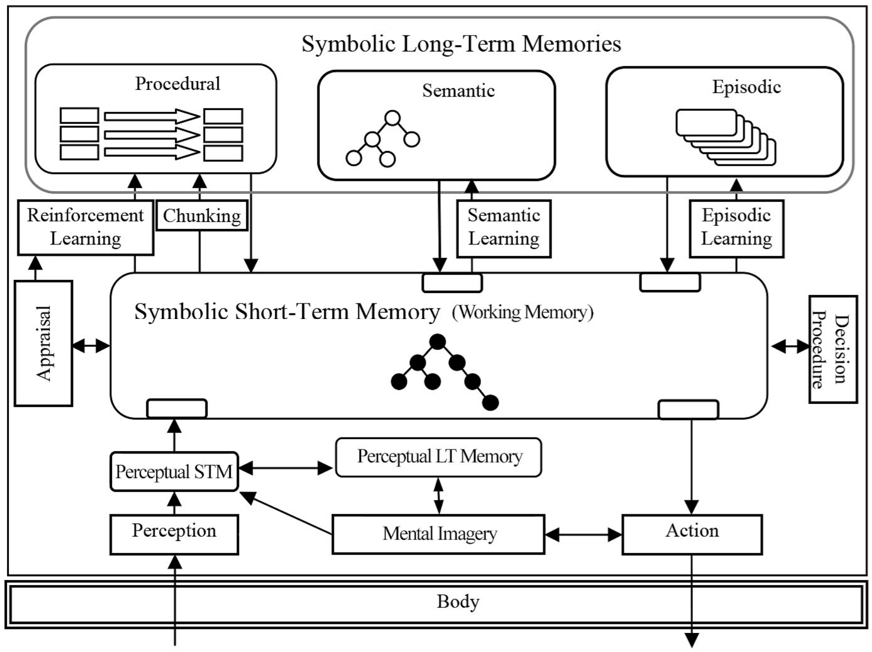

Figure 1: The SOAR Cognitive Architecture. The semantic memory module of the architecture represents the information by using symbolic chunks (graph-like structures). The representational assumption of the architecture is entirely symbolic.

representations strongly rely on the compositionality of meaning: where we can distinguish between a set of primitive, or atomic, symbols and a set of complex symbols. However, compositionality, despite being an important aspect of human conceptual systems, is somewhat at odds with the representation of concepts regarding typicality [27]. As a consequence of this representational commitment, SOAR agents are not equipped with common-sense knowledge components concerning, for example, prototypical or exemplar-based representations[11]. Therefore the system is not able to deal with prototype and exemplar-

[11]And this problem arises despite the fact that the chunks in SOAR can be represented as a sort of frame-like structures containing some common-sense (e.g. prototypical) information. In fact, the main problem of this architecture w.r.t. the heterogeneity assumption relies on the fact that it does not specify how the typical knowledge components of a concept, and the corresponding non monotonic-reasoning strategy, can interact with a possibly conflict-

based categorization which, as described above, are two forms of common-sense conceptual reasoning well established in human cognition and assumed to co-exist in the heterogeneous perspective.

3.2. ACT-R

ACT-R (see figure 2 below) is a cognitive architecture explicitly inspired by theories and experimental results coming from human cognition. Here the cognitive mechanisms concerning the knowledge level emerge from the interaction of two types of knowledge: declarative knowledge, which encodes explicit facts that the system knows, and procedural knowledge, which encodes rules for processing declarative knowledge. In particular, the declarative module is used to store and retrieve pieces of information (called chunks, composed of a type and a set of attribute-value pairs, similar to frame slots) in declarative memory. ACT-R employs a subsymbolic activation of symbolic conceptual chunks representing the encoded knowledge. Finally, the central production system connects these modules by using a set of IF-THEN production rules.

Differently from SOAR, ACT-R allows to represent the information in terms of prototypes and exemplars and allow to perform, selectively, either prototype or exemplar-based categorization. This means that the architecture allows the modeller to manually specify which kind of categorization strategy to employ according to his specific needs. Such an architecture, however, only partially addresses the homogeneity problem since it does not allow to represent, jointly, these different types of common-sense representations conveying different types of information for the same conceptual entity (i.e. it does not assume a heterogeneous perspective). As a consequence, it is also not able to autonomously

---

ing representational and reasoning procedures characterizing other conceptualisation of the same conceptual entity. In short it assumes, like most of the symbolic-oriented CAs, the availability of a monolithic conceptual structure (e.g., a frame-like prototype or a “classical” concept) without specifying how such information can be integrated and harmonized with other knowledge components to form the whole knowledge spectrum characterizing a given concept.

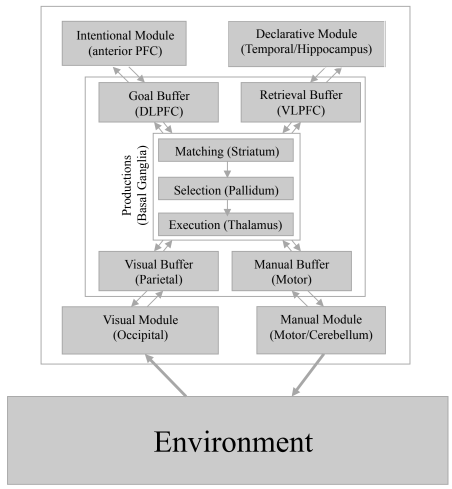

Figure 2: The ACT-R Cognitive Architecture. The declarative memory module of the architecture represent the information by using symbolic chunks (graph-like structures). The representational assumption of the architecture is a combination of symbolic and subsymbolic.

decide which of the corresponding reasoning procedures to activate (e.g. prototypes or exemplars) and to provide a framework able to manage the interaction of such different reasoning strategies (however its overall architectural environment provides, at least in principle, the possibility of implementing cascade reasoning processes triggering one another).

Even if some attempts exist concerning the design of harmonization strategies between different types of common-sense conceptual categorizations (e.g. exemplar-based and rule-based, see [28]) however they do not handle the problem concerning the interaction of the prototype or exemplar-based processes according to the results coming from experimental cognitive science (for example: the old item effect, privileging exemplars w.r.t. prototypes is not modeled. See footnote 3 on this aspect.). Summing up: w.r.t. the heterogeneity problem, the components needed to fully reconcile the Heterogeneity approach with ACT-R are available, however they have not been fully exploited yet.

Regarding the size problem: as for SOAR, ACT-R agents are usually equipped with task-specific knowledge and not with general cross-domain knowledge. In this respect some relevant attempts to overcome this limitation have been recently done by extending the Declarative Memory of the architecture. They will be discussed in section 3.5 along with their current implications.

3.3. CLARION

CLARION is a hybrid cognitive architecture based on the dual-process theory of mind. From a representational perspective, processes are mainly subject to the activity of two sub-systems, the Action Centered Sub-system (ACS) and the Non-Action Centered Sub-system (NACS). Both sub-systems store information using a two-layered architecture, i.e., they both include an *explicit* and an *implicit* level of representation. Each top-level chunk node is represented by a set of (micro)features in the bottom level (i.e., a distributed representation). The (micro)features (in the bottom level) are connected to the chunk nodes (in the top level) so that they can be activated together through bottom-up or top-down activation. Therefore, in general, a chunk is represented by both levels: using a chunk node at the top level and distributed feature representation at the bottom level.

W.r.t. to the size and the heterogeneity problems, CLARION encounters problems with both levels since i) there are no available attempts aiming at endowing such architecture with a general and cross-domain knowledge ii) the dual-layered conceptual information does not provide the possibility of encoding (manually or automatically via learning cycles) the information in terms of the heterogeneous classes of representations presented in section 2. In particular, the main problematic aspect concerns the representation of the common-sense knowledge components. As for SOAR and ACT-R, in CLARION the possible co-existence of typical representations in terms of prototypes, exemplars and theories (and the interaction among them) is not treated. In terms of reasoning

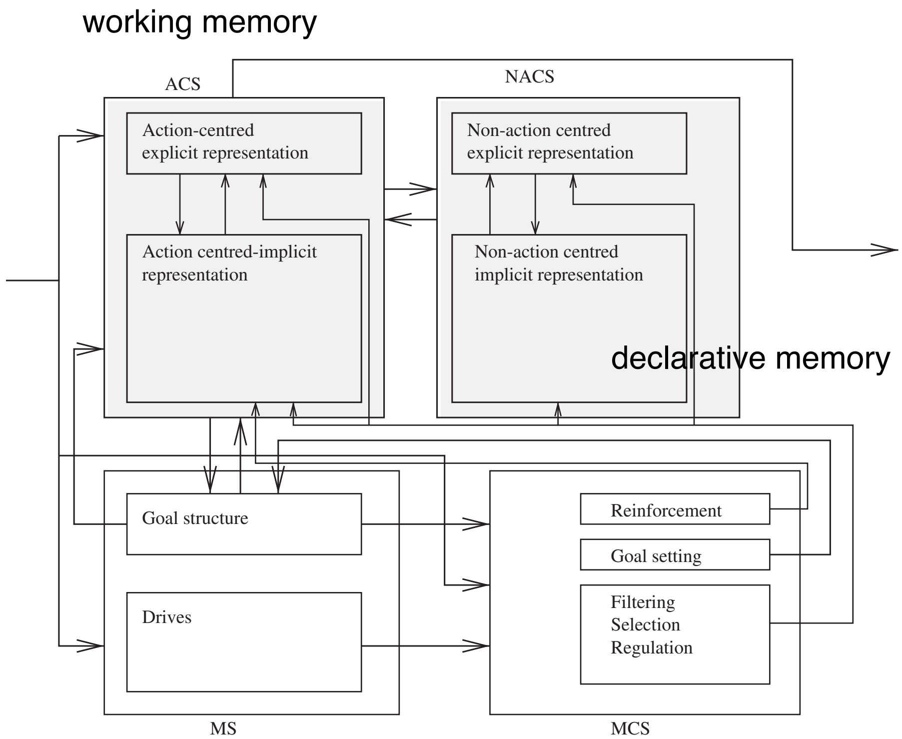

Figure 3: The CLARION Cognitive Architecture (adapted from [20]).

strategies, notwithstanding that the implicit knowledge layer, based on neural network representations, can provide forms of non-monotonic reasoning (e.g., based on similarity), such kind of similarity-based reasoning is currently not grounded in the mechanisms guiding the decision choices followed, for example, by prototype or exemplar-based reasoning.

3.4. Vector-LIDA

Vector-LIDA is novel version of the LIDA cognitive architecture (depicted in figure 3) employing, at the representational level, high-dimensional vectors and reduced descriptions.

High-dimensional vector spaces have interesting properties that make them attractive for representations in cognitive models. The distribution of the distances between vectors in these spaces, and the huge number of possible vectors,

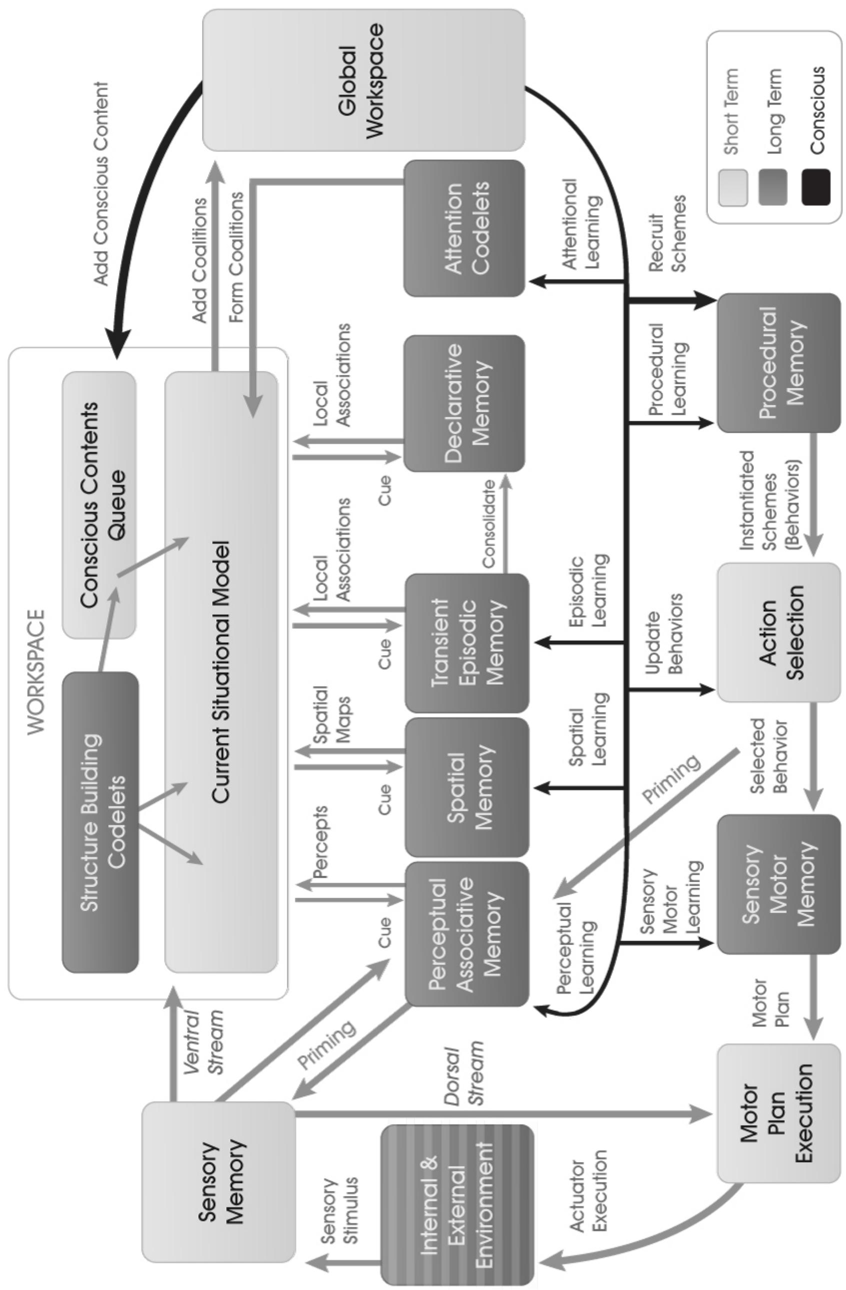

Figure 4: The LIDA Cognitive Architecture. The novel version of the architecture employs high-dimensional vector spaces to represent knowledge in its Declarative Memory module.

allow noise-robust representations where the distance between vectors can be used to measure the similarity (or dissimilarity) of the concepts they represent. Moreover, these high-dimensional vectors can be used to represent complex structures, where each vector denotes an element in the structure. However, a single vector can also represent one of these same complex structures in its entirety by implementing a reduced description, a mechanism to encode complex hierarchical structures in vectors or connectionist models. These reduced description vectors can be expanded to obtain the whole structure, and can be used directly for complex calculations and procedures, such as making analogies, logical inference, or structural comparison. Vectors in this framework are treated as symbol-like representations, thus enabling different kind of operations executed on them (e.g. simple forms of compositionality via vectors blending). Vector-LIDA encounters the same limitations of the other CAs since i) its agents are not equipped with a general cross-domain knowledge and therefore can be only used in very narrow tasks (their knowledge structure is either built or learned ad hoc). Additionally, this architecture does not address the problem concerning the heterogeneity of the knowledge typologies. In particular its knowledge level does not represent the common-sense knowledge components such as prototypes and exemplars (and the related reasoning strategies). In fact, as for CLARION, despite vector representations allowing to perform many kind of approximate comparisons and similarity-based reasoning (e.g., in tasks such as categorization), the functionalities concerning prototype-based or exemplar-based representations (along with the design of the interaction between their different reasoning strategies) are not provided [12].

[12]In this respect, however an element that is worth-noting is represented by the fact that the Vector-LIDA representational structures are very close to the framework of Conceptual Spaces. Conceptual Spaces are a geometric knowledge representation framework proposed by Peter Gärdenfors. They can be thought as a particular class of vector representations where knowledge is represented as a set of *quality dimensions*, and where a geometrical structure is associated to each quality dimension. They are discussed in detail in section 5. The convergence of the Vector-LIDA representation towards Conceptual Spaces could enable, in such

3.5. Attempts to Overcome the Knowledge Limits

The problem concerning the limited knowledge availability for agents endowed with cognitive architectures has been recently pointed out [16] and some technical solutions for filling this knowledge gap have been proposed. In particular the use of ontologies and of semantic formalisms has been seen as a possible solution for providing effective content to the structural knowledge modules of the cognitive architectures. Some initial efforts have been done in this direction [13]. In particular, within the 'Mind's Eye' program (a 2010-2012 DARPA-funded project), the knowledge layers of the ACT-R architecture have been semantically extended with external ontological content coming from three integrated semantic resources composed of the lexical databases WordNet [29], FrameNet [30] and by a branch of the top level ontology DOLCE [31] related to event modeling. In this case, the amount of semantic knowledge selected for the realization of the Cognitive Engine (one of the systems developed within the 'Mind's Eye' program) and for its evaluation, despite being far larger w.r.t. the standard ad-hoc solutions, was tailored to the specific needs of the system itself. In fact, it was aimed at solving a precise task of event recognition through a video analysis intelligent process; therefore only the ontological knowledge about the events was selectively embedded in it. While this is a reasonable approach in an applicative context, it still does not allow to test the general cognitive mechanisms of a Cognitive Architecture on a general, multi-faceted and multi-domain, knowledge. Therefore it does not allow to evaluate *strictu sensu* to what extent the designed heuristics allowing to retrieve and process, from a massive and composite knowledge base, conceptual knowledge can be considered *satisfying* w.r.t. human performance.

architecture, the possibility of dealing with at least prototype and exemplars-based representations and reasoning, thus overcoming the knowledge homogeneity problem.

[13] It is worth noting that most of the attempts have been performed in ACT-R that seems to be currently the available CA paying more attention to carefully constrain its knowledge infrastructure to the insights coming from the results of experimental cognitive science w.r.t. the representation and reasoning procedures operating at the knowledge level.

More recent works have tried to completely overcome at least the size problem of the knowledge level. To this class of works belongs that one proposed by Salvucci [32] aiming at enriching the knowledge model of the Declarative Memory of ACT-R with a world-level knowledge base such as DBpedia (i.e. the semantic version of Wikipedia represented in terms of ontological formalisms) and a previous one proposed in [33] presenting an integration of the ACT-R Declarative and Procedural Memory with the Cyc ontology [34] (one of the widest ontological resources currently available containing more than 230,000 concepts). Both wide-coverage integrated ontological resources, however, represent conceptual information in terms of symbolic structures and encounter the standard problems affecting this class of formalisms: i) they are not well equipped to deal with common-sense knowledge representation and reasoning (since approximate comparisons are hard and computationally intensive to implement with graph-like representations), and ii) the typology of encoded knowledge is biased towards the "classical" (but unsatisfactory) representation of concepts in terms of necessary and sufficient conditions ([27]). In other terms, these ontology-based systems, if considered in isolation, only allow *de facto*, to represent (and reason on) one part of the whole spectrum of conceptual information[14]. On the other hand, the so-called common-sense knowledge components (i.e., those that allow to characterize and process conceptual information in terms of prototypes, exemplars or theories and described above) are largely absent. Common-sense conceptual knowledge, however, is exactly the type of cognitive information cru-

[14] In concrete applications, in fact, the information usually taken into account by adopting ontological knowledge resources is that concerning the taxonomical relations between concepts since it is based on necessary and sufficient conditions and allows to perform efficiently forms of automatic monotonic reasoning. The remaining common-sense characterization of concepts are not modeled since, despite in the field of logic-oriented KR various fuzzy and non-monotonic extensions of DL formalisms have been designed to deal with such aspects [35], various theoretical and practical problems remain unsolved and, in general, an acceptable KR framework able to provide a practically usable trade-off regarding language expressivity and complexity has not yet been achieved [27].

cially used by humans for heuristic reasoning and decision making and therefore represents a necessary aspect to be integrated in CAs aiming at providing an explanatory role of some sort in the science of mind.

It is worth noting that, at least in principle, some of the described limitations are intended to be overcome by the above-mentioned works, since the integration of such wide-coverage ontological knowledge bases with the ACT-R Declarative Memory allows to preserve the possibility of using the common-sense conceptual processing mechanisms available in that architecture (e.g. prototype and exemplars based categorization). Still, however, the main problem faced by these external integrations remains: the lack of the representation of common-sense information to which such common-sense architectural processes can be applied. For instance, a conceptual retrieval based on prototypical traits (i.e. a prototype-based categorization, see footnote 3) simply cannot be performed on such integrated ontological knowledge bases if these symbolic systems do not represent at all the typical information associated to a given concept (and, as we will see in more detail in section 7, this phenomenon is largely majoritarian). In addition, as mentioned in section 3.2, the problem concerning the interaction, in a general and principled way, of the different types of common-sense processes involving different representations of the same conceptual entity remains open.

In the light of the elements presented above it can be argued, therefore, that the current proposed solutions for dealing with the knowledge problems in CAs are not completely satisfactory. In particular, the integration with huge world-level ontological knowledge bases can be considered a necessary solution for solving the size problem. It is, however, insufficient for dealing with the knowledge homogeneity problem and with the integration of the common-sense conceptual mechanisms activated on heterogeneous bodies of knowledge, as assumed in the heterogeneous representational perspective.

In the next sections we outline three possible alternative solutions that, despite being not yet fully developed are, in perspective, suitable to account for both for the heterogeneous aspects in conceptualization and for the size problems. In doing so we will outline how they share the same insights about the neural level of representation (adopted in most CAs because of its efficacy in perceptual-based tasks). Namely, such approaches converge on the idea that the problems affecting the knowledge level can be better addressed by focusing on more abstract levels of representations w.r.t that available in neural networks. In this perspective, the interesting aspect concerning neural representations consists in the definition and development of transformation methods allowing to pass from low-level representations to more abstract ones. As we will show, some methods going in this direction already exist and have been successfully employed in the area of computational cognitive science in systems aiming at providing a reconciled and unified view of the theories of concepts based on prototypes, exemplars, and theory-like structures.

4. Semantic Pointers

The Semantic Pointers approach is a representational perspective currently investigated in the context of the biologically inspired SPAUN architecture [36]. A Semantic Pointers architecture sees concepts as symbol-like vectorial representations that result from different kinds of transformation processes of low-level neural representations in further high-level representations that function to support cognitive processes like categorization, inference, and language use. The core idea behind this approach is that the activity of a population of neurons at any given time can be interpreted as representing a vector.

The SPAUN architecture, assuming this perspective, has been successfully used to replicate three paradigmatic categorization studies concerning prototype-based categorization, exemplar-based categorization and theory-theory based categorization [37]. Such results show that the provided representational approach can account for different kinds of categorization processes assumed in the heterogeneous perspective. However, it is agnostic w.r.t. *how* such processes interact with each other in the case of multiple available representations for the same conceptual entity. From the size perspective, on the other hand, this approach has been currently exploited for representing the human-sized lexical knowledge structured in the Wordnet taxonomy in terms of biologically plausible and scalable neural network representations [38].

In this approach, the interpretation of neural representations as vectors is obtained through different kind of transformation operations, namely: circular convolution, vector addition and involution [38]. In the circular convolution operation two input neural populations, each representing a vector, are connected to an intermediary population that projects to an output population a vector that is the convolution of the two input vectors. The Vector addition operation plays, on the other hand, the role of a superposition operator. In particular, it allows multiple bindings to be stored in a single vector. Finally, the vector involution operation is an approximate inverse with respect to the circular convolution. As reported by the authors 'the circular convolution, vector addition, and involution operations can be thought of as vector analogs of the familiar algebraic operations of multiplication, addition, and taking the reciprocal, respectively' [38]. In this sense, the Semantic Pointer perspective seems to provide an effective set of operational tools to proceed from a lower level of representation to another, more abstract, one.

Summing up: concerning the size problem, as mentioned, this approach has proven to be usable to neurally represent human-level lexical knowledge. On the other hand, i.e. w.r.t. the heterogeneity problem, it represents a more powerful, but still incomplete, account of the common-sense typicality-based processes executable on conceptual representations. In particular, the framework has been proven able to replicate the full spectrum of typicality effects studied in human cognition including (and differently from the other proposals reviewed) the theory-theory approach [15]. However it still does not provide

[15] In this respect it is worth noting that the methods employed by the Semantic Pointers Architecture to provide an abstract interpretation of neural mechanisms and representation are completely compatible (and integrable) with some mechanisms provided by cognitive architectures dealing with the neural representations. For example: they can be easily reproduced within the subsymbolic activation mechanisms of a cognitive architecture such as ACT-R.

any account concerning the dynamics of interaction and harmonization of the plethora of processes involving the conceptual representations assumed to co-exist according to the heterogeneous hypothesis. Therefore, in this sense, the same objection raised for the current state of development of the knowledge level of the standard CAs remains unanswered. As a consequence, the explanatory power of the Semantic Pointer Architecture w.r.t. the cognitive theories and the experimental results that it is able to replicate is currently very limited (since the replication of such categorization experiments did not lead to any kind of additional explanation of these, already known, phenomena). This aspect represents a symptom that, in order to account for the interaction of the heterogeneous mechanisms operating over different, but interlinked, representations the focus on the neural level is, in some sense, unnecessary and can be switched to other classes of representations having the advantage of being less opaque.

5. Conceptual Spaces as Intermediate Level

Conceptual Spaces [39] have been proposed by Peter Gärdenfors as an intermediate level of representation between the subsymbolic and symbolic levels. It has been argued that the integration of this level enables to overcome some classical problems specifically related to the subsymbolic and symbolic representations considered in isolation [40]. Conceptual Spaces are a geometrical framework for the representation of knowledge[16] and can be thought as a *metric* space in which entities are characterized by quality dimensions [39]. To each quality dimension is associated a geometrical (topological or metrical) structure. In some cases, such dimensions can be directly related to perceptual mechanisms; examples of this kind are temperature, weight, brightness, pitch. In other cases, dimensions can be more abstract in nature. In this setting,

[16] In the last fifteen years, such framework has been employed in a vast range of AI applications spanning from visual perception [41] to robotics [42], from question answering [43] to music perception [44] (see [45] for a recent overview).

concepts correspond to convex regions, and regions with different geometrical properties correspond to different sorts of concepts [39]. Here, prototypes and prototypical reasoning have a natural geometrical interpretation: prototypes correspond to the geometrical center of a convex region (the centroid). Also exemplar-based representation can be represented as points in a multidimensional space, and their similarity can be computed as the intervening distance between two points, based on some suitable metrics (such as Euclidean and Manhattan distance etc.).

Recently some available conceptual categorization systems, explicitly assuming the heterogeneous representational hypothesis and coupling Conceptual Spaces representations and ontological knowledge bases, have been developed. For our purposes, we will consider here the DUAL PECCS system [46]: that system has been integrated with available CAs by explicitly designing the flow of interaction *between* the common-sense, non-monotonic, categorization strategies (based on prototypes and exemplars and operating on conceptual spaces representations) and the standard deductive processes (operating on the ontological conceptual component). The harmonization regarding such different classes of mechanisms has been devised based on the tenets coming from the dual process theory of reasoning (see section 7). Additionally, in this system, the flow of interaction occurring *within* the class of non monotonic categorization mechanisms (i.e. prototypes and exemplar-based categorization) has been devised and is dealt with at the Conceptual Spaces level. This latter aspect is of particular interest in light of the problem concerning the heterogeneity of the encoded knowledge. In fact, since the design of the interaction of the different processes operating with heterogeneous representations still represents a largely ignored problem in current CAs, this system shows that Conceptual Spaces represent a relatively effortless framework to model the dynamics between prototype and exemplar-based processes.

Concerning the size problem, the possible grounding of the Conceptual Spaces representational framework with symbolic structures enables the integration with wide-coverage knowledge bases such CYC (as provided, for example, in DUAL PECCS [46]), DBpedia or similar.

An additional element of interest concerning the advantages provided by introducing the adoption of Conceptual Spaces as intermediate representational level in CAs regards its capability to address a classical problem in formal conceptualization: namely the problem of reconciling compositionally and typicality effects (for more details on this issue we remind to [47]) [17]. This aspect does not affect, per se, the size problem but the problem concerning the knowledge heterogeneity (since it assumes the existence of typicality-based representations) and has been shown to be problematic for symbolic/logic-oriented approaches [49]) as well as, according to the well-known argument by Fodor and Phylishin [50], for classical connectionist approaches. On the other hand this aspect can be formally handled by recurring to Conceptual Spaces (as shown in [47, 51]). Interestingly enough, this problem can also be treated by the Semantic Pointers perspective (once, in this framework, the more abstract level of representation is reached through the transformation operations mentioned above). The similarity between the two approaches is another indirect suggestion that the neural level of representation, per se, can be considered not directly necessary to deal with the problematic aspects affecting the conceptual representation and processing capabilities in CAs.

[17] Broadly speaking this aspect regards the problem of dealing, in a coherent way, with the compositionality of prototypical representations. According to a well-known argument ([48]; [49]), prototypes are not compositional. In brief, the argument runs as follows: consider a concept like *pet fish*. It results from the composition of the concept *pet* and the concept *fish*. However, the prototype of *pet fish* cannot result from the composition of the prototypes of a pet and a fish: a typical pet is furry and warm, a typical fish is grayish, but a typical pet fish is neither furry and warm nor grayish. The possibility of explaining, in a coherent way, this type of combinatorial and generative phenomenon highlights a crucial aspect of the conceptual processing capabilities in human cognition and concerns some crucial high-level cognitive abilities such as that ones concerning conceptual composition, metaphor generation and creative thinking. Dealing with this problem requires the harmonization of two conflicting requirements in representational systems: the need of syntactic, generative, compositionality (typical of logical systems) and that one concerning the exhibition of typicality effects.

Summing up: endowing CAs with Conceptual Spaces seems, in principle, a promising way to deal with both the size and the heterogeneity problems of conceptual representations. There is, however, still an open problem to explicitly face for such approach. In particular, concerning the size issue, there is still a lack of knowledge bases encoded in terms of Conceptual Spaces comparable with the sizes of the ontological KB. Some initial attempts to automatically learn and encode wide-coverage Conceptual Spaces knowledge bases by starting by linguistic resources such as BabelNet[18] and ConceptNet[19] have been developed [52, 53], but still there is a huge gap to cover and this aspect requires further investigations.

6. Neural-Symbolic Integrations and Extended Declarative Memories

As mentioned in section 3.5. different efforts have been attempted to implicitly address the size and the knowledge heterogeneity problems in CAs. Notably such efforts, that share the same limitations and possibilities as the others, have been developed within an architecture such as ACT-R that presents a hybrid approach to conceptual representation and reasoning combining subsymbolic-based activation mechanisms, operating on classical symbolic structures, and rule-based representational structures (see section 3.2). Since the current state of the art and the possible future developments of ACT-R have already been discussed in section 3.5, we focus here on showing how its underlying assumptions are compliant with both the Semantic Pointers Perspective, and with the approach claiming the advantages that an intermediate Conceptual Spaces representation connecting subsymbolic and symbolic levels would provide. With respect to the first approach, in particular, it has been showed how the integration of ACT-R with a connectionist architecture allows to learn without supervision associations in object recognition between percepts and categorical labels. [54]. The way in which such elements are integrated is fully compliant

[18] http://babelnet.org/
[19] http://conceptnet5.media.mit.edu/

with the Semantic Pointer perspective and is based on the shared assumption that leveraging and abstracting more high-level forms of representation is a key element to produce advances that cannot be achieved by operating exclusively at the neural level. With respect to the Conceptual spaces approach, on the other hand, the neuro-symbolic integration allows to deal with the classical problem concerning the need for reconciling compositionally and typicality effects in conceptualization. The approach developed in ACT-R, in fact, belongs to the class of so-called *neo-connectionist* approaches that, unlike the classical connectionist systems, are able to deal with limited forms of compositionality in neural networks (see [55] on this point).

Interestingly enough, there are also attempts that have shown how the neuro-symbolic approach adopted by ACT-R can be used as an intermediate functional level in a complex system combining different Cognitive Architectures, such as SODAS [56] and SOAR, that can perform different cognitive tasks (e.g. the high level symbolic and knowledge-drive reasoning in SOAR and the low-level perceptual one to SODAS) that are more naturally dealt with in different environments [57]. This idea is somehow similar to that of using Conceptual Spaces as an intermediate level of representation: from a knowledge processing perspective, the types of tasks that such hybrid architecture is able to account are essentially the same.

ACT-R can also generalize over perceptual transductions by applying fine-grained models of the world to concrete scenarios. As already discussed, in order to fulfill this goal, ACT-R needs to properly encapsulate those models – or *ontologies* – and exploit them for pattern recognition and high-level reasoning. Since the ACT-R declarative module supports a relatively coarse-grained semantics based on slot-value pairs, and the procedural system is not optimal to effectively manage complex logical constructs (e.g., 2^{nd} order), specific extensions have been designed to make ACT-R suitable to fulfill knowledge-intensive tasks. Accordingly, the work outlined in [16] proposed an expansion of ACT-R with SCONE [58]. SCONE is an open-source knowledge-base system intended for use as a component in many different software applications: it provides a

LISP-based framework to represent and reason over symbolic common sense knowledge. Unlike most diffuse KB systems (e.g. ontologies), SCONE is not based on Description Logics [59]: its inference engine adopts marker-passing algorithms [58] (originally designed for massive parallel computing) to perform fast queries at the cost of losing logical completeness and decidability. In particular, SCONE represents knowledge as a *semantic network* whose nodes are locally weighted (*marked*) and associated to arcs (*wires*[20]) in order to optimize basic reasoning tasks (e.g. class membership, transitivity, inheritance of properties, etc). The philosophy that inspired SCONE is straightforward: from vision to speech, humans exploit the brain's massive parallelism to fulfill all recognition tasks; similarly, if we want to build a general AI system that is able to deal with the large amount of knowledge required in common-sense reasoning, we need to rely on a mechanism that is fast and effective enough to simulate parallel search. Shortcomings are not an issue since humans are not perfect inference engines either. Accordingly, SCONE implementation of marker-passing algorithms aims at simulating a pseudo-parallel search by assigning specific marker bits to each knowledge unit. For example, if we query a KB to get all the parts of cars, SCONE would assign a marker M1 to the A-node CAR and search for all the statements in the knowledge base where M1 is the A-wire (domain) of the relation PART-OF, returning all the classes in the range of the relation (also called 'B-nodes'). SCONE would finally assign the marker bit M2 to all B-nodes, also retrieving all the inherited subclasses[21]. The modularization and implementation of an ontology with SCONE allows for an effective formal representation and inferencing of core ontological properties of world entities. Note that the integration of SCONE into ACT-R respects the requirements of the cognitive architecture, especially in terms of limited-capacity buffers constraining the communication between a dedicated SCONE module and ACT-R's default

[20]In general, a *wire* can be conceived as a binary relation whose domain and range are referred to, respectively, as A-node and B-node.

[21]We refer the reader to [58] for details concerning marker-passing algorithms.

modules. Also, the SCONE marker-passing algorithms are comparable to ACT-R spreading activation, leaving open the possibility of a deeper integration of the two frameworks in future work. The integration of ACT-R with SCONE represents, in other words, a suitable way to connect architectural mechanisms to a symbolic knowledge base. With respect to the external extensions provided with wide-coverage KBs (and discussed in section 3.5), however, such approach still needs to face the problem concerning the size aspect (since the SCONE KBs are not comparable with Cyc or DBpedia). Concerning the heterogeneity problem, on the other hand, such integration seems to provide a straightforward way to combine common-sense reasoning operating with symbolic knowledge structures. Still, however, the problem concerning the integration of heterogeneous processes acting on different bodies of knowledge is not addressed.

Summing up: all the approaches presented in the sections 4, 5 and 6 can be seen as alternative – and yet compliant – solutions in order to develop a more comprehensive (and constrained to human cognition) account of conceptual representation and processing mechanism in Cognitive Architectures.

A further axis that we will consider as an element for dealing with the heterogeneity issues, in the context of CAs endowed with wide-coverage knowledge bases, is represented by the so-called dual process hypothesis of reasoning and rationality. This aspect will be considered in the next section.

7. A Dual Process Approach for the Heterogeneous Integration of Cognitive Mechanisms

The approaches presented in the previous sections converge on the insight that designing the interaction (and integration) of the heterogeneous processes operating with different representations (i.e. the heterogeneity problem) can be made more effective by operating at more abstract levels of representation than the one proposed by connectionist representations.

In our opinion an additional element that is worth considering to determine, at the architectural level, the interaction strategies between different types of mechanisms operating on heterogeneous representations corresponds to the dual process hypothesis of reasoning and rationality. According to *dual process* theories ([60], [61], [62]) two different types of cognitive processes and systems exist, which have been called respectively *System(s) 1* and *System(s) 2*.

System 1 processes are automatic. They are phylogenetically older and shared by humans and other animal species. They are innate and control instinctive behaviors, so they do not depend on training or particular individual abilities and, in general, they are cognitively undemanding. They are associative and operate in a parallel and fast way. Moreover, *System 1* processes are not consciously accessible to the subject.

System 2 processes are phylogenetically recent and are peculiar to the human species. They are conscious and cognitively penetrable (i.e. accessible to consciousness) and based on explicit rule following. As a consequence, if compared to *System 1*, *System 2* processes are sequential and slower, and cognitively demanding. Performances that depend on *System 2* processes are usually affected by acquired skills and differences in individual capabilities.

The dual process approach was initially proposed to account for systematic errors in reasoning. Such errors (consider, e.g., the classical examples of the selection task or the conjunction fallacy) should be ascribed to fast, associative and automatic *type 1* processes, while *type 2* is responsible for the slow and cognitively demanding activity of producing answers that are correct concerning the canons of normative rationality.

In general, many aspects concerning the psychology of concepts have presumably to do with fast, type 1 system and processes, while others can be plausibly ascribed to type 2. In particular, the ability to make explicit, high-level inferences involving conceptual knowledge, and capacity to justify them, can be considered as a type 2 process while the common-sense mechanisms operating with typical representations (e.g. prototype, exemplars or theory-based categorization) can be considered type 1 processes.

A possible way to evaluate the importance of dual process strategies in knowledge processing can be provided by testing to what extent an AI system designed with this perspective and endowed with a common knowledge base "has the methods for making obvious inferences from this knowledge" [63]. Such common-sense based evaluation task is known to be one of the grand challenges of AI and Cognitive Modeling in general [64]. In so doing we can account for the importance of the dual process approach by analyzing the results obtained by the system by executing S1 or S2 processes alone or in combination.

By following the general suggestions presented in [65] we tested the DUAL-PECCS categorization system (see section 5), integrated with the ACT-R mechanisms, in a conceptual categorization task very similar to the psychological test known as "Word Reasoning" [22]. For human subjects, the Word Reasoning task consists in identifying a concept based on one to three clues. The participant might be told 'You can see through it' as a first clue; 'It is square and you can open it', and so on. The processing required by a Word Reasoning item goes beyond retrieval because the participant has to integrate the clues and choose among alternative hypotheses. Unfortunately, as reported by [65], the standard specific questions provided for this task in the Wechsler Preschool and Primary Scale of Intelligence are proprietary. Nonetheless, the general structure of each sentence is public. For this purpose we have therefore re-used a dataset composed of 112 linguistic descriptions (corresponding to very simple riddles) designed by a team of linguists and neuroscientist in the frame of a research project investigating neural correlates of lexical processing and already used for previous comparisons between humans and systems performances [23] [43].

Such descriptions exhibit a structure similar to that of the Word Reasoning task: on average, no more than 3 cues are present in each riddle. An example of such descriptions is "The mice hunter with whiskers and long tail", where

[22]For this experiment the system relies on a Conceptual Spaces KB of 300 concepts, executing S1 processes, integrated with the corresponding classes in the Open Cyc ontology via Wordnet IDs (see [46] for the details abut the integration). The S2 processes are operated on the ontological knowledge base and work as control mechanism w.r.t. the categorization results provided by type 1 processes which are non-monotonic in nature.

[23]The full list of descriptions is publicly available at: `http://goo.gl/EYJozw`.

the expected category to be retrieved was *cat*, and in particular its representation corresponding to the "prototype of *cat*"; conversely, a description such as "The feline mice hunter without fur" was expected to lead to the answer of "exemplar of *canadian-sphynx*". The *expected* categorical targets represent a gold standard, since they correspond to the results provided by 30 human subjects in a psychological experimentation and already described and presented elsewhere [66, 46].

For such descriptions we have recorded the categorization capabilities of the system by analyzing: i) when the expected categorical target is obtained by S1 processes in isolation ii) which role is played by the S2 types of processes iii) whether the S2 types of processes considered in isolation would have been able to provide the same, or better, results w.r.t. the S1 processes considered in isolation. The test of the effectiveness of S2 types of processes in isolation (the third condition mentioned above) has been executed by querying large ontological knowledge bases such as Cyc [34] and DBpedia. The differences between the two systems are reported as well. For querying both Cyc and DBpedia we have manually extracted the information from the text and have transformed then in SPARQL queries. For example: the description "A big, black and white sea bird that swims and cannot fly" corresponds to the following SPARQL query in DBpedia (provided here with a N3 notation to favour the readability) [24]

```
SELECT DISTINCT ?animal
WHERE {?animal
   dbpedia-owl:class dbpedia:Bird;
   dcterms:subject ?s1;
   dcterms:subject ?s2.
   ?page dbpedia-owl:family ?animal.
```

[24] The complete list of queries is available: https://goo.gl/fnwwqO.

```
FILTER(contains(bif:lower(str(?page)), "white")
  || contains(bif:lower(str(?page)), "black")).
FILTER(contains(bif:lower(str(?s1)), "fly")
  || contains(bif:lower(str(?s1)), "flight")
  && contains(bif:lower(str(?s1)), "less")).
FILTER(contains(bif:lower(str(?s2)), "sea")).
```

Table 1: Experimental results assessing the usefulness of the S1-S2 integration processes in a categorisation task w.r.t. the S1 or S2 processes considered in isolation.

Cases where $\mathcal{S}2$ confirmed the category returned by $\mathcal{S}1$	99.0% (111/112)
Cases where $\mathcal{S}1$ (alone) returned the expected category	77.7% (87/112)
Cases where $\mathcal{S}2$ (Cyc) alone returned the expected category	1.6% (2/112)
Cases where $\mathcal{S}2$ (DBpedia) alone returned the expected category	2.7% (3/112)

The results obtained by this experimentation are reported in table 1 [25].

7.1. Discussion

An interesting aspect revealed by this analysis is that the tested DUAL-PECCS system (explicitly based on both a heterogeneous representational hypothesis and on the dual process assumption) is able to categorize, thanks to the S1 component, stimuli with typical, though ontologically incoherent, descriptions. An example of such a case is the result obtained for the stimulus “The big fish that eats plankton”. In this case the expected prototypical answer is whale. However, whales are mammals, not fishes. In the adopted system, the S1 component returns the “whale” answer by resorting to the prototypical

[25] The results for the S1 categorization performance cover the full pipeline of the DUAL-PECCS system including the information extraction step from the natural language. Therefore some errors are due to the difficulty of this step. Without IE step the performance of the S1 system increase to the 89.3%.

knowledge. However, when the output of S1 is then checked with S2 processes against the Open Cyc ontology (the symbolic KB used in DUAL-PECCS), an inconsistency is detected and explained as follows:

```
subClassOf( (cetacean), (placental mammal) )
subClassOf( (mammal), (warm-blooded animal) )
subClassOf( (newClass), (whale) )
type( newIndiv, (newClass) )
subClassOf( (newClass), (fish) )
disjointWith( (warm-blooded animal), (cold-blooded animal) )
subClassOf( (fish), (cold-blooded animal) )
subClassOf( (placental mammal), (mammal) )
subClassOf( (whale), (cetacean) )
```

```
Laconic Explanation: Class (whale) is not (cold-blooded animal)
but is (warm-blooded animal)
```

As shown in the example above, the S2 processes activated by the ontological component provides the whole logical path leading to the inconsistency of the S1 result (it also provides a summary of the complete explanation, a laconic explanation that is easier to read and understand for human users). Due to the detected inconsistency, the first result of S1 is withdrawn and the next best result provided by S1 is tested [26]. This example shows in which cases the cycle of interaction between S1 and S2 processes can lead to revised and interesting conclusions.

An additional *datum* coming out of this evaluation is that S1 mostly provided an output coherent with the model in the S2 component (there is only one case, i.e. the one described above, where the S2 component corrects the

[26] In this case, the final output of the categorization system is a pair obtained by the first, fast and typical, S1 result and the second, slow and S2 compliant one. The details about the categorization algorithm and the termination conditions are in [46]

output of S1). This datum is of interest in that, although it is postulated that the reasoning check performed by S2 is beneficial to ensure a refinement of the categorization process, in this experimentation S2 did not reveal any significant improvement to the output provided by S1. This is, on the other hand, in line with the assumption that most of the common-sense answers can be successfully addressed, in the heterogeneous perspective, by the typical representational components adopting S1 processes. In addition, this datum can be additionally explained by considering the fact that the adopted dataset contains, as above mentioned, exclusively common-sense linguistic descriptions to be categorized. In cases of datasets with a different type of descriptions and involving, for example, the categorization of items based on necessary and sufficient condition (e.g., as happens in the mathematical domain) our prediction is that the S2 processes operating on the classical ontological representations could categorize very well the correct answer since, in this case, the activated reasoning process would correspond to a very simple and classical form of deductive categorization that has a different nature w.r.t. the S1 processes.

Finally, the current analysis showed that the S2 knowledge components (considered in isolation) are not able, de facto, to provide answers to most of the provided common-sense queries. The completely inadequate, or absent, answers provided by the tested large-scale ontological systems (Cyc and DBpedia) is a result compatible with the problems mentioned in section 3.5 and affecting this class of ontological structures (namely the fact that, due to the Tarskian-like semantics assumed by the underlying formalisms, the representation of common-sense information is largely absent in such knowledge bases). In other words, this is a symptom of the fact that such representational frameworks need to be integrated with other frameworks in order to be able to represent and reason on common-sense information. In general, the results obtained by this preliminary analysis suggests that, for common-sense reasoning and retrieval, the improvement provided by the adoption of the S2 mechanisms operating on classical symbolic structures is very limited.

However, it is also clear that it is not possible to explain the entire cognition of a cognitive agent exclusively in terms of S1 processes. Therefore, given the importance of the dual process approach in explaining how to harmonize and integrate different kind of reasoning processes assumed to co-exist in a heterogeneous representational perspective, additional investigations are needed.

In particular, in our opinion, such analyses should investigate: i) in which cases the S2 processes play a more relevant role w.r.t. the one proposed here ii) in which cases the S2 processes are not at all evoked by a cognitive system, since the need to react in real time is more pressing. Since there is not a clear answer to such questions, such aspects will involve, in our opinion, the future research agenda of both the cognitive psychology and cognitive (artificial) systems research.

8. Conclusions

In this paper we identified and characterized two main aspects concerning the knowledge level of the current CAs, namely the *size* and the homogeneous *typology* of the encoded knowledge. We have argued that, on the basis of the results coming from the experimental research in cognitive science, such aspects need to be addressed in order to structurally bind the knowledge level of the Cognitive Architectures to the constraints and challenges faced by human cognition in everyday knowledge processing tasks. Additionally, we have argued that these issues represent, from a technological perspective, a crucial challenge to address in order to be able to build cognitive agents able to operate and make decisions in general scenarios by exploiting a plethora of integrated reasoning mechanisms. Based on these assumptions we have provided an analysis of the most relevant CAs in the state of the art: we showed how all of them encounter, at different levels of granularity, some problems in dealing, jointly, with the above mentioned aspects. In the final part of the paper we have presented three different, but compatible, approaches that converge on the insight that, in order to address the problems affecting the knowledge level in CAs, the focus of attention should be posed on a more abstract level of representation w.r.t. the one addressed by neural representations (the analyzed approaches are: the Semantic Pointers approach; the approach based on Conceptual Spaces as intermediate representational level; and the novel Neuro-Symbolic Approach embedded in ACT-R).

Finally, since a crucial problem in the heterogeneous representational perspective is represented by the harmonization of different kinds reasoning processes, we have preliminarily investigated the usefulness of the dual process approach of reasoning by analyzing the results obtained by the DUAL-PECCS system in a categorization tasks. The obtained results suggest that, while the general heuristic provided by the dual approach represents a suitable way to integrate different reasoning mechanisms, it is still not clear (from a theoretical or an applicative point of view) if both dual process mechanisms are always activated. Therefore it is still an open question as to whether the hypothesized dual processes are worth considering as a general architectural mechanism (and, as such, worth implementing in the CAs processes operating on the conceptual structures of a cognitive agent) or as a local mechanism, activated under certain circumstances. As mentioned above, an answer to this question will require a joint investigation effort by both the cognitive psychology and the cognitive modeling community.

9. References

[1] G. Gigerenzer, P. M. Todd, Simple heuristics that make us smart, Oxford University Press, USA, 1999.

[2] H. A. Simon, A behavioral model of rational choice, The quarterly journal of economics (1955) 99–118.

[3] A. Newell, Physical symbol systems, Cognitive science 4 (2) (1980) 135–183.

[4] A. Newell, The knowledge level, Artificial intelligence 18 (1) (1982) 87–127.

[5] A. Newell, Unified theories of cognition, Harvard University Press, 1994.

[6] R. Cordeschi, The discovery of the artificial: Behavior, mind and machines before and beyond cybernetics, Vol. 28, Springer Science & Business Media, 2002.

[7] M. Miłkowski, Explaining the computational mind, Mit Press, 2013.

[8] A. Lieto, D. P. Radicioni, From human to artificial cognition and back: New perspectives on cognitively inspired ai systems, Cognitive Systems Research 39 (2016) 1–3.

[9] H. Putnam, Minds and machines, MacMillan, 1960.

[10] E. Rosch, Cognitive representations of semantic categories, J. Exp. Psychol. Gen. 104 (3) (1975) 192–233.

[11] G. L. Murphy, The big book of concepts, MIT press, 2002.

[12] E. Machery, Doing without concepts, OUP, 2009.

[13] B. C. Malt, An on-line investigation of prototype and exemplar strategies in classification., Journal of Experimental Psychology: Learning, Memory, and Cognition 15 (4) (1989) 539.

[14] M. Frixione, A. Lieto, Representing non classical concepts in formal ontologies: Prototypes and exemplars, in: New challenges in distributed information filtering and retrieval, Springer, 2013, pp. 171–182.

[15] A. Lieto, A computational framework for concept representation in cognitive systems and architectures: Concepts as heterogeneous proxytypes, Procedia Computer Science 41 (2014) 6–14.

[16] A. Oltramari, C. Lebiere, Pursuing artificial general intelligence by leveraging the knowledge capabilities of act-r, in: Artificial General Intelligence, Springer, 2012, pp. 199–208.

[17] D. Vernon, Artificial cognitive systems: A primer, MIT Press, 2014.

[18] J. Laird, The Soar cognitive architecture, MIT Press, 2012.

[19] J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, Y. Qin, An integrated theory of the mind., Psychological review 111 (4) (2004) 1036.

[20] R. Sun, The CLARION cognitive architecture: Extending cognitive modeling to social simulation, Cognition and multi-agent interaction (2006) 79–99.

[21] S. Franklin, F. Patterson Jr, The lida architecture: Adding new modes of learning to an intelligent, autonomous, software agent, pat 703 (2006) 764–1004.

[22] J. Snaider, S. Franklin, Vector lida, Procedia Computer Science 41 (2014) 188–203.

[23] D. Vernon, G. Metta, G. Sandini, A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents, IEEE Transactions on Evolutionary Computation 11 (2) (2007) 151.

[24] P. Langley, J. E. Laird, S. Rogers, Cognitive architectures: Research issues and challenges, Cognitive Systems Research 10 (2) (2009) 141–160.

[25] K. Thórisson, H. Helgasson, Cognitive architectures and autonomy: A comparative review, Journal of Artificial General Intelligence 3 (2) (2012) 1–30.

[26] A. Newell, H. A. Simon, Computer science as empirical inquiry: Symbols and search, Communications of the ACM 19 (3) (1976) 113–126.

[27] M. Frixione, A. Lieto, Representing concepts in formal ontologies: Compositionality vs. typicality effects, Logic and Logical Philosophy 21 (4) (2012) 391–414.

[28] J. R. Anderson, J. Betz, A hybrid model of categorization, Psychonomic Bulletin & Review 8 (4) (2001) 629–647.

[29] G. A. Miller, Wordnet: a lexical database for english, Communications of the ACM 38 (11) (1995) 39–41.

[30] C. F. Baker, C. J. Fillmore, J. B. Lowe, The berkeley framenet project, in: Proceedings of the 17th international conference on Computational linguistics-Volume 1, Association for Computational Linguistics, 1998, pp. 86–90.

[31] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, A. Oltramari, Wonderweb deliverable d18, ontology library (final), ICT project 33052.

[32] D. D. Salvucci, Endowing a cognitive architecture with world knowledge, in: Procs. of the 36th Annual Meeting of the Cognitive Science Society, 2014.

[33] J. Ball, S. Rodgers, K. Gluck, Integrating act-r and cyc in a large-scale model of language comprehension for use in intelligent agents, in: AAAI workshop, 2004, pp. 19–25.

[34] D. B. Lenat, Cyc: A large-scale investment in knowledge infrastructure, Communications of the ACM 38 (11) (1995) 33–38.

[35] L. Giordano, V. Gliozzi, N. Olivetti, G. L. Pozzato, A non-monotonic description logic for reasoning about typicality, Artificial Intelligence 195 (2013) 165–202.

[36] C. Eliasmith, T. C. Stewart, X. Choo, T. Bekolay, T. DeWolf, Y. Tang, D. Rasmussen, A large-scale model of the functioning brain, Science 338 (6111) (2012) 1202–1205.

[37] P. Blouw, E. Solodkin, P. Thagard, C. Eliasmith, Concepts as semantic pointers: a framework and computational model, Cognitive science.

[38] E. Crawford, M. Gingerich, C. Eliasmith, Biologically plausible, human-scale knowledge representation, Cognitive science.

[39] P. Gärdenfors, Conceptual spaces: The geometry of thought, MIT press, 2000.

[40] P. Gärdenfors, Symbolic, conceptual and subconceptual representations, in: Human and Machine Perception, Springer, 1997, pp. 255–270.

[41] A. Chella, M. Frixione, S. Gaglio, A cognitive architecture for artificial vision, Artificial Intelligence 89 (1) (1997) 73–111.

[42] A. Chella, M. Frixione, S. Gaglio, Anchoring symbols to conceptual spaces: the case of dynamic scenarios, Robotics and Autonomous Systems 43 (2) (2003) 175–188.

[43] A. Lieto, D. P. Radicioni, V. Rho, A common-sense conceptual categorization system integrating heterogeneous proxytypes and the dual process of reasoning, in: In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), Buenos Aires, AAAI Press, 2015, pp. 875–881.

[44] A. Chella, A cognitive architecture for music perception exploiting conceptual spaces, in: Applications of Conceptual Spaces, Springer, 2015, pp. 187–203.

[45] F. Zenker, P. Gärdenfors, Applications of conceptual spaces - The case for geometric knowledge representation, Springer, 2015.

[46] A. Lieto, D. P. Radicioni, V. Rho, Dual peccs: a cognitive system for conceptual representation and categorization, Journal of Experimental & Theoretical Artificial Intelligence 29 (2) (2017) 433–452.

[47] A. Lieto, A. Chella, M. Frixione, Conceptual spaces for cognitive architectures: A lingua franca for different levels of representation, Biologically Inspired Cognitive Architectures 19 (2017) 1–9.

[48] J. A. Fodor, The present status of the innateness controversy, in: J. A. Fodor (Ed.), Representations: Philosophical Essays on the Foundations of

Cognitive Science, MIT Press, Cambridge, MA, 1981, Ch. 10, pp. 257 – 316.

[49] D. N. Osherson, E. E. Smith, On the adequacy of prototype theory as a theory of concepts, Cognition 9 (1) (1981) 35–58.

[50] J. A. Fodor, Z. W. Pylyshyn, Connectionism and cognitive architecture: A critical analysis, Cognition 28 (1-2) (1988) 3–71.

[51] M. Lewis, J. Lawry, Hierarchical conceptual spaces for concept combination, Artificial Intelligence 237 (2016) 204–227.

[52] J. Derrac, S. Schockaert, Inducing semantic relations from conceptual spaces: a data-driven approach to plausible reasoning, Artificial Intelligence 228 (2015) 66–94.

[53] A. Lieto, E. Mensa, D. P. Radicioni, A resource-driven approach for anchoring linguistic resources to conceptual spaces, in: AI* IA 2016 Advances in Artificial Intelligence, Springer, 2016, pp. 435–449.

[54] Y. Vinokurov, C. Lebiere, D. Wyatte, S. Herd, R. O'Reilly, Unsurpervised learning in hybrid cognitive architectures, in: Workshops at the twenty-sixth AAAI conference on artificial intelligence, 2012.

[55] R. C. O'Reilly, A. A. Petrov, J. D. Cohen, C. J. Lebiere, S. A. Herd, T. Kriete, I. P. Calvo, J. Symons, How limited systematicity emerges: A computational cognitive neuroscience approach.

[56] H. V. D. Parunak, P. Nielsen, S. Brueckner, R. Alonso, Hybrid multi-agent systems: integrating swarming and bdi agents, in: International Workshop on Engineering Self-Organising Applications, Springer, 2006, pp. 1–14.

[57] R. Wray, C. Lebiere, P. Weinstein, K. Jha, J. Springer, T. Belding, B. Best, V. Parunak, Towards a complete, multi-level cognitive architecture, in: Proceedings of the eighth international conference on cognitive modeling, Ann Arbor, MI, 2007, pp. 325–330.

[58] S. Fahlman, Using scones multiple-context mechanism to emulate human-like reasoning, in: First International Conference on Knowledge Science, Engineering and Management (KSEM'06), Springer–Verlag (Lecture Notes in AI), Guilin, China, 2006.

[59] F. Baader, D. Calvanese, D. L. Mcguinness, D. Nardi, P. F. Patel-Schneider (Eds.), The Description Logic Handbook : Theory, Implementation and Applications, Cambridge University Press, 2003.

[60] K. E. Stanovich, R. F. West, Advancing the rationality debate, Behavioral and brain sciences 23 (05) (2000) 701–717.

[61] J. S. B. Evans, K. E. Frankish, In two minds: Dual processes and beyond., Oxford University Press, 2009.

[62] D. Kahneman, Thinking, fast and slow, Macmillan, 2011.

[63] E. Davis, Representations of commonsense knowledge, Morgan Kaufmann, 2014.

[64] M. Minsky, The emotion machine: Commonsense thinking, artificial intelligence, and the future of the human mind, Simon and Schuster, 2007.

[65] S. Ohlsson, R. H. Sloan, G. Turán, D. Uber, A. Urasky, An approach to evaluate ai commonsense reasoning systems., in: FLAIRS Conference, 2012.

[66] D. P. Radicioni, F. Garbarini, F. Calzavarini, M. Biggio, A. Lieto, K. Sacco, D. Marconi, On Mental Imagery in Lexical Processing: Computational Modeling of the Visual Load Associated to Concepts, in: G. Airenti, B. Bara, G. Sandini, M. Cruciani (Eds.), Proceedings of the EuroAsian-Pacific Joint Conference on Cognitive Science (EAP-COGSCI 2015), 2015, pp. 181–186.