The Levels of Understanding framework, revised

Tomaso Poggio

The *Levels of Understanding* framework, revised.

Tomaso Poggio
McGovern Institute for Brain Research, Center for Biological & Computational Learning, Department of Brain & Cognitive Sciences,
Massachusetts Institute of Technology,
Cambridge, MA 02139, U.S.A.

Abstract *I discuss the "levels of understanding" framework described in Marr's Vision and propose a revised and updated version of it to capture the changes in computation and neuroscience over the last 30 years.*

1. The "levels of understanding" manifesto

The levels of understanding manifesto is one of the best known in Marr's *Vision (*Marr 2009): it has been mentioned as one of the most enduring constructs of twentieth century cognitive science and computational neuroscience (Wilems 2011). The argument is that a complex system -- like a computer and like the brain -- should be understood at several different levels. For the purpose of this note, let me list just three levels: *the hardware, the algorithms and the computations*. In the *Vision* book, David emphasizes that explanations at different levels are largely independent of each other: a software engineer does not need to know the hardware in any great detail. The message was important at the time, thirty years ago: the study of the problems to be solved -- and of the associated computations -- is relevant in its own right and is needed for a full understanding of the brain. I will argue, however, that it is now time to re-emphasize the connections between levels and to extend the range of levels, if we want to make progress in computational neuroscience.

To explain let me recount the background of the argument (see also Poggio, 1981, 2010). The section in the *Vision* book about levels of understanding is directly based on a paper (Marr and Poggio, 1977) that we wrote together for a booklet of NRP (the influential Neuroscience Research Program founded at MIT by Frank Schmitt). That paper was the original "manifesto" of our computational approach to the brain. Its content was a summary of long discussions that David and I had in the spring of '76 about the levels of analysis of a complex system. We started from an argument described in a paper (Reichardt, and Poggio 1976) on the visual system of the fly by Werner Reichardt and myself. We distinguished the three levels of single cells and circuits, of algorithms and of behavior (of the organism). David insisted, correctly, in replacing (in Marr and Poggio (1977)) the behavior level with the level of computation and computational analysis. This was very important for defining the approach of computational neuroscience. One key aspect of the original argument in Reichardt, and Poggio (1976) however almost disappeared in the process. In

Reichardt, and Poggio (1976) we stressed that one ought to study the brain at different levels of organization, from the behavior of a whole animal to the signal flow, i.e. the algorithms, to circuits and single cells. In particular, we expressed our belief -- and Werner had written about it even earlier -- that a) insights gained on higher levels help to ask the right questions and to do experiments in the right way on lower levels and b) it is necessary to study nervous systems at all levels simultaneously. From this perspective, the importance of coupling experimental and theoretical work in the neurosciences follows directly: without close interaction with experiments, theory is very likely to be sterile.

I believe that David would also think that it is time to look again at the levels of understanding framework -- now *emphasizing the connections between levels* and their synergies. In particular, I believe that neuroscience can help computational theory and even computer science as suggested by recent models of visual cortex, which are leading to interesting approaches in computer vision. In '79, when David wrote *Vision*, our belief was that computational theories may help neuroscientists. The rise of computational neuroscience during the last several years showed that this has happened. Importantly, the table is now turning: in the near future, neuroscience may well be providing new ideas and approaches to Artificial Intelligence.

2. Levels of understanding: a revision?

There is a surprising omission in Marr's *Vision* quest to understand intelligence and the brain, which I am sure would have been corrected if David would have lived longer. Of course it is important to understand the computations and the representations used by the brain – this is the main objective of the book -- but it is also important to understand how an individual organism -- and in fact a whole species – learns and evolves them from experience of the natural world. One could even argue that a description of the learning algorithms and their *a priori* assumptions is deeper and more useful than a description of the details of *what* is actually learned. I have been arguing for the last two decades that the problem of learning is at the core of the problem of intelligence and of understanding the brain. Learning, I think, should have been included *explicitly* in Turing's operational definition of intelligence – his famous Turing test. Not surprisingly, the language of modern statistical learning, including regularization, SVMs, graphical models, hierarchical Bayes, which is permeating various areas of computer science, is at the basis of many recent success stories in building intelligent artifacts such as DeepBlue, Watson, Siri and MobilEye, and is a key component of today's computational neuroscience. It is worth noticing that knowing how to learn to do a specific set of computations -- say visual recognition of objects -- allow us to replicate the ability of recognition in machines, even without an understanding of which algorithms and explicit constraints are exploited! Of course, David Marr's (1969, 1970, 1971) early work was on theories of associative memories (cerebellum, hippocampus and neo-cortex) which are still very interesting today. Here I propose --

and I am sure David would agree -- that *learning* should be added to the list of levels of understanding, above the computational level.

3. Revising the levels of understanding: what it means

I think that any attempt to provide a complete list of all levels and sublevels of understanding is arbitrary -- to say the least. Most taxonomies are. Notice that we had four levels in the original paper with David but I focused here on three of them. At least since Goedel, we have to accept that taxonomies of intelligence have a certain degree of personal taste and internal inconsistency. Several of my friends had logically compelling arguments against adding other levels to the original ones and about organizing a taxonomy in different ways. These alternative revisions of the levels of understanding taxonomy may in fact be logically better (see remarks 1 and 2). In this revision however, my main motivation is to capture is to capture what "understanding" intelligence means for today's science.

Science is a social phenomenon: what is accepted as a good explanation or proof has changed considerably through the last three centuries. In the same way, what people call "understanding" changes quickly. Think about molecular biology in its first fifty years or so: identifying a gene as responsible for a disease or a phenotype was a perfectly acceptable "understanding". Today, we know that genes are not binary switches and are usually part of complicated networks of chemical reactions: the type of required understanding is changing. When David and I wrote the levels of understanding essay, he was really adding a new *frame of mind* -- a new level of explanation -- to older and more classical ones. The new one was the computational level, appropriate for the new science of computational vision and eventually for computational neuroscience. The "old" frames of mind were hardware, circuits and even algorithms -- the frames of mind of physicists and electronic engineers. In the last three decade, the frame of mind of our society of scientists has progressed again in a Hegelian-like helix of increasing sophistication. Theoretical and practical successes -- for instance in vision for cars -- of Machine Learning suggests that it is possible to solve difficult problems in vision (and in intelligence) without a computational and algorithmic understanding. This is (often) an oversimplification but it is also not completely far off. It implies that some researchers may argue -- and they do -- that understanding at the level of learning is sufficiently powerful to solve a problem and therefore perfectly adequate as an explanation *all by itself.* The same argument could be used at the higher level of evolution: knowing how to evolve learning machines frees one from having to understand specific learning modules. Of course full understanding for a scientist requires understanding at all the different levels, of a physicist, a chemist, a computer scientist, a machine learning aficionado and an evolutionist. More importantly, going back to one of the main points in Reichardt and Poggio (1976), I would argue that todays theories and models of the brain should be testable at multiple levels. For instance, a theory of vision should address the computational, algorithmic and biophysical levels and make predictions at all of them. It should also address how objects are learned for recognition and describe how circuits and cells learn from visual experience during development. Ideally, the theory should also suggest plausible ways in which evolution could have

evolved the computations and the modules associated with the ventral stream. To address so many different levels of understanding and how they interact is a tall order for any model but it is a necessary prerequisite for any serious theory of intelligence.

4. *Levels of understanding revised: learning*

On the computational side, the theory of statistical learning has achieved a remarkable degree of completeness and of practical success. Within it, many interesting problems remain open and are a fertile ground for interesting and useful mathematics. I am not discussing here theoretical frameworks other than statistical learning theory in the style of Vapnik and Smale, around the key tools of regularization and Reproducing Kernel Hilbert Spaces. The reasons are that I do not believe that other theoretical frameworks for learning are relevant for understanding learning in the brain, at least in this discussion. In particular, 1) PAC learning has been effectively subsumed by the more powerful and deeper statistical learning theory; 2) a strict Bayesian theory of learning is useful in terms of providing a unifying framework for a phenomenological description of intelligent behaviors but as a theory of learning it does not lead directly to hard results on sample complexity and generalization (Epsilon and Delta, 2012); (3) Deep Learning Networks are likely to be just a distraction despite their current popularity, for the same reasons spelled out by David Marr (1975) (*"Neural net theory.... This combines the limitations of the two previous theories...Again, the primary unresolved issue is what functions you want implemented, and why. In the absence of this knowledge, a neural net theory, unless it is closely tied to the known anatomy and physiology of some part of the brain and makes some unexpected and testable predictions, is of no value.).*

Here I want to take a broad perspective and ask: what is next in the theory of statistical learning *from the point of view of understanding intelligence*? One could for instance argue that the most important aspect of intelligence and of the amazing performance of real brains is in fact the ability to learn. How then does statistical learning theory, as developed so far, compare with brains? One of the most obvious differences, is the ability of people and animals to learn from very few examples. A comparison with real brains offers another, and probably related, challenge to learning theory (Poggio and Smale 2003). Regularization-based "learning algorithms" correspond to one-layer architectures. Are hierarchical architectures with more layers justifiable in terms of learning theory?

4.1 *Hierarchical architectures and learning*

It seems that learning theory does not offer any general argument in favor of hierarchical learning machines for regression or classification. This is somewhat of a puzzle since the organization of cortex – for instance visual cortex – is strongly hierarchical. At the same time, hierarchical learning systems show superior performance in several engineering applications. Very recent theoretical work (Poggio, et al 2011) suggests that the hierarchical architecture of visual cortex

follows from the need to learn, during development, a representation which simplifies the supervised classification problem. The hierarchy is an important preprocessing stage, learned in an unsupervised way during development, transforming images into a representation which should provide a much lower sample complexity. The argument is as follows. The theory assumes that the main problem for biological object recognition is invariance to geometric image transformations. To motivate this assumption let me estimate whether the cardinality of the universe of possible images generated by an object originates more from intraclass variability – eg different types of dogs – or more from the range of possible viewpoints – including scale, position and rotation in 3D. Assuming a granularity of a few minutes of arc in terms of resolution and a visual field of say 10 degrees, one would get > 10^3 different images of the same object from x, y translations, another factor of > 10^3 from rotations in depth, a factor of > 10 from rotations in the image plane and another factor of > 10 from scaling. This gives on the order of at least 10^8 distinguishable images for a single object. On the other hand, how many different distinguishable (for humans) types of dogs exist within the "dog" category? It is unlikely that there are more than, say, $10^2 - 10^3$. From this point of view, it is a much greater win to be able to factor out the geometric transformations than the intracategory differences. The argument then is that main computational goal of the ventral stream is to learn (during development) and discount (during adult vision) image transformations. It is in fact possible to prove that hierarchical, multilayer architectures can compute signatures of objects that are invariant for *Aff (2;R)* (the affine group in R^2), both locally (for parts) and globally (for wholes).

There are a few general points that would follow if the theory were true for the brain. First, I conjecture that the above theoretical justification of *hierarchical architectures may be be more general*. Hierarchies would be built during *development* of a brain and they would represent the result of stages of unsupervised learning of invariant features -- in the case of vision by learning how images transform (continuous motion in time is the principle used to discover transformations which do not change identity) and how to compute "signatures" that are invariant to them. Invariant features would then make the classifier task much easier in terms of sample complexity because they would factor out much of the variability due to local and global transformations. In general terms, this is in itself a conjecture for learning theory. The intuition is the following. Consider distinguishing a specific face from other ones (thus this is a recognition problem of the *identification* type and not of the *categorization* type) with a distance-based classifier. Assume that faces of different individuals are separated in distance more than the noise in the data. Then the sample complexity, e.g. in this case the number of examples needed for correct classification, goes from about 10^8 (no invariant representation) to 10^0 (representation invariant to pose). Thus the tentative answer to the Poggio and Smale (2003) puzzle above is that hierarchical architectures may be needed to solve a key issue in learning theory: how to learn a "good" representation. In the vision case the representation is learned in an unsupervised mode and is "good" in the sense that is invariant to transformations in the world that change the image but not the identity, thereby reducing the sample complexity of recognition.

Second, the additional assumption of online, Hebbian-like learning mechanisms in various visual areas, implies that in the theory the tuning of neurons in the ventral stream (V1, V2, V4 and IT) depends qualitatively and quantitatively on the spectrum of the covariance of the learned transformations for different receptive fields sizes. Thus the theory predicts that the ventral system should be a mirror of the symmetry/invariance/conservation properties (in the sense of physics) of image transformations in the physical world.

It is thus natural to ask whether also in other sensory modalities, such as the auditory system, and possibly in other aspects of intelligence in the brain, *the tuning properties of neurons reflect desirable invariance to transformations of the input signals learned in an unsupervised way during development in order to reduce the complexity of recognition.*

5. *Levels of understanding revised: evolution*

If it is important to understand how an individual organism learns to perform useful recognition from experience of the natural world, it is clearly important to understand how a species evolves the learning algorithms themselves. If we could understand how to evolve learning, we would have a way to obtain a learning machine even without an understanding of which specific algorithms are used to learn and what will be learned. There is an interesting algorithmic question here: how do mechanistic pieces assemble into a whole that serves to provide innovative solutions to complex problems as diverse as smell, speech, vision ... whose solutions characterize intelligence? From the point of view of statistical learning theory, it seems that the incredible effectiveness with which humans (and many animals) learn from and perform in the world *cannot* result only from superior learning algorithms, but *also* from a huge platform of knowledge and priors. This is right in the spirit of Marr's computational approach: constraints, "discovered" by evolution, allow the solution of the typically ill-posed problems of intelligence. Thus evolution is responsible for intelligence, and should be at the top of our levels of understanding. The key questions here are specifically about the path taken by evolution to develop intelligence. For instance: *did intelligence, as the ability to learn, evolve from associative reflexes and memories with the addition of (neurally simple) primitive operations such as composition of different memories?*

6. *Levels of understanding: the 2012 revised list*

The revised list is then:

Levels of Understanding (1977--2012)

Evolution

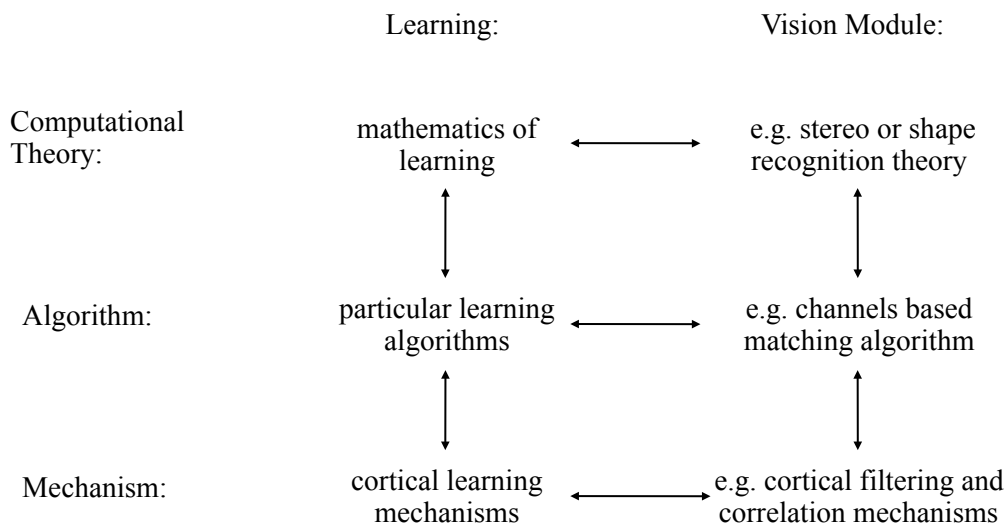Learning and development

Computation

Algorithms

Wetware, hardware, circuits and components

In conclusion, not only we need to know how transistors and synapses work, which algorithms are used for computations such as stereo and what are the goals and the constraints of computations such as object recognition but we also need to understand also how a child may learn recognition and how evolution could have discovered the learning algorithms themselves. In short, in addition to the "old, classical" levels, we also need to understand what is the role of nature and nurture in the development of intelligence. Only then we may be able to develop intelligent machines that could learn to see – and think -- without the need to be programmed to do it.

Remarks

(1) Keith Nishihara writes: Consider these questions: Is there a computational theory for learning? Are there algorithms for learning? Are there mechanisms in the neuroanatomy for learning? I think the answer to these is yes. I think it might be more useful to think of learning as having a computational, algorithmic and mechanisms level, as in the diagram below:

	Learning:		Vision Module:
Computational Theory:	mathematics of learning	⟷	e.g. stereo or shape recognition theory
	↕		↕
Algorithm:	particular learning algorithms	⟷	e.g. channels based matching algorithm
	↕		↕
Mechanism:	cortical learning mechanisms	⟷	e.g. cortical filtering and correlation mechanisms

(2) Joel Leibo writes: What is the difference between the learning and evolution levels? Are they the same as one another? Isn't evolution just a specific learning algorithm?   If they only differ in algorithm, then aren't they just different kinds of answers that you might give to questions on the 2nd (algorithmic) level?...I do not see how you could give a complete description of a complex system on the "learning" or "evolution" level. I do not immediately see how these "levels" could be added without radically revising the meaning of "level" in this context.

REFERENCES

Epsilon T., and Delta L., 2013 "The Bayes Delusion" *In preparation*.

Marr D., 2010 *Vision* (Cambridge: MIT Press (reissued))

Marr D., 1969 "A theory of cerebellar cortex." *J. Physiol.*, 202:437-470.

Marr D., 1970 "A theory for cerebral neocortex." *Proceedings of the Royal Society of London B*, 176:161-234.

Marr D., 1971 "Simple memory: a theory for archicortex." *Phil. Trans. Royal Soc. London*, 262:23-81.

Marr D., 1975 "Approaches to Biological Information Processing", *Science,* Vol. 190, No. 4217 pp. 875-876.

Marr D., Poggio T., 1977 "From Understanding Computation to Understanding Neural Circuitry" In: Neuronal Mechanisms in Visual Perception, E. Poppel, R. Held and J.E. Dowling (eds.), Neurosciences Res. Prog. Bull., 15, 470-488.

Poggio T., 1981 "Marr's Computational Approach to Vision", *Trends in Neurosciences*, 10, 258-262.

Poggio T., 2010 "Afterword by Tomaso Poggio" re: Marr, David, Vision, MIT Press, ISBN-10: 0-262-51462-1; ISBN-13: 978-0-262-51462-0 1982, reissued July, 2010; see also http://cbcl.mit.edu/publications/ps/marr-vision-afterword.pdf

Poggio T., (sections with Mutch J., Leibo JZ and Rosasco L.), 2011 "The Computational Magic of the Ventral Stream: Towards a Theory" *Nature Precedings*, doi:10.1038/npre.2011.6117.1

Poggio T., Smale S., 2003 "The Mathematics of Learning: Dealing with Data", *Notices of the American Mathematical Society (AMS)*, Vol. 50, No. 5, 537-544.

Reichardt W., Poggio T., 1976 *"*Visual Control of Orientation Behavior in the Fly: A Quantitative Analysis" *Quarterly Review of Biophysics*, 3, 311-375.

Wilems RM, 2011 " Re-appreciating the why of cognition: 35 years after Marr and Poggio" *Frontiers in Psychology*, 2 1-5.