

# Turing<sup>++</sup> Questions: A Test for the Science of (Human) Intelligence

Tomaso Poggio and Ethan Meyers

Center for Brains, Minds and Machines

MIT, Cambridge, MA, USA

## Summary

It is becoming increasingly clear that there is an infinite number of definitions of intelligence. Machines that are intelligent in different narrow ways have been built since the 50s. We are entering now a *golden age for the engineering of intelligence* and the development of many different kinds of intelligent machines. At the same time there is a widespread interest among scientists in understanding a specific and well defined form of intelligence, that is *human intelligence*. For this reason we propose a stronger version of the original Turing test. In particular, we describe here an open-ended set of *Turing<sup>++</sup> Questions* that we are developing at the Center for Brains, Minds and Machines at MIT — that is questions about an image. Questions may range from *what is there* to *who is there*, *what is this person doing*, *what is this girl thinking about this boy* and so on. The plural in questions is to emphasize that there are many different intelligent abilities in humans that have to be characterized, and possibly replicated in a machine, from basic visual recognition of objects, to the identification of faces, to gauge emotions, to social intelligence, to language and much

more. Recent advances in cognitive neuroscience has shown that even in the more limited domain of visual intelligence, answering these questions requires different competences and abilities, often rather independent from each other, often corresponding to separate modules in the brain. The term Turing<sup>++</sup> is to emphasize that our goal is understanding human intelligence at all Marr's levels — from the level of the computations to the level of the underlying circuits. Answers to the Turing<sup>++</sup> Questions should thus be given in terms of models that match human behavior and human physiology — the mind and the brain. These requirements are thus well beyond the original Turing test. A whole scientific field that we call the *science of (human) intelligence* is required to make progress in answering our Turing<sup>++</sup> Questions. It is connected to neuroscience and to the engineering of intelligence but also separate from both of them.

### **Definitions of Intelligence**

We may call a "person" intelligent and even agree among us. But what about a colony of ants and their complex behavior? Is this intelligence? Were the mechanical computers built by Turing to decode the encrypted messages of the German U-boats, actually intelligent? Is Siri intelligent? The truth is that the question of *What is intelligence* is kind of ill-posed as there are many different answers, an infinite numbers of different kinds of intelligence. This is fine for engineers who may be happy to build many different types of intelligent machines. The scientists among us may instead prefer to focus on a question that is well defined and can be posed in a scientific way, on the question of *human intelligence*. In the rest of the paper we use the term intelligence to mean human intelligence.

## **Understanding Human Intelligence**

Consider the problem of visual intelligence. Understanding such a complex system requires understanding it at different levels (in the Marr sense, see Poggio 2012), from the computations to the underlying circuits. Thus we need to develop algorithms that provide answers of the type humans do. But we really need to achieve more than just simulate the brain's output, more than what Turing asked. We need to understand what understanding an image by a human brain means. We need to understand the algorithms used by the brain, but we also need to understand the circuits that run these algorithms. This may also be useful if we want to be sure that our model is not just faking the output of a human brain by using a giant look-up table of what people usually do in similar situations, as hinted at the end of the movie *Ex Machina*. Understanding a computer means understanding the level of the software *and* the level of the hardware. Scientific understanding of human intelligence requires something similar — understanding of the mind *as well as* of the brain.

### **Using behavior and physiology as a guide**

In order to constrain our search for intelligent algorithms, we are focusing on creating computational models that match human behavior and neural physiology. There are several reasons why we are taking this approach. The first reason, as hinted above, is to avoid superficial solutions that mimic intelligent behavior under very limited circumstances, but that do not capture the true essence of the problem. Such superficial solutions have been a prominent approach to the traditional Turing Test going back to the ELIZA program written in the 1960's, (Weizenbaum 1966). While these approaches might occasionally fool humans, they do not address many of the fundamental issues and thus this approaches will fail to match many aspects of human behavior. A second related reason is that algorithms might appear to be perform well when tested under limited circumstances, but when compared to the full range of human abilities they might not do nearly as well. For example, deep neural networks very well on object recognition tasks, but also fail in simple ways that would never

been seen in human behavior (Szegedy, et al, 2006). By directly comparing computer systems' results to human behavioral results we should be able to assess whether a system that is displaying intelligent behavior is truly robust (Sinha et al, 2006). A final reason is that studying primate physiology can give us guidance about how to approach the problem. For example, to recognize people based on their faces appears to occur in discrete face patches in the primate brain (see Freiwald and Tsao 2010, and section below). By understanding the computational roles of these patches we aim to understand the algorithms that are used by primates to solve these tasks (Meyers et al, 2015).

### **Human Intelligence is one word but many problems**

Recent advances in cognitive neuroscience have shown that different competencies and abilities are needed to solve visual tasks, and that they seem to correspond to separate modules in the brain. For instance, the apparently similar questions of object and face recognition (what is there vs who is there) involve rather distinct parts of visual cortex (e.g., the lateral occipital cortex vs. a section of the fusiform gyrus). The word intelligence can be misleading in this context, like the word life was during the first half of the last century when popular scientific journals routinely wrote about the problem of life, as if there was a single substratum of life waiting to be discovered to completely unveil the mystery. Of course, speaking today about the problem of life sounds amusing: biology is a science dealing with many different great problems, not just one. Thus I think that intelligence is one word but many problems, not one but many Nobel prizes. This is related to Marvin Minsky's view of the problem of thinking, well captured by the slogan Society of Minds. In the same way, a real Turing test is a broad set of questions probing the main aspects of human thinking. Because "Intelligence" encompasses a large set of topics, we have chosen Visual Intelligence in human and non-human primates as a primary focus. Our approach at the Center for Brains, Minds and Machines (CBMM) to Visual Intelligence includes connections to some developmental, spatial, linguistic, and social questions. To further sharpen our

focus, we are emphasizing measuring our progress using questions, described in more detail below, that might be viewed as extensions of the Turing test. We have dubbed these Turing<sup>++</sup> Questions. Computational models we develop will be capable of responding to queries about visual scenes and movies – who, what, why, where, how, with what motives, with what purpose, and with what expectations. Unlike a conventional engineering enterprise that tests only absolute (computational) performance, we will require that our models exhibit *consistency with human performance/behavior, with human and primate physiology, and with human development*. The term Turing<sup>++</sup> refers to these additional levels of understanding that our models and explanations must satisfy.

### **The Turing<sup>++</sup> Questions**

Our choice of Questions follows in part from our understanding of human intelligence grounded in the neuroscience of the brain. Each question roughly corresponds to a distinct neural module in the brain. We have begun defining an initial set of such problems/questions about visual intelligence, since vision is our entry point into the problem of intelligence. We call such questions Turing<sup>++</sup> Questions because they are inspired by the classical Turing test but go well beyond it. Traditional Turing Tests permit counterfeiting and require matching only a narrowly defined level of human performance. Successfully answering Turing<sup>++</sup> Questions will require us not only to build systems that emulate human performance, but also to ensure that such systems are consistent with our data on human behavior, brains, neural systems, and development. An open-ended set of Turing<sup>++</sup> Questions can be effectively used to measure progress in studying the brain-based intelligence needed to understand images and video.

As an example consider the image shown below. A deep learning network might locate faces and people. One could not interrogate such a network, however, with a list of Turing<sup>++</sup> Questions such as these:

- *What is there?*
- *Who is there?*
- *What are they doing?*
- *How, in detail, are they performing actions?*
- *Are they friends or enemies or strangers?*
- *Why are they there? What will they do next?*
- *Have you seen anything like this before?*



**Figure 1**

Courtesy of Boris Katz, CBMM, from the LableMe data base.

We effortlessly recognize objects, agents, and events in this scene. We, but not a computer program, could recognize that this is an amusement park; several people are walking; there is a stroller in front of the fence; two women are carrying bags; very few people, if any, are riding the carousel. We, but not a computer program, could generate a narrative about the scene. It's a fairly warm, sunny day at the amusement park. A blonde young mother or caregiver in rolled-up blue jeans is waiting, presumably with a baby, by the carousel. One or two friends may be walking up to meet her.

We would assess the performance of a model built to answer questions like these by evaluating a) how similarly to humans our neural models of the brain answer the questions, and b) how well their implied physiology correlates with human and primate data obtained by using the same stimuli.

Our Turing<sup>++</sup> Questions require more than a good imitation of human behavior; our computer models should also be human-like at the level of the implied physiology and development. Thus the CBMM test of models uses Turing-like questions to check for human-like performance/behavior, human-like physiology, and human-like development.

Because we aim to understand the brain and the mind and to replicate human intelligence, the challenge intrinsic to the testing is not to achieve best absolute performance, but performance that correlates strongly with human intelligence measured in terms of behavior and physiology. We will compare models and theories with fMRI and MEG recordings, and will use data from the latter to inform our models. Physiological recordings in human patients and monkeys will allow us to probe neural circuitry during some of the tests at the level of individual neurons. We will carry out some of the tests in babies to study the development of intelligence.

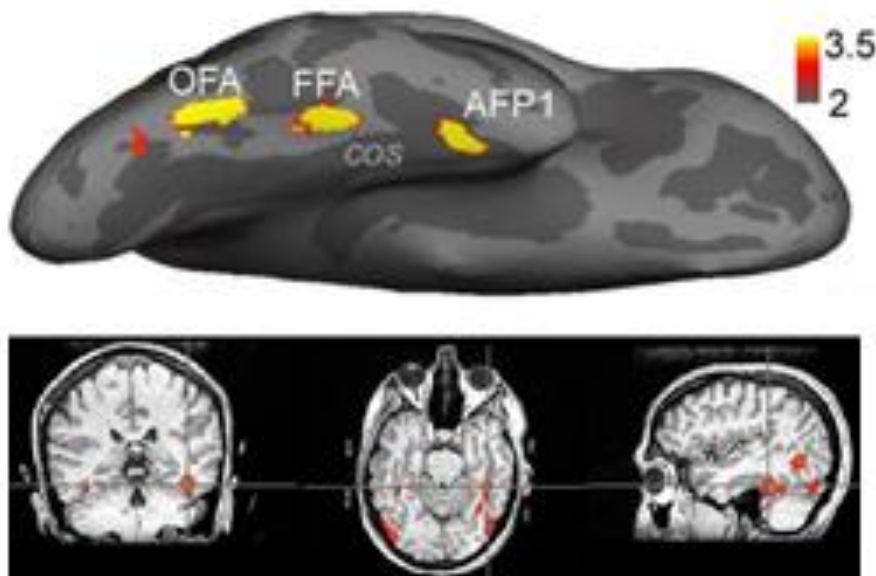
The series of tests is open-ended. The initial ones, e.g. face identification, are tasks that computers are beginning to do and where we can begin to develop models and theories of how the brain performs the task. The later ones, e.g. generating stories explaining what may have been going on in the videos and answering questions about previous answers, are goals for the next few years of the Center and beyond.

The modeling and algorithm development will be guided by scientific concerns, incorporating constraints and findings from our work in cognitive development, human cognitive neuroscience, and systems neuroscience. These efforts likely would not produce the most effective AI programs today (measuring success against objectively correct performance); the core assumption behind this challenge is that by

developing such programs and letting them learn and interact, we will get systems that are ultimately intelligent at the human level.

### **An example of a Turing<sup>++</sup> Question: whois there, e.g. face identification**

The Turing<sup>++</sup> Question which is most ripe, in the sense of possibility to answer it at all the required levels, is face identification. We have data about human performance in face identification — from a field which is called psychophysics of face recognition. We know which patches of visual cortex in humans are involved in face perception by using fMRI techniques as shown in Figure 2.



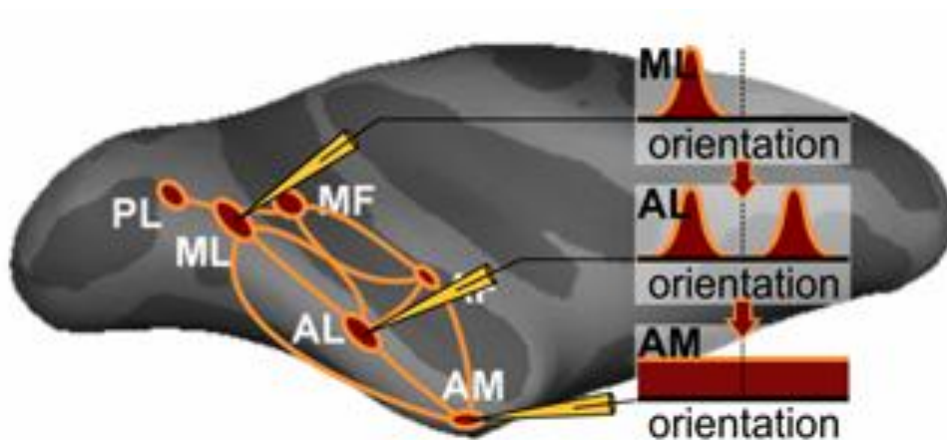
**Figure 2**

Courtesy of Nancy Kanwisher, CBMM. Modified from: Tsao DY, Moeller S, Freiwald WA. Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;105(49):19514–9.



We can identify the homologue areas in the visual cortex of the macaque where there is a similar network of interconnected patches shown in Figure 4. In the monkey it is possible to record from individual neurons in the various patches and characterize their properties. Neurons in patch ML are view and identity tuned, neurons in AM are identity specific but more view invariant. Neurons in the intermediate patch AL tend to be mirror symmetric: if they are tuned to a view they are also likely to be tuned to the symmetric one.

We begin to have models that perform face identification well and are consistent with the architecture and the properties of face patches (i.e., we can make a correspondence between stages in the algorithm and properties of different face patches). The challenge is to have performance that correlates highly with human performance on the same data sets of face images and that predict the behavior of neurons in the face patches for the same stimuli.



**Figure 3**

Courtesy of Winrich Freiwald, CBMM. Modified from: Tsao DY, Moeller S, Freiwald WA. Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;105(49):19514–9.

CBMM is organizing in September 2015 the first Turing<sup>++</sup> Questions workshop, focused on face identification. The title of the workshop is *A Turing<sup>++</sup> Question: Who is there?*. The workshop will introduce databases and review the states of existing model to answer the question *who is there* at the levels of performance and neural circuits.

## **The Science of Intelligence**

For the Center for Brains Minds and Machines the main research goal is the *science of intelligence* rather than the engineering of intelligence — the hardware and software of the brain rather than just absolute performance in face identification. Our *Turing<sup>++</sup> Questions* reflect fully these research priorities.

The emphasis on answers at the different levels of behaviour and neural circuits reflects the levels of understanding paradigm (Marr 2010). The argument is that a complex system -- like a computer and like the brain/mind -- must be understood at several different levels, such as hardware and algorithms/computations. Though Marr emphasizes that explanations at different levels are largely independent of each other, it has been argued (Poggio, 2012) that it is now important to re-emphasize the connections between levels, which was described in the original paper about levels of understanding (Marr and Poggio, 1977). In that paper we argued that one ought to study the brain at different levels of organization, from the behavior of a whole organism to the signal flow, i.e. the algorithms, to circuits and single cells. In particular, we expressed our belief 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. Otherwise there are not enough constraints for a unique solution to the problem of human intelligence.

## REFERENCES

- Freiwald, W. A., and Tsao, D. Y. 2010. Functional Compartmentalization and Viewpoint Generalization Within the Macaque Face-Processing System. *Science* 330: 845– 851.
- Marr, D. 2010. *Vision*. Cambridge, MA: The MIT Press.
- Marr, D., and Poggio, T. 1977. From Understanding Computation to Understanding Neural Circuitry. In *Neuronal Mechanisms in Visual Perception*, ed. E. Poppel, R. Held, and J. E. Dowling. *Neurosciences Research Program Bulletin* 15: 470–488.
- Meyers, E.; Borzello, M.; Freiwald, W.; Tsao, D. 2015. Intelligent Information Loss: The Coding of Facial Identity, Head Pose, and Non-Face Information in the Macaque Face Patch System. *Journal of Neuroscience* 35(18): 7069–81.
- Poggio, T. 1981. Marr's Computational Approach to Vision. *Trends in Neurosciences* 10(6): 258–262.
- Poggio, T. 2012. The Levels of Understanding Framework, Revised. *Perception* 41(9): 1017–1023.
- Reichardt, W., and Poggio, T. 1976. Visual Control of Orientation Behavior in the Fly: A Quantitative Analysis. *Quarterly Review of Biophysics* 9(3): 311–375.
- Sinha, P.; Balas, B.; Ostrovsky, Y.; and Russell, R. 2006. Face Recognition by Humans: 19 Results All Computer Vision Researchers Should Know About. *Proceedings of the IEEE* 94(11): 1948–1962.
- Szegedy, C.; Zaremba, W.; Sutskever, I.; Bruna, J.; Erhan, D.; Goodfellow, I. J.; and Fergus, R. 2013. Intriguing Properties of Neural Networks. *CoRR (Computing Research Repository)*, abs/1312.6199. Association for Computing Machinery.
- Weizenbaum, J. 1966. ELIZA—A Computer Program for the Study of Natural Language Communication Between Man and Machine. *Communications of the ACM* 9(1): 36–45.

## AUTHOR BIOGRAPHIES

Tomaso A. Poggio, is the Eugene McDermott Professor in the Dept. of Brain & Cognitive Sciences at MIT and the director of the new NSF Center for Brains, Minds and Machines at MIT of which MIT and Harvard are the main member Institutions. He is a member of both the Computer Science and Artificial Intelligence Laboratory and of the McGovern Brain Institute. He is an honorary member of the Neuroscience Research Program, a member of the American Academy of Arts and Sciences, a Founding Fellow of AAI and a founding member of the McGovern Institute for Brain Research. Among other honors he received the Laurea Honoris Causa from the University of Pavia for the Volta Bicentennial, the 2003 Gabor Award, the Okawa Prize 2009, the AAAS Fellowship and the 2014 Swartz Prize for Theoretical and Computational Neuroscience. He is one of the most cited computational scientists with contributions ranging from the biophysical and behavioral studies of the visual system to the computational analyses of vision and learning in humans and machines. With W. Reichardt he characterized quantitatively the visuo-motor control system in the fly. With D. Marr, he introduced the seminal idea of levels of analysis in computational neuroscience. He introduced regularization as a mathematical framework to approach the ill-posed problems of vision and the key problem of learning from data. In the last decade he has developed an influential hierarchical model of visual recognition in the visual cortex. The citation for the recent 2009 Okawa prize mentions his "...outstanding contributions to the establishment of computational neuroscience, and pioneering researches ranging from the biophysical and behavioral studies of the visual system to the computational analysis of vision and learning in humans and machines." His research has always been interdisciplinary, between brains and computers. It is now focused on the mathematics of learning theory, the applications of learning techniques to computer vision and especially on computational neuroscience of the visual cortex. A former Corporate Fellow of Thinking Machines Corporation and a former director of PHZ Capital Partners, Inc., is a director of Mobileye and was involved in

starting, or investing in, several other high tech companies including Arris Pharmaceutical, nFX, Imagen, Digital Persona and Deep Mind.

Ethan Meyers is an Assistant Professor of Statistics and Hampshire College. He received his BA from Oberlin College in Computer Science, and his PhD in Computational Neuroscience from MIT. His research examines how information is coded in neural activity, with a particular emphasis on understanding the processing that occurs in high level visual and cognitive brain regions. To address these questions, he develops computational tools that can analyze high dimensional neural recordings.