

The Grand Challenge Is Called: Robotic Intelligence

Angel Pasqual del Pobil

Department of Computer Science
Campus Penyeta Roja, Jaume-I University,
E12071 Castellon, Spain
pobil@inf.uji.es

Abstract. The role of perception and action in current AI systems is analyzed and some points concerning present AI methodologies are discussed. It is argued that if we accept as a central goal for AI to build artificial systems that behave in an intelligent way in the real world, then additional evaluation criteria for intelligent systems should be used. Finally, some of these criteria are proposed.

1. Motivation

One of the present grand challenges for AI concerns artificial behavior for agents that have to deal with the real world through perception and motor actions. Nowadays, there exist an enormous lack of balance between existing AI systems in some aspects of their competence. Whereas in some formal microworlds AI systems have reached the highest human level of competence, or there exist knowledge-based systems exhibiting human expert competence in narrow technical domains, few systems exist surpassing the competence of an insect in moving around pursuing a goal in an unstructured world. This enormous distance between pure abstract intellectual tasks at one end, and those that involve sensorimotor interaction with the physical world at the other, calls for an emphasis on research on robotic agents.

The traditional Turing vision of a disembodied, abstract, symbol-processing intelligence has been recently questioned. New proposals —such as those put forward by Harnad or Brooks— are worth consideration: robotic capacities must be added to an extended version of the Turing test and the symbol grounding problem has to be approached by grounding a system's representations in the physical world via sensory devices with the result of emergent functionalities.

It has been customary to relegate issues such as manipulation, vision or robot motion out of the so-called *core* IA. To verify that this point of view is still alive to a great extent, we only have to examine the current AI journals having more impact, and analyze the percentages of published articles. It can be argued, on the contrary, that such issues, relevant to *Robotic Intelligence*, should be situated in the very center of

the core of AI. Consequently, new evaluation criteria are called for to be applied to those AI systems requiring an actual interaction with the world to be fully completed. In the rest of this article arguments of a fundamental as well as a methodological nature are provided in this regard.

2. Extending the Turing Test

Since it was formulated in 1950, the so-called Turing Test (T2) was generally accepted as the right, definitive test to discriminate an intelligent computer from a non-intelligent one [Turing 50]. Basically, in the test there is a person communicating via a teleprinter with a computer and another person, but ignoring what line is connected to the person and what to the computer. If the person cannot tell how the lines are connected after a dialogue through both lines, then the computer is said to have passed T2 and it can be rated as intelligent. It has to be noted that T2 is implicitly assuming that intelligence is the ability to reason and to communicate by means of language.

In 1980 John Searle put forward a thought experiment with the intention of showing that computers cannot really understand what they do [Searle 80]. Since then, it has raised much controversy among the AI community, and discussions still keep on and on as to its correctness. In essence, Searle's Chinese Room argument assumes that in T2 the language used for communicating is Chinese instead of English, and the computer is replaced by a person, Searle himself, called the operator, locked in the Chinese Room. The operator understands no Chinese at all, but he is provided with a set of instructions in English to manipulate Chinese written symbols in such a way that, following these instructions, the operator is able to produce a set of symbols as the output to a given input set of Chinese symbols. Now, if this operator is able to pass T2 in Chinese, we should conclude that he understands Chinese, while the operator, being Searle himself, does not know a single word of this language, on the contrary, all he has done is following the instructions for manipulating meaningless symbols. An immediate consequence of this argument is that T2 is not the definitive test for intelligence: a computer passing it understands no more what it is doing than Searle understands Chinese.

Stevan Harnad has proposed the Total Turing Test (T3) as an extension of T2 that is not invalidated by Searle's argument [Harnad 89]. In T3 the computer is replaced by a robot, and the person carrying out the test is not communicating through a teleprinter but actually seeing (and hearing) the candidate robot and a real person, while both are operating directly on the world. If after a certain amount of time (as long as desired) the observer is not able to say which is the robot and which the person, then the robot has passed the test and it can be rated as intelligent. The key point is that now, in addition to reasoning and communicating by language, the candidate must exhibit all robotic capacities a person has, including the ability to see, grasp,

manipulate, move, hear, recognize, etc., in a way that is indistinguishable from those of a person.

Some people seem to be bothered by the fact that T3 would imply a robot with perfect human-like appearance. In the context of Harnad's argument this is rather beside the point, but, by way of relief, we can think of the person disguised as a robot and trying to distinguish the *real* robot from the *false* robot. Moreover, T3 is exactly the common-sense test we perform many times a day as a way of solving in practice the *other minds* problem [Harnad 91], and it is behavior rather than appearance which makes us judge rightly (think of a severely lobotomized person, for instance, versus a person disguised as whatever you may think of).

What really matters in T3 is that robotic capacity has been integrated as an inseparable part of the definitive test for intelligence. If we want an intelligent system to pass T3, it must be endowed with Robotic Intelligence, understood as the subset that is common to both Robotics and AI and that deals with the intelligent connection of perception to action [Brady 85], or to put it another way: the particular class of *artificial*, intelligent systems that are *robotic* as well.

3. Where Are the Symbols of Pure Symbol Systems Grounded?

Classical AI is based on the use of pure symbol systems, i.e., following the traditional distinction between a symbol level and its implementation in a particular computer. The *symbol grounding problem* is yet another challenge to pure symbolic AI [Harnad 90] ([Mira et al. 95] also deals with the same problem in terms of description domains). The symbols in a symbol system are systematically interpretable as meaning something; however, in a typical AI system, that interpretation is not intrinsic to the system, it is always given by an external interpreter (e.g., the designer of the system). Neither the symbol system in itself nor the computer, as an implementation of the symbol system, can ground their symbols in something other than more symbols. The operator in the Chinese Room will never be able to understand Chinese because it is somebody else who knows the interpretation of Chinese symbols, the one who designed the instructions for manipulating them. And yet, when we think, unlike computers, we use symbol systems that need no external interpreter to have meanings. The meanings of our thoughts are intrinsic, the connection between our thoughts and their meanings is direct and causal, it cannot be mediated by an interpreter, otherwise it would lead to an infinite regress if we assume that they are interpretable by someone else.

Again, the solution to this paradox may lie in Robotic Intelligence (RI) systems instead of pure symbolic AI systems [Harnad 93], [Brooks 90]. In an RI system, with T3-level performance, the symbols are grounded in the system's own capacity to interact robotically with what its symbols are about. Such an RI system should be able to perceive, manipulate, recognize, classify, modify, ..., and reason about the

real-world objects and situations that it encounters. In this way, its symbols would be grounded in the same sense that a person's symbols are grounded, because it is precisely those objects and situations that their symbols are about. If we think of a symbol that corresponds to a word, we ground it when we first learn our mother tongue through interaction with the outer world, because we cannot obviously ground it in more words. In this respect, for a blind child the meanings of its symbol system must necessarily differ from those of a normal child, because its interaction with the world is severely handicapped.

A possible answer to the question of how to ground basic spatial concepts is the use of connectionism. Neural nets can be a feasible mechanism for learning the invariants in the analog sensory projection on which categorization is based [Harnad 93]; see also [Cervera & del Pobil, 97], [Cervera & del Pobil, 98], [Mel90], [Martin & del Pobil 94], [Heikkonen 94].

4. Cockroaches vs. Grandmasters

Another point in which AI has failed to take the right orientation is in choosing the appropriate level of competence an AI system should exhibit. This has led to an enormous lack of balance between existing AI systems and natural systems in some aspects of their competence. For example, there are chess-playing systems that are able to reach grandmaster level of competence, only being defeated by a few persons in the world; or expert systems that show an expert competence in, say, diagnosing infectious diseases. And, on the other hand, there are no existing system that surpasses the competence of a cockroach in moving around with a goal in an unstructured world. This enormous distance tends to be always between pure abstract intellectual tasks at one end, and robotic tasks, at the other, i.e., those that involve sensorimotor interaction with the real world. In the case of human level competence, not to speak of cockroaches, the gap between these two levels of competence is still larger. Our simplest, everyday, common-sense robotic capacities are very far from what robots can currently do: our artificial grandmaster would inevitably *die* in case of fire just because it would not be able to find the exit door and turn the handle to open it (turning door handles is one of the tasks that current robots cannot do in a general case).

One characteristic that all natural living systems seem to exhibit is autonomy (though it may be argued that not all living systems are *intelligent* —I think that cockroaches are at least more competent in some respect than AI systems— it is undeniable that all natural intelligent systems are living systems). An interesting approach is the so-called *Artificial Life* [Steels & Brooks 93]. In my opinion, some of its start points go beyond the requirements of intelligence: namely, reproduction, self-maintenance or evolution do not seem to us to be essential features inseparable from intelligence. Autonomy, on the other hand, appears to be hardly separable from full Robotic Intelligence, if autonomy is understood as the ability to conduct itself in an unsupervised way within an unstructured world.

In this respect, Rodney A. Brooks' artifacts [Brooks 90] can be situated at just the opposite end with respect to symbolic AI. He rules out any notion of representation [Brooks 91] to rely exclusively on interfacing to the real world through perception and action. By approaching intelligence in an incremental manner, he pursues the building of robots that exhibit insect level behavior, and proposes this methodology as being closer to the ultimate right track for AI.

5. Reasoning about Perception and Perception-Based Reasoning

Even accepting that perception and action as robotic capacities must play a fundamental role in an intelligent system, there is a frequent misconception regarding approaches to machine perception and motor actuation. Namely, it is usually assumed that sensorimotor abilities can be isolated from the rest of the system and just be implemented as input/output modules that interface to the main processing—or pure abstract reasoning—unit. This is the Turing vision of AI that is usually assumed as a start point. It can be encountered, as such, in classical AI introductory textbooks, in [Charniak & McDermott 85] for instance, an intelligent system is depicted as one consisting of input, output and inside modules.

This *modularity hypothesis* cannot be justified by any evidence whatsoever, neither from neurophysiology, nor from biology or cognitive psychology. It is just a traditional problem-reduction methodology that is systematically used to tackle hard problems. This simplification has led to a vitiated state of affairs: researchers in perception assume that what they do may be useful someday to core AI researches via some kind of interface (or representation), while people working in pure abstract reasoning accept without questioning that the perceptionists will be able to provide the system with the (over)simplified model of the world they are taking for granted.

The assumption about the subsequent development of such an interface is, at least, dangerous, and the predictable difficulties in obtaining it may render useless many of previous research efforts in perception and reasoning. In addition, in natural living systems, cognitive abilities are intimately and inseparably tied to perception and (maybe to a smaller extent) action capacities. Psychophysical evidence suggests this fact and, moreover, that the representation of the world used by an intelligent system is directly dependent on its perception skills [Brooks 91].

This accounts yet for another fact that has attracted the attention of researches in spatial reasoning: we often use spatial concepts to reason about domains of an abstract nature that, in principle, have nothing to do with space [Freksa 91]. For instance, we represent a data structure as a tree, or a hierarchy as a pyramid, or we plot the evolution of whatever variable we are interested in to see how it *goes up or down*, (see [Lakoff 87] for more examples). The explanation seems straightforward: the way we perceive things is conditioning the way we reason to such an extent that we have to express knowledge in terms of spatial notions to be able to understand it.

An immediate consequence is that reasoning models or representations dealing with physical space that include no reference to perception may be misleading and, in addition to reasoning about perception [Musto & Konolige 93], the right approach should be *perception-based reasoning*, because perception is not just a direct source of information about the physical properties of the environment, but it is, in a way that is still to be discovered, tightly integrated with our reasoning procedures and motor actions.

This has been our personal experience after working on three robotics problems from this perspective: gross motion planning for a robot arm refers to the problem of generating its movement so as to avoid collision with obstacles or other robots; fine motion planning deals with motion involving contact with other objects; and in grasp planning we try to figure out how a robot arm with a gripper must stably grasp an object (a complete statement of these problems along with a survey of the state of the art can be found in [del Pobil & Serna 95]). There are many published solutions to these three problems that work with computer simulations: the perception and motor action parts are just taken for granted. If we put aside this assumption to deal with a partially unknown environment by using actual sensors and effectors, then the problems change to such an extent that most of the simulated approaches are no longer applicable [Cervera & del Pobil 97], [Cervera & del Pobil 98], [Sanz et al. 98], [Gupta & del Pobil, 98].

Another consequence of the incorrectness of the modularity hypothesis would be that T3 turns out to be immune to the Chinese Room Argument. Searle himself considers “The Robot Reply”: what would happen if the computer is replaced by a robot, wouldn’t it have genuine understanding?. He answers that the *addition* of perceptual and motor capacities adds *nothing* by way of understanding. To justify this answer he modifies his thought experiment by *adding* a perceptual apparatus and a motor apparatus in such a way that some of the Chinese symbols that come to the operator come from the perceptual apparatus, and some of the Chinese symbols he is giving out serve to make the motor apparatus move [Searle 80]. As in the original version of the argument, the operator doesn’t know what is going on. If the modularity hypothesis is wrong, as [Harnad 89] argues, then this answer makes no sense, since it is assuming that the motor and perceptual capacities can be encapsulated in independent modules that are just added to the main reasoning unit.

6. Some Evaluation Criteria for Intelligent Systems

The replacement of T2 by T3 is not only a question of theoretical foundations, on the contrary, it has profound methodological implications, just as T2 had. Although Turing himself spoke about *the imitation game*, T2 and T3 are scientific criteria: they represent eventual empirical goals to be accomplished scientifically [Harnad 92]. Although nowadays AI is far from building a system that passes T2, it has oriented AI research in the last decades and it is to a great extent responsible for the central role of

core pure abstract symbolic AI. The advent of T3 should imply a change in the focus of AI research to the effect that Robotic Intelligence must lie in its *core*. Although passing T3 is an ultimate goal, moving in the right direction here and now means that not so many things should be taken for granted: we have to turn our faces to the real world as it is perceived.

The previous considerations suggest that new criteria must be added to traditional evaluation criteria for intelligent systems and, moreover, they should partly replace them. As an overall recommendation, T3 should orient research in AI, at least in the same way as T2 and the Turing vision of AI have influenced AI and still do. It is not clear which is the right methodology towards T3, this is an open question, but at least some criteria that are closer to the right direction can be proposed at the moment.

First, we should ask about the role of perception and action in an intelligent system that is embodied and situated inside a physical environment: they may be completely obviated, they may be just taken for granted, they may be simulated, they may be added to the system in a straightforward way at the current state of transducer/effector technologies, they may be fully integrated within the rest of the system, or the system may be built in such a way that its reasoning mechanisms are based on its perception and action capacities. Obviously, all these possibilities would be evaluated in a different way.

Second, we should ask how the symbols in the system are grounded. This question is closely related to the previous one, but it is of a deeper nature: what is the link between symbols and meanings?, is it just the designer of the system who decides what meanings to attach to such and such symbols, or is there a direct, causal way (probably through sensorimotor interaction) to establish these connections?

Third, the level of competence of the system and its underlying assumptions about cognitive modularity should be taken into account. Isolating a certain intelligent capacity does not seem a proper methodology: if we ignore the possible interaction of this capacity with other capacities, and the lower-level capacities on which it is based, representing it in a particular fashion may be completely irrelevant to the ultimate goal of AI as the science of the artificial, and its interest would just be as an instance of a problem-solving methodology. In this respect, Brooks' proposal in the sense of achieving AI through building robotic artifacts seems reasonable, although excluding all kinds of internal symbols may be going too far.

When evaluating the competence of a situated system, a critical question should be: to what extent it interacts with the real world, and if it doesn't, whether it could presumably scale up to a system that really interacts with the world. The autonomy of the system should also be a crucial point, considering at least a somewhat partial autonomy as a first step towards total T3 autonomy.

Finally, we must insist on the fact that taking the biological paradigm seriously implies building on top of an integrated and distributed sensorimotor system, since the coordination of our movement is done mainly in an *unconscious* way, relying on perception without central processors coming into play. Neural networks have proven

to be an adequate paradigm for approaching this kind of problems as well as others at the subsymbolic level. I believe that the connectionist and symbolic perspectives to AI should be taken as mutually supporting approaches to the same problems, rather than as competitive areas, as is often the case. Hybrid systems integrating both perspectives appear as the right track to follow towards what A. Sloman calls the *General Science of Intelligence* [Sloman 93], [Mira & Delgado 97].

To conclude, I would like to make a final reflection on the long-term aim of research in Robotic and Artificial Intelligence. A related longer discussion—that is worth reading—can be found in [McKerrow 91, pp. 14-23]. Our work, together with the work of many researches in this field, will be a contribution towards more autonomous and useful robots. We could even say more intelligent, in the same way we say an ape is more *intelligent* than an insect. We may also state that in the future robots will become more and more similar to persons, if by similar we understand that they will be able to mimic many capabilities of human beings. However, science will not be able to go further, since there will always be an *ontological discontinuity* between human beings and the rest of living creatures and artifacts, in the sense E. F. Schumacher discusses it in his posthumous essay *A Guide for the Perplexed* (1977). This gap in the level of being is summarized by Schumacher's expression for human beings:

$$m + x + y + z,$$

z will always be missing for a robot. This z includes personality, fortitude, fidelity, love, the appreciation of beauty, hope, the ability for personal relationships, faith, dignity, creativity, justice, magnanimity, solidarity, ... What is this z ? How can it be defined? Answering these questions would take us too far now. Let us only remind that we are not complex machines, and we are not just evolved apes either.

Acknowledgments

The author is grateful to Stevan Harnad for his comments on an earlier draft of this paper and to Aaron Sloman for an insightful discussion. Special gratitude is due to Prof. José Mira for the time he has spent teaching graduate courses at Jaume-I University and for many penetrating discussions.

References

- Brady, M. (1985), "Artificial Intelligence and Robotics", *Artificial Intelligence* 26, 79-121, 1985.
- Brooks, R.A. (1990) "Elephants Don't Play Chess", *Robotics and Autonomous Systems*, 6, 3-15.
- Brooks, R.A. (1991) "Intelligence without Representation", *Artificial Intelligence* 47, 139-159.

- Cervera, E., del Pobil, A.P. (1997) "Programming and Learning in Real-World Manipulation Tasks", in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'97)*, Grenoble, France.
- Cervera, E., del Pobil, A.P. (1998) "Eliminating sensor ambiguities via recurrent neural networks in sensor-based learning", in *Proc. IEEE International Conference on Robotics and Automation*, Leuven, Belgium.
- Charniak, E., McDermott, D. (1985) *Introduction to Artificial Intelligence*, Addison-Wesley.
- del Pobil, A.P., Serna, M.A. (1995), *Spatial Representation and Motion Planning*, Lecture Notes in Computer Science Vol. 1014, Springer, Berlin.
- Freksa, C. (1991) "Qualitative Spatial Reasoning", in *Proc. IMACS Inter. Symposium on Decision Support Systems & Qualitative Reasoning*, Toulouse, France.
- Gupta, K.K., del Pobil, A.P. (eds.) (1998) *Practical Motion Planning in Robotics*, John Wiley & Sons, Chichester, UK.
- Harnad, S. (1989) "Minds, Machines and Searle", *Journal of Theoretical and Experimental Artificial Intelligence* 1: 5-25.
- Harnad, S. (1990) "The Symbol Grounding Problem", *Physica D* 42: 335-346.
- Harnad, S. (1991) "Other bodies, Other minds: A machine incarnation of an old philosophical problem". *Minds and Machines* 1: 43-54.
- Harnad, S. (1992) "The Turing Test is not a Trick: Turing Indistinguishability is a Scientific Criterion", *ACM SIGART Bulletin* 3(54) 9-10.
- Harnad, S. (1993) "Grounding Symbolic Capacity in Robotic Capacity", in L. Steels, R.A. Brooks (eds.) *The Artificial Life route to Artificial Intelligence*, Lawrence Erlbaum.
- Heikkonen, J. (1994) *Subsymbolic Representations, Self-Organizing Maps and Object Motion Learning*, Research paper No. 36, Lappeenranta University of Technology, Finland.
- Lakoff, G. (1987) *Women, fire, and dangerous things: what categories reveal about mind*, University of Chicago Press, Chicago.
- Martin, P., del Pobil, A.P. (1994) "Application of Artificial Neural Networks to the Robot Path Planning Problem", in *Applications of Artificial Intelligence in Engineering IX* edited by G. Rzevski, R.A. Adey and D.W. Russell, Computational Mechanics Publications, Boston, pp. 73-80.
- McKerrow, P.J. (1991), *Introduction to Robotics*, Addison-Wesley, Sidney.
- Mel, B.W. (1990), *Connectionist Robot Motion Planning*, Academic Press, San Diego, California.
- Mira, J., Delgado, A.E., Boticario, J.G., Díez, F.J. (1995), *Aspectos básicos de la Inteligencia Artificial*, Sanz y Torres, Madrid.
- Mira, J., Delgado, A.E. (1997) "Some Reflections on the Relationships Between Neuroscience and Computation", in *Biological and Artificial Computation: From Neuroscience to Technology*, edited by J. Mira, R. Moreno-Díaz and J. Cabestany, Lecture Notes in Computer Science Vol. 1240, Springer, Berlin.
- Musto, D., Konolige, K. (1993) "Reasoning about Perception", *AI Communications* 6: 207-212.
- Sanz, P.J., del Pobil, A.P., Iñesta, J.M., Recatalá, G. (1998), "Vision guided grasping of unknown objects for service robots", in *Proc. IEEE International Conference on Robotics and Automation*, Leuven, Belgium.
- Schumacher, E. F. (1977) *A Guide for the Perplexed*. (Spanish edition titled *Guía para los perplejos* published by Editorial Debate, Madrid, 1981).

- Searle, J. R. (1980) "Minds, Brains and Programs", *Behavioral and Brain Sciences* 3: 417-424.
- Sloman, A. (1993) "Prospects for AI as the General Science of Intelligence", in *Proc. AISB Conf.*, Birmingham.
- Steels, L., Brooks, R.A. (eds.) (1993) *The Artificial Life route to Artificial Intelligence. Building Situated Embodied Agents*. Lawrence Erlbaum.
- Turing, A.M. (1950) "Computing Machinery and Intelligence", *Mind* LIX, no.2236, 433-60.