The Need of an Interdisciplinary Approach based on Computational Modelling in the Study of Categorization.

Francesco Gagliardi (francesco.gagliardi@libero.it)
Department of Philosophical and Epistemological Studies
University of Rome *La Sapienza* — Via Carlo Fea — I-00161 Roma, Italy

Categorization in Cognitive Psychology and the Prototypes-Exemplars Debate

The main theories (Murphy, 2002) concerning the study of categorization and the nature of concepts are: the classical theory also known as Aristotelian, the prototypes theory, the exemplars theory and theory-theory. The theories of prototypes and exemplars, jointly taken, constitute the so called *typicality view* on concepts. In fact, both theories, even if in contrast, are based on experimental evidences that as a whole they show the existence of a "phenomenon" of typicality in categorization processes (see *"Typicality as phenomenon"* in Murphy; 2002 pg. 28). Prototypes and exemplars theories supersede the limitations and the experimental inadequacy of the classical theory, based on logical predicates, but when considered separately turn out to be incomplete and unsatisfactory (Murphy, 2002; pg. 4). Nevertheless, in the past thirty years some literature concerning experimental psychology focused on the comparison between prototypes theory and exemplar theory and carrying out experiments in order to demonstrate the correctness of one theory or the other one. For example we can consider the following two papers in conflict (Minda, Smith, 2002) and (Zaki et al., 2003). In the former is supported prototypes theory, while in the latter the exemplars theory, even if they make use of the same data set. As matter of fact, the research line related to the diatribe of prototypes vs. exemplars appears to be a dead end because it is fruitless and not decisive and also because it is based on the *naïve* epistemology of pursuing a so called *experimentum crucis*.

Theories of Categorization and Machine Learning

The ultimate aim of the researches about categorization is the understanding of representations of categories (Murphy, 2002; pg. 3) that we build, the concepts, and by which we perform different cognitive tasks. A common aspect of prototypes theory and exemplar theory is the idea that each category is represented by instances belonging to the class: in one case the instances are the prototypes abstracted from observations, and in the other case are the same previously observed instances. In the field of machine learning (Witten, Frank, 2005) (Duda, Hart, Stork, 2000) and automatic classification, one of the learning methodologies known in literature is the so called *instance based learning,* for which the classes, learnt by the classifier system, are represented by instances of the corresponding class. Therefore, the field of machine learning, and in particular of instance-based learning, is the natural context where to study the theories of human categorization based on prototypes or exemplars, from both the theoretical viewpoint of the computational statistics, and the empirical viewpoint of the synthetic method (Cordeschi, 2001), consisting in the realization of classifier systems which embody theories of categorization. Within instance-based learning it is possible to connect the characteristics of robustness and sensibility of a classifier system with categories representation based, respectively, on prototypes or exemplars. In fact, prototypes based classifiers, such as the *Nearest Prototype Classifier* (*NPC*) and the *Nearest Multiple-Prototype Classifier* (*NMPC*), construct the representative instances of the class, called *prototypes*, as the barycentres of an observations subset. These systems obtain robust classifications, that is, not sensitive to noisy and atypical observations. On the other way, classifiers based on exemplars, such as the *Nearest Neighbour Classifier* (*NNC*) and its well known generalization *k-NNC,* use as the set of representative instances the whole set of observations of classes, without any elaboration or abstraction. These systems, which are entirely based on the ability to save all observations in memory, obtain classifications extremely sensible and not at all robust. In the family of instance-based systems the classifiers *NPC* and *NNC* represent the limit cases of maximum robustness and maximum sensibility respectively and they use types of classes representations that can be related to the theories of prototypes and of exemplars, respectively. As it is well known in computational statistics a classifier system, whether natural or artificial, is the result of the trade-off between the two contrasting requisites of robustness and sensibility. More formally this problem is linked with the *Bias-Variance theorem* and with the *Bias-Variance dilemma*, e.g. (Duda, Hart, Stork, 2000, Chap.9). Thus *Prototype-Theory* and *Exemplar-Theory* have not to be considered as two conflicting theories, but they are two limit cases of a same technique to categorize called *Instance-Based Learning*. This technique is used both by natural systems as human minds, and by some artificial systems as instance-based classifier systems. From these simple theoretical considerations it is then clear that it is absolutely groundless to assert the correctness of one of the two theories against the other; a theory which subsumes both of them should be sought just in the trade-off between robustness and sensibility. In fact there are some classifier systems, such as the *Varying Abstraction Model* (Vanpaemel, 2005), the *Mixture Model* (Rosseel, 2002) or the *Prototype-Exemplar Learning Classifier* (Gagliardi, 2008), which are able to subsume both the prototypes and exemplars theories and, hence, they can help to realize a *theory of typicality* which would explain the phenomenon of typicality. In summary, we can affirm that when framing the problem of categorization in the field of machine learning, the prototypes-exemplars diatribe reveals completely unfounded for the general theoretical considerations about the bias-variance dilemma, and also for experimental evidences due to the existence of some hybrid classifier systems. Therefore, the aforementioned diatribe is ill-posed, because of a poor formalization of the subject and the *naïve* epistemology of *experimentum crucis*. These drawbacks could be superseded, by using results of machine learning and computational statistics, and by embracing the synthetic method, as it would be required by the interdisciplinary nature of the categorization problem.

Machine Learning and Cognitive Plausibility of Representations

Classification algorithms strongly depend on the kind of classes' representation that they infer from data, known as *concepts description* (Witten, Frank, 2005; pg. 42) and that they then use to classify new instances. In fact, in the field of machine learning, one can distinguished different family of classifier systems according to the kind of used representations (e.g. instances, decision trees, logical predicates, support vectors, etc.). As it is known in cognitive psychology, the instance-based representation is the only one that coheres with both the prototypes and exemplars theories and therefore, it is the representation to be used in accordance with the *typicality view*. Instead, the most used type of knowledge representation in the machine learning is the one based on rules or decision trees: *"Induction of decision trees is probably the most extensively researched method of machine learning used in data mining"* (Witten, Frank, 2005; pg. 199), although these kinds of representations lack of a true cognitive plausibility, in fact they can be thought as models of the classical theory of categorization, since they represent concepts as logical predicates. As matter of fact, many researches in machine learning as well as machine learning handbooks completely neglect the connections with cognitive psychology and ignore concepts theories, or they do it, let say, in a "superficial" manner. This attitude is well exemplified by Witten and Frank who affirm, with regard to the different possible categories representations that: *"instances do not really «describe» the patterns in data"* (Witten, Frank, 2005; pg. 79) and with regard to the instances based categories learning that: *"in a sense this violates the notion of «learning»"* (Witten, Frank, 2005; pg. 79). This position, followed till its extreme consequences, leads to the paradoxical idea that humans, since represent categories by instances, do not have real learning abilities and do not really have concepts; conversely these abilities are hold only by machines that represent the classes in a anti-psychological way, as for example, with rules and decisions trees. Machine learning researches underestimate possible theoretical and applicative involvements with cognitive sciences although it seems natural that who studies artificial learning of categories should do it in parallel with, or at least not ignoring, the studies about natural learning of categories.

Concluding Remarks

In the previous sections we put in evidence how the "mono-disciplinary" use of cognitive psychology and machine learning produces disappointing results. In fact, from a hand, cognitive psychology produced thirty years of an unfruitful prototypes-exemplars diatribe, which could be avoided if one had not limited oneself to a superficial use of mathematics for the development of cognitive theories, instead of a more foundational use of it, based on machine learning and synthetic method. On the other hand, the field of machine learning disdains the experimental evidences produced by the psychological research. These errors have to be ascribed to a very disciplined and closed methodological praxis, inside the respective scientific communities, in an almost "corporatist" way. Instead, the problem of categorization, as for many of the problems dealt in cognitive sciences, is the same whether one considers natural systems, as human minds, or artificial systems so an interdisciplinary approach in the study of categorization is the natural setting to conduct researches and it is able to progress both field.

References

Cordeschi, R. (2001), *The Discovery of the Artificial. Behavior, Mind and Machines Before and Beyond Cybernetics*. Kluwer, Dordrecht.

Duda, R., Hart, P., Stork, D. (2000) *Pattern Classification (2nd Ed)*. John Wiley & Sons, New York, NY.

Gagliardi, F. (2008) *A Prototype-Exemplars Hybrid Cognitive Model of "Phenomenon of Typicality" in Categorization.* In Proceedings of the 30th Conference of the Cognitive Science Society. Austin, TX. Pp. 1176–1181.

Minda, J.P., Smith, J.D. (2002) *Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation*. Journal of Experimental Psychology: Learning, Memory, Cognition. 28:275–292.

Murphy, G.L. (2002). *The big book of concepts*. MIT Press, Cambridge, MA.

Rosseel, Y. (2002) *Mixture models of categorization.* Journal of Mathematical Psychology. 46: 178-210.

Vanpaemel, W., Storms, G., Ons, B. (2005) *A Varying Abstraction Model for Categorization*. In Proceeding of the XXVII Conference of the Cognitive Science Society. Pp. 2277-2282.

Witten, I.H., Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. *(2nd Ed.)* Kaufmann, San Francisco, CA.

Zaki, S.R., Nosofsky, R.M., Stanton, R.D., Cohen, A.L. (2003) *Prototype and Exemplar Accounts of Category Learning and Attentional Allocation: A Reassessment*. J Exp Psychol Learn Mem Cognit. 29(6), pp. 1160–1173