# A Computational Framework for Concept Representation in Cognitive Systems and Architectures: Concepts as Heterogeneous Proxytypes

Antonio Lieto[1]

[1] *University of Torino, Dept. of Computer Science, Italy.*
*ICAR – CNR, Palermo, Italy.*
*lieto@di.unito.it, lieto.antonio@gmail.com*

---

**Abstract**
In this paper a possible general framework for the representation of concepts in cognitive artificial systems and cognitive architectures is proposed. The framework is inspired by the so called proxytype theory of concepts and combines it with the heterogeneity approach to concept representations, according to which concepts do not constitute a unitary phenomenon. The contribution of the paper is twofold: on one hand, it aims at providing a novel theoretical hypothesis for the debate about concepts in cognitive sciences by providing unexplored connections between different theories; on the other hand it is aimed at sketching a computational characterization of the problem of concept representation in cognitively inspired artificial systems and in cognitive architectures.

*Keywords:* Knowledge Representation, Concept Representation, Cognitive Systems, Cognitive Architectures, Proxytypes.

---

## 1 Introduction

The representation and processing of conceptual information plays a crucial role in many cognitive tasks (e.g. knowledge organization, reasoning, learning etc.) each involving different brain areas and several neural connections. In this paper, a possible general framework for the representation of concepts in cognitive artificial systems and cognitive architectures is proposed. The proposed framework is introduced at the computational level of detail according to Marr [11]. This means that, despite some representational details are sketched (see section 4), the main goal of this article is that one of providing a possible common ground between the theoretical and the computational cognitive science concerning the problem of concept representation. In particular, our proposal is inspired by the so called proxytype theory of concepts [16] and combines it with the heterogeneity approach to concept representations, according to which concepts do not constitute a unitary phenomenon (see [9]). The contribution of the paper is twofold: on one hand, it is aimed at providing a possible computational characterization of the problem of concept representation in cognitively inspired artificial systems and in cognitive architectures; on the other hand, it aims at furnishing a renewed theoretical framework for the debate about concepts in cognitive sciences, by providing unexplored connections and possible "bridges" between different theories. The rest of the paper is organized as follows: in section 2 a brief overview of the theoretical debate about concepts in cognitive science is provided; in section 3 the proxytype theory of concepts, as well as its connection with biological arguments, is introduced and the theoretical connection with the heterogeneous approach is proposed. Section 4 shows how the proposed framework can be plausibly implemented in cognitively inspired artificial systems and architectures by adopting different frameworks for representing the conceptual knowledge. Section 5, finally, provides a short review of the related works. Conclusions end the paper.

# 2 Concepts in Cognitive Science

In Cognitive Science different theories about the nature of concepts have been proposed. According to the traditional view, known as "classical" or "aristotelian" theory, concepts can be simply defined in terms of sets of necessary and sufficient conditions. Such theory was dominant in philosophy and psychology from the antiquity until the mid '70s of the last century when, the empirical results of Eleanor Rosch [15] – historically preceded by the philosophical analyses by Ludwig Wittgenstein [21] – demonstrated the inadequacy of such a theory for ordinary – common sense – concepts. Such results showed, on the other hand, that ordinary concepts are characterized and organized in our mind in terms of prototypes.

The Rosch's results have had a crucial importance for the development of different theories of concepts trying to explain different representational and reasoning aspects concerning the problem of typicality. Usually, such theories, are grouped in three main classes: prototype theory (developed with the above mentioned work of Rosch), exemplars theory and theory-theories (see e.g. [13] and [9] for a detailed review of such approaches). All of them are assumed to account for (some aspects of) the typicality effects in conceptualization.

According to the prototype view, knowledge about categories is stored in terms of prototypes, i.e. in terms of some representation of the "best" instances of the category. For example, the concept CAT should coincide with a representation of a typical cat. 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 set of specific exemplars explicitly stored within memory: the mental representation of the concept CAT is the set of the representations of (some of) the cats we encountered during our lifetime.

Theory-theories 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 CAT is individuated by the role it plays in our mental theory of zoology. In other version of the approach, concepts themselves are identified with micro-theories of some sort. For example, the concept CAT should be identified with a mentally represented micro-theory about cats.

Despite such approaches have been historically considered as "competitors", since they propose different models and have different predictions about how the humans organize and reason on conceptual information, different works (the first in this line was proposed by [10]) show that they turned out to be not mutually exclusive. Rather, they seem to succeed in explaining different classes of cognitive phenomena. In particular, the empirical data – i.e. behavioral measures such as categorization probability and reaction time - suggest that subjects use different representations to categorize. Some use exemplars, a few rely on prototypes, and others appeal to both exemplars and prototypes. In addition, this distinction seems to have also neural plausibility, as it is witnessed by various empirical researches (the first in this line is that one of Squire and Nolton [18]).

Such experimental results led, in the area of theoretical cognitive science, to the development of the so called "heterogeneous hypothesis" about the nature of concepts. Such approach tries to unify under the same umbrella the different theories and is based on the assumption that concepts do not constitute a unitary phenomenon (see [9]). In particular, different types of concepts are assumed to exists: prototypical concepts, exemplar based concepts and theory-theory concepts. All such concepts constitute different bodies of knowledge and may contain different levels of representation of the information associated to the same conceptual entity (i.e. the different bodies of knowledge are “semantic pointers”, in the sense intended by [4] and [19], towards the same conceptual entity; see section 5 for further details). Furthermore, each body of conceptual knowledge obeys to specific processes in which such representations are involved (e.g. in cognitive tasks such as recognition, learning, categorization etc.). An illustrative example of such a situation is reported, in a very simplified way, in the figure below; it shows that different types of concepts constitute different bodies of knowledge referred to the same conceptual entity. On such body of knowledge specific cognitive processes take places. Of course it is not necessary that all the different bodies of knowledge are filled since for certain concepts does not exists a characterization in prototypical, exemplars or theory-theory based terms.

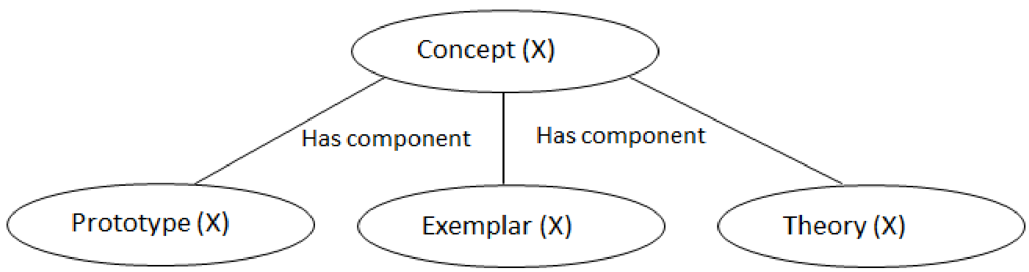


Figure 1. Heterogeneous Hypothesis of Concepts*.

# 3 Coupling the Heterogeneous Approach and the Proxytype Theories of Concepts

In more recent years, the philosopher Jesse Prinz [16] proposed an alternative theory of concepts: the proxytype theory. In particular such theory, directly inspired by what proposed by Lawrence Barsalou (see e.g. [1]), has the quality of introducing in a systematic way a biological characterization within an explanatory theory of concepts. Such characterization is particularly interesting for the explanation of some phenomena such as, for example, that one regarding the activation (and the retrieval) of the conceptual information.

*In primis*, Prinz draws a distinction, quite common in the area of cognitive architectures, between long term and working memory (thus hypothesizing a biological localization and interaction between different brain areas for dealing with conceptual structures). By adopting this distinction, Prinz identifies concepts with proxytypes. A proxytype is any element of a complex representational network stored in long-term memory corresponding to a particular category that could be tokened in working memory to “go proxy” for that category. This view stems from the consideration that concepts can be possessed even if they are not currently being used, together with the consideration

* In the figure 1 it is not reported the whole spectrum of representations that can potentially point to a given conceptual entity. For example it is not reported the body of knowledge containing a “classical” characterization of the concept X in terms of necessary and sufficient conditions (assuming that such characterization exists for the concept under consideration).

that working memory cannot activate the entire network of stored memories at any given time. Context will determine which stored memories will be recalled on any given occasion. With respect to concepts, what exists in long-term memory are complex networks of representations. What binds together the elements of these networks are causal connections. The elements are causally connected in that activation of any one element of the network (an activation that involves its tokening in working memory) will typically cause the activation of some other element. These networks stored in long-term memory correspond to categories of things in the outside world. For example, there is a network corresponding to cats. Such a network was constructed over time on the basis of perceptual interactions with cats. In other terms, Prinz's proxytype theory considers concepts as temporary constructs in working memory having form of multimodal perceptual representation (or copies of a particular activation of long term representations).

By proposing his approach, however, Prinz considers proxytypes themselves as being essentially prototype-like. In our opinion, however, the idea that different kinds of proxytypes concepts can be tokenized in the working memory can be plausibly linked to the assumption proposed by the heterogeneous approach to conceptual representations, according to which a wider range of representation types can be available to serve as concepts than just prototypes (such representations, in line with Prinz theory, should be stored in long term memory). Summing up, a plausible biologically and cognitive inspired theoretical framework for the representation of concepts can be hypothesized by combining the proxytype theory of concepts with the heterogeneous approach to conceptual representations. Namey: proxytypes can be assumed to be heterogeneous in nature, based on the different representations from which they depend on[†]. In this view, the activation of specific representations, and their process of "going proxy" in working memory, can be seen as a function of the perceptual stimulus (verbal or not verbal) coming from the external environment. A simple example of the explanatory advantage provided by the proposed framework is presented in the figure 2 below. For the sake of simplicity the example only concerns the prototypes and exemplars representations.

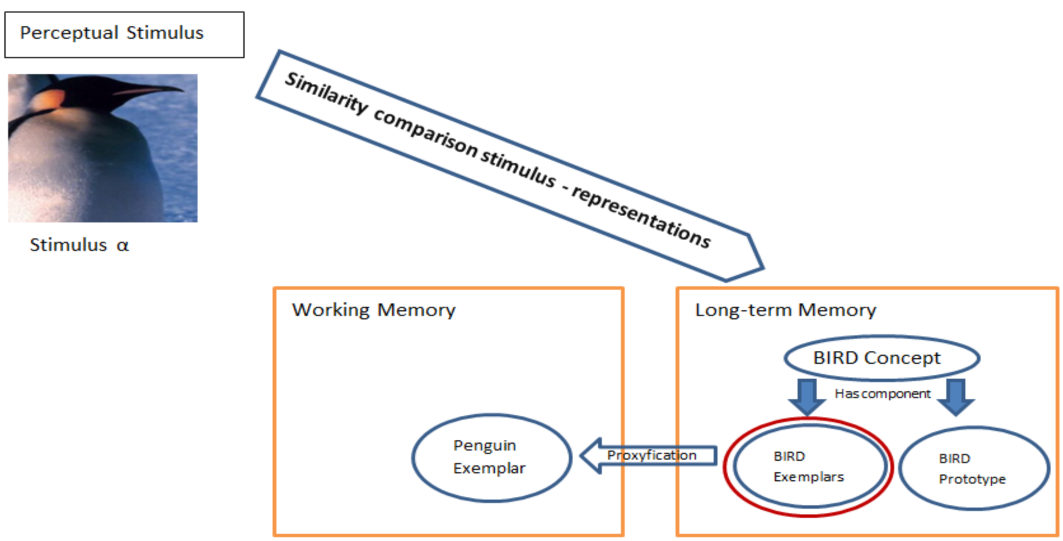


Figure 2. A pictorial representation of the *proxification* process involving *heterogeneous* representations

[†] From a theoretical point of view this approach differs from the pluralistic approach to concepts [20] since the different conceptual structures composing a concept are assumed to act as semantic pointers towards a common conceptual entity. Therefore they are not considered separated concepts.

Let us consider to perceive the visual[‡] input α presented in the figure 2 above and that we need to categorize it. By adopting the approach based on heterogeneous proxytypes, such visual input can be compared with a set of representations available in the long term memory.

The most similar representation w.r.t. the given input will be then made available in the working memory as a proxy. In the case presented in the figure, the item α corresponds to a rather atypical bird since it is a penguin (and, for example, the penguins are heavy and are not flying birds). This means that, in such case, the result of the similarity comparison between the external percept and heterogeneous internal conceptual representation will favor the tokenization of an exemplars-based representation[§], as also suggested by the psychological literature [10].

# 4 Computational Representation of Concepts

In Artificial Intelligence (AI), and in particular in the area of Knowledge Representation, different proposals have been made in literature for representing concepts and for dealing with different aspects of the conceptual information (e.g. in processes such as learning and reasoning). In particular, during the ’70s-‘80s of the last century, in the context of theoretical debates on connectionism, a classical distinction was in auge between symbolic and sub-symbolic models and processes of representation. While sub-symbolic, or connectionist models, were used for embedding knowledge structures and processed more closely to human-like organizations and processes, on the other hand, many forms of logic-based systems were developed which were mostly oriented on the possibility of providing a clear formal semantics, thus enabling forms of logically-valid automatic reasoning. Examples of this systems are the KL-ONE semantic networks [2]. In more recent years, the cognitive scientist Peter Gärdenfors [7] proposed a novel distinction based on a tripartition of representational levels where, instead of a symbolic/sub-symbolic dichotomy, a further level is considered: namely the conceptual level. This level of representation is intermediate between the other two, and is characterized by a representation in terms of conceptual spaces, i.e. geometrical representations of knowledge that consist of a number of quality dimensions where concepts are represented as convex regions whose center corresponds to prototypical representations.

Such a tripartition can be useful for providing different representational insights, from a computational point of view, regarding how the proposed framework based on heterogeneous proxytypes can be effectively applied in practice. From an operational point of view, in fact, it is possible to argue that each body of conceptual knowledge can be represented by recurring to different framework and formalisms already available and used in the AI and Cognitive Modelling literature.

The prototypical body of knowledge for a given concept C, for example, can be represented by adopting different computational frameworks. From a symbolic perspective, prototypical representations can be encoded in terms of frames [12] or semantic networks [14]; from a conceptual space perspective prototypes are geometrically represented as centroids of a convex region; and from a sub-symbolic perspective the prototypical knowledge concerning a concept is typically represented as reinforced patterns of connections emerging according to classical Hebbian mechanisms in artificial neural networks (ANN)[**]. Similarly, for the exemplars-based body of knowledge, both symbolic and conceptual space representations can be used (for the hypothesis regarding the representation of exemplar-based representations in conceptual space see [5]) as well as the sub-symbolic paradigm. In

[‡] As already mentioned, the same represented process is also feasible in the case of linguistic input. In such case, the similarity comparison implies a phase of **linguistic parsing** and a phase of **concept grounding** between the stimulus and the corresponding conceptual representation.

[§] The “similarity comparison” can be assumed to start in parallel for all the representations and end with when a certain similarity threshold is satisfied. Therefore, as demonstrated by the works mentioned in section 2 on human categorization, different time responses can be hypothesized and tested according to the type of stimulus presented in input (e.g. in the case of penguin, the exemplar representation will be activated first and tokenized, thus stopping the whole process of comparisons).

[**] For a detailed review of the AI approaches to the common sense computing we remind at [3].

particular: exemplars can be represented as instances of a concept in symbolic systems, as points in a conceptual space or as a particular pattern of activation in a ANN. Also for the theory-theory based component is, at least in principle, possible to use all such different levels of representation. However this seems to be the case where symbolic and conceptual level allow a most promising descriptive way w.r.t. the sub-symbolic one.

Summing up, all the different types of conceptual representations can be implemented, with different methods and techniques, in cognitive artificial systems and architectures. In addition, different computational mechanisms for "proxyfying" the conceptual representations can be applied.

The simplest version of the proxification process can be obtained, for example, by implementing IF-THEN rules able to activate the working memory tokenization of a certain representation when a given similarity threshold is satisfied. However, in such scenario, more complex and cognitively-inspired methods can be hypothesized, adopted and tested in the computational arena.

# 5 Related Works

One of the main difference of the proposed framework w.r.t. other biologically inspired proposals to concept representation, in given by the fact that, in our case, the conceptual representations are specifically assumed to be heterogeneous in nature (i.e. are assumed to be composed of different bodies of knowledge referred to the same conceptual entity). Therefore, from a computational perspective, the different types of representational solutions briefly presented in the previous section (e.g. symbolic solutions, conceptual spaces and ANN) can be combined and integrated in order to represent different semantic aspects of the same conceptual entity. For example: the "classical" or "theoretical" body of knowledge of a given concept can be represented by using standard symbolic frameworks (such as the computational ontologies), while the prototypical and exemplar-based bodies of knowledge can be demanded to representations encoded in terms of conceptual spaces or ANN. The description of the way in which such different knowledge modules, and their corresponding processes, interact between them is out of the scope of the current work. One of the current research directions for studying such mechanisms is based on the dual process theory of reasoning and rationality (see [6]), but several different directions may be taken and deserve further investigations.

In order to better explain the difference of the proposed approach w.r.t. other existing proposals let us consider the figure 3 below.

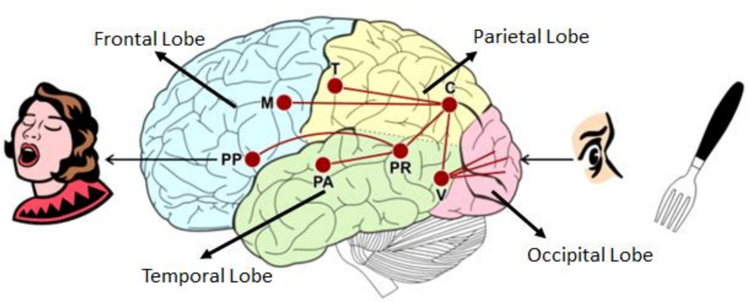


Figure 3. Neurocognitive network for "Fork"

In such figure, taken from Sowa [17] and presenting a neurocognitive network used in a biologically inspired framework for the computational representation of the information related to

“Fork”, the node **labeled C** represents the concept of a fork. In such framework, **C** is assumed to occur in the parietal lobe, which is closely linked to the primary projection areas for all the sensory modalities. For the image of a fork, **C** is connected to node V, which has links to percepts for the parts and features of a fork in the visual cortex (occipital lobe). For the tactile sensation of a fork, **C** links to node T in the sensory area for input from the hand. For the motor schemata for manipulating a fork, **C** links to node M in the motor area for the hand. For the phonology for recognizing the word Fork, **C** links to node PR in Wernicke’s area. Finally, PR is linked to node PA for the sound /fork/ in the primary auditory cortex and to node PP in Broca’s area for producing the sound. The main differentiation of such proposal w.r.t. what discussed in this article is given by the fact that, in our framework, different proxytypes C can be activated each corresponding to different representations in the long term memory. In addition, according to the tokenized representation in the working memory, different neurocognitive “conceptual” networks can be activated which are functions of the activated representation.

Our framework is also apparently similar to another biological proposal considering concepts as “semantic pointers” ([4] and [19]) and implemented in the field of the computational modelling of brain. Also in such approach, in fact, different informational components are supposed to point to a unifying conceptual entity. The similarity, however, is limited to this point. In fact, while [4] and [19] consider, as “conceptual components”, the different information channels through which the content of the information is provided (e.g. their components are divided in: sensory, motor, verbal etc.; and, for each type of carrier, a mapping function to a brain area is supposed to be activated), in our framework, the heterogeneity of the representational components is assumed by considering the different typologies of the conceptual content vehiculated (e.g. typical vs classical vs exemplars-based etc.). Therefore our proposal focuses on the heterogeneity regarding the *content* of the represented information, while [4] and [19] propose an heterogeneity at the level of the information channel. In our opinion our axis of representation provide an additional level of generality since the *content* distinction is cross-channel.

# 6 Conclusions

In this paper a possible general framework for the representation of concepts in cognitive artificial systems and cognitive architectures is proposed. The proposed framework is inspired by the so called proxytype theory of concepts and combines it with the heterogeneity approach to concept representations according to which concepts do not constitute a unitary phenomenon. This proposal is a first theoretical attempt to bridge, in a plausible way, different theories of concepts in Cognitive Science by providing also some insights concerning how the proposed framework can be implemented in cognitive artificial systems and architectures. As future - short term – goal a computational model of such framework will be devised in the field of lexical processing. In such a model the input of the system will be represented by a natural language definition and the output by the activation and retrieval of specific proxytypes according to the given definition. Such a framework will represent an extension of a question answering system already developed [8] and used for the tasks of concept identification and retrieval. A further development of this line of research will consists in testing the hypothesis of the proposed framework in the area of cognitive architectures.

## Acknowledgements

I am indebted to Marcello Frixione and Leonardo Lesmo for the fruitful and deep discussions regarding the ideas and the topics presented in this article. I also would like to thank the anonymous reviewers for their valuable comments. This work is partially supported by the Ateneo-San Paolo project number TO_call03_2012_0046, The role of visual imagery in lexical processing (RVILP), by the IT for Law Project (University of Torino) and by the ICAR-CNR F.A.C.I.L.E. Project (ICT.P08.003.001, Framework for the management and intelligent fruition of knowledge and perceptual information in cognitive agents). I would like to thank everyone who worked to make these opportunities available to me.

# References

[1] Barsalou L.W., Santos A., Simmons W.K., Wilson C.D., (2008). Language and Simulation in Conceptual Processing, Symbols, embodiment and meaning, pp. 245-283.

[2] Brachman R. and J.G. Schmolze J.G., (1985). An overview of the KL-ONE knowledge representation system, Cognitive Science, 9, pp. 171-216.

[3] Cambria E., Hussain A., Havasi, Eckl C., (2009) Common Sense Computing: from the Society of Mind to Digital Intuition and Beyond in Proceedings of Joint COST 2101 and 2102 International Conference, BioID_MultiComm 2009, Madrid, Spain, September 16-18.

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

[5] Frixione M. and Lieto A., (2013). Exemplars, prototypes and conceptual spaces. Advances in Intelligent Systems and Computing, Proceedings of BICA 2012, Volume 196, 131-136, Springer.

[6] Frixione M. and Lieto A., (2014). Towards an Extended Model of Conceptual Representations in Formal Ontologies: A Typicality-Based Proposal in Journal of Universal Computer Science, Vol. 20, Issue 3, pp. 257-276.

[7] Gärdenfors P., (2000). Conceptual Spaces: The Geometry of Thought. The MIT Press/Bradford Books, Cambridge, MA, 2000.

[8] Lieto A., Minieri A., Piana A., Radicioni D.P. A Knowledge based system for prototypical reasoning, (2015) to appear in Connection Science. DOI: http://dx.doi.org/10.1080/09540091.2014.956292

[9] Machery E., (2010). Doing without Concepts. Oxford University Press, Oxford.

[10] Malt B.C., (1989). An on-line investigation of prototype and exemplar strategies in classification, Journal of Experimental Psychology 15 (4), pp. 539-555

[11] Marr, D., (1977). Artificial intelligence - A personal view. Artificial Intelligence 9: 37-48. Marr, D. 1982 . Vision. San Francisco : W . H . Freeman.

[12] Minsky M., (1975). A framework for representing knowledge, in Patrick Winston (ed.), The Psychology of Computer Vision, McGraw-Hill, New York.

[13] Murphy G.L., (2002). The Big Book of Concepts. The MIT Press, MA.

[14] Quillian M.R., (1968). Semantic memory. In M. Minsky (ed.), Semantic Information Processing, The MIT Press, Cambridge, MA.

[15] Rosch E., (1975). Cognitive representation of semantic categories, Journal of Experimental Psychology, 104, pp. 573-60.

[16] Prinz J., (2004). Furnishing the Mind: Concepts and Their Perceptual Basis, MIT Press.

[17] Sowa J., (2012). Cognitive Architectures for Conceptual Structures, Conceptual Structures for Discovering Knowledge Lecture Notes in Computer Science Volume 6828, pp 35-49.

[18] Squire L., Knowlton B., (1995). Learning about categories in the absence of memory. Proc. Nat. Acad. Sc. USA, 92(26): 1247012474.

[19] Thagard, P., (2012). The cognitive science of science: Explanation, discovery, and conceptual change. Cambridge, MA: MIT Press.

[20] Weiskopf, D. A; (2009). The plurality of concepts. Synthese 169.1 (2009): 145-173.

[21] Wittgenstein L., (1953). Philosophische Untersuchungen, Blackwell, Oxford.