KI – Künstliche Intelligenz manuscript No.
(will be inserted by the editor)

Shakey ever after?

Questioning tacit assumptions in robotics and artificial intelligence

Alexandra Kirsch

Received: date / Accepted: date

Abstract Shakey the robot was a milestone of autonomous robots and artificial intelligence. Its design principles have dominated research until now. Tacit philosophical and architectural assumptions have impoverished the space of research topics and methods. I point out ways to overcome this impasse with sideglances to other scientific fields.

Keywords Shakey · Architectures · Design · Objectivism

1 Are we making progress?

Shakey the robot [23] was a milestone for both robotics and artificial intelligence. Most of the principles developed for or before Shakey, like A*-Search for navigation, the STRIPS representation of actions or layered command-and-control architectures, are still used in a similar form in today's systems. In contrast to many of those, Shakey spanned the whole range of perception, state estimation, planning, and motor control.

In line with other early success stories of AI, making progress from the high level that Shakey has set, has proven to be a lot more difficult than expected. The progress in performance from Shakey to today's robots is not too obvious for ousiders of robotics, especially when considering the size, price and capabilities of today's sensors and actuators together with the research effort spent in the last 50 years. Autonomous robots are still slow, clumsy and unreliable.

I suggest that the reason for the slow progess is—apart from the problem being really tough—that the research communities both in AI and robotics have failed to question assumptions and design decisions from the time of Shakey. My critique comprises assumption of different kinds:

A. Kirsch
Independent Scientist, E-mail: ak@alexkirsch.de

- **Technical assumptions** are the most obvious legacy of Shakey. This paper discusses some of the assumptions underlying common robot architecture: spliting perception and action and the strict command-and-control pattern of hierarchical structures.
- **Procedural assumptions** comprise the metrics-oriented, linear-process mindset in science (and business) in western societies, leaving little room for creative exploration of the space of technical alternatives.
- **Philosophical assumptions** have a much longer tradition than Shakey, but they are a major cause for the technical impasse.

My goal is to make readers aware that much of the research of AI and robotics follows habits that need to be rethought now and then.

2 Cutting cognition into pieces

Shakey, as any other autonomous robot, faced the challenge of covering the whole range of cognitive capabilities: determining what is going on outside the robot, interpreting it, determining something useful to do, and then really doing it without causing harm. These capabilities have been nicely chained in a sense-plan-act loop that is rarely mentioned, but in some form or other tacidly accepted.

But the split into perception, action and something in between causes hard engineering problems. What exactly is the purpose of the components and what are the interfaces between them? Which component is supposed to control the heading of the camera? Since this involves motor control, one might assume it is action. But if the robot is performing a search task, you would want an active vision module to determine the camera position. Or think about communication. Is that perception (you hear sound or see movement) or action (you produce sound or movement), or rather cognition

(you have to interpret and generate utterances or gestures in the context of the conversation)?

One of my favourite brain teasers is how to combine action and perception for robot navigation. The typical interface is 2D or 3D points in some coordinate system. For me this is one of the clumsiest representations of space. The question of how humans and other animals represent space is still under debate in psychology and neuroscience, but I have not seen a theory of coordinate neurons so far. Most people are rather bad at 3D geometry and no language on earth has to my knowledge any constructs for coordinates. There are differences in global and local reference systems, but certainly nothing resembling the numerical representations we plague ourselves with when we program robots.

Robots do not have to work like living beings, so why shouldn't robots use coordinates, since they have a calculation machine instead of a brain? The problem is that every step in this (rather shallow) hierarchy adds to the robot's errors. A camera that we nowadays put on a robot has a much better resolution than the human eye and robots can use additional sensing such as laser range finders. So the sensors are are not the limiting factor. But the process of trying to determine the one coordinate that we want the robot to move to, involves uncertainty in the sensor interpretation as well as arbitrary choices. For example, when a robot is supposed to grasp an object from a table, there are many possible coordinates for the robot to move to and grasp. But with the typical interface of passing one coordinate from the perception module to the action module, we have to single out one of the possible coordinates. Now we switch to the action side and again the robot has to make uncertain, underspecified decisions. It can chose any path to its goal coordinate, some of the paths may be more suitable for the subsequent action than others (for example moving in a way that the robot can observe the object and determine if another robot or a person is taking it away). Also the accuracy for reaching the goal position is usually set to some arbitrary radius around the goal point. But a small deviation in one direction may make no difference, while into another direction it may cause the robot to be unable to lift its arm for grasping. And the reason is that the goal passed to the navigation module contains the wrong information [22]. What was needed was "move to a position from which you can grasp the object" rather than "move to position $\langle x, y, \varphi \rangle$".

This problem is part of the larger one of hierarchies. The cognitive part usually starts at some symbolic representation. But putting the world into symbols means you have to discard details and make arbitrary or close choices. A robot may not be quite sure if there is a living human or a statue in front of it. By setting a threshold it will pass one symbol or the other to the next module, rather than its uncertainty, which could possibly be resolved by another module. We can pass a probability distribution, but it still hides information about reasons that have led to the assumptions.

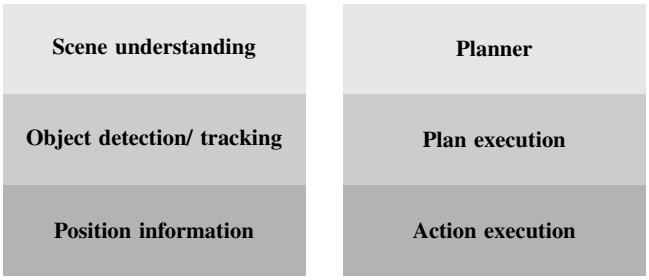


Fig. 1 Basic robot architecture that has been commonly accepted in some variation since Shakey. It assumes a split into perception and action, possibly a third cognition component, and a strict command-and-control hierarchy.

On the one hand, intelligence obviously needs some kind of abstraction, but on the other hand the cut of modules into a strict command-and-control hierarchy (Figure 1) has serious drawbacks as just discussed.

Since humans are a working system that handles this contradiction well, we may find inspiration from the natural model. Neuroscience has not found any processing pipelines in the brain, but instead interconnected modules that communicate bi-directionally or over complex network structures. Having modules communicate in more directions raises a new set of engineering challenges: what are the data structures to share between the modules? How do we avoid infinite loops? But the effort may pay off with more stable behavior. After all, human sensing is far from perfect, but in the process of interpretation and abstraction the errors get smaller.

These and other problems of the standard architecture have been noted before. The most explicit breach with these principles was announced by Brooks [1]. While his subsumption architecture never gained ground, it inspired work on behavior-based architectures [21]. Behavior-based architectures integrate systemic principles from biology and reject linear processing pipelines. Their greatest challenge is to produce higher-level behavior, such as explicit reasoning and human-level communication. Maes [20] has, however, shown a way how such architectures can include explicit arbitration between goals and actions.

Blackboard architectures [2] were inspired by cognitive models of humans [8], and were taken over to machines [7]. These architectures have a central storage—the blackboard—to allow any module to communicate with any other. Thus, blackboards enable, in principle, the communication of modules on different levels of abstraction and they have been found to be rather robust [19]. But as reported in [19], the common data format is hard to define, so that in the end, the interaction of components is more engineered than the principles of the architecture would suggest. It seems that blackboard architectures are rather popular in autonomous

robots, but there seems to be little progress on a theoretical level, even though many open questions remain. For example, the combination of several blackboards may be a way to overcome the problem of the common data structures [6, 12].

Cognitive architectures originated by modeling human behavior [18, 15]. They usually include explicit notions of memory and an active arbitration between action choices. Several were used on robots [26, 24], but the integration usually just adds another fixed level in a command-and-control hierarchy with the drawbacks described above.

Even though there is some work on architectures going on, the cut of the basic modules is never really questioned. This is possibly a reason why the integration of cognitive and robot architectures is still so clumsy. Cognitive architectures have no fitting slots for the algorithms developed for robots, while robot architectures have no room for higher-level cognitive functions. If we stick to the traditional cut of functions, we can hardly expect to come up with revolutionary architectures. The examples above show that interesting ideas are around, but they need to be given room for further exploration.

3 Science is a nonlinear process

Shakey and other early AI systems have shown that the engineering of cognitive systems is hard. Consequently, the fields of AI and robotics have split into subfields that each tackle specific modules such as perception, knowledge representation, planning or action selection. But when we look at the difficulties that the architectural choices in Shakey imply, we have to ask whether the modules that were optimized by the specialized communities, were the right ones.

This possibly erroneous split of modules is aggravated by the dynamics that each community has developed. By the isolation from other communities and from the overall goal of building complete systems, the criteria for research topics and evaluation metrics, are targeted only towards specific, isolated tasks, while the fit to a system is not even mentioned. This is comparable with what happens in big companies when departments are evaluated by short-sighted, isolated criteria [3]. For example, if the purchasing department is instructed to buy services and parts as cheap as possible, the quality of the products and customer satisfaction will suffer. But since these are performance indicators of other departments, the purchase does not care. It is good to have quality criteria in single departments or research communities, but these criteria should include quantitative or qualitative measures for the success of the joint goal.

The bottom line is that 50 years after Shakey we have lots of algorithms for a faulty default architecture, optimized along criteria that may or may not matter in a complete autonomous system.

Robotics and AI face the same problem as any scientific discipline today: in western societies scientific results are treated like industry products. Scientists are expected to publish at high speed, just as Charlie Chaplin in *Modern Times* has to fix some screw at the highest possible speed. Getting a paper published is easiest if you can present results that are somehow "better" than previous results. This implies that 1) you should not be the first to tackle a problem, otherwise you have nothing to compare to, 2) you should choose a problem with a metric that is easy to measure and compare, such as runtime or some theoretical optimum. Isolated algorithms are therefore a much more rewarding research target than complete systems or architectures, because they not only fulfill points 1 and 2, they are much faster developed and tested; if you are lucky someone has started a competition and you do not even have to worry about instances to evaluate on.

And since the reviewer is locked to a conveyer belt just like the writer of a paper, it is best to keep to standard evaluation criteria. It does make sense to let an algorithm run and see whether and how it works, but not every quality can be objectively measured. Some algorithm may be more flexible than another, allowing it to be used in a wider range of situations. You can try to evaluate this generality by applying your algorithm to a wide range of situations, but this puts you into a severe disadvantage to the competition using the standard quantitative measures. It will take you more time to do the work, thus slowing down your publishing rate, and all your nice tests will probabily not fit into the 6- or 8-page space of a robotics or AI conference.

All this pressure on science seems to assume some linear process of development. Just like new releases of cars are expected to be better than the model before, people assume that every published approach is better than the last one. The difference between cars and autonomous robotics is that the technology for cars has been well understood for the last hundred years, while intelligent robots are still a complete puzzle. Today's cars are not derived from the very first prototype of a car. Early inventors tried different approaches, some of which worked, others did not. Only then could the car industry get into this streamlined process[1].

The problem we have now is that one particular approach—the Shakey-like architecture—has been optimized as far as it can. If someone seriously tried to cut the task of cognition in some other way, not just adjusting the architecture with the components as they are standardly cut, but in a really new way that would require different types of modules, the result (measured in whatever way) would be worse than that of

[1] It is also questionable whether such optimization of known technology is beneficial for industries. Brooks (`https://rodneybrooks.com/the-end-of-moores-law/`) criticizes the circuit industry for not thinking about new concepts of computer architectures.

the prevailing standard, simply because the standard has 50 years of optimization included, while the new one has possibly a year. In addition, the approach can in fact be worse than the Shakey-like architecture. When you try something new it can turn out to be wrong, that is what "trying" implies.

To really make progress, we have to detach ourselves from the linear development pattern. Science is all about trial and error. We need to explore to get better. In this respect science can learn a lot from design [11]. David Kelley, founder of the Hasso Plattner Institute of Design at Stanford University and the design and consulting firm IDEO, has popularized the method of design thinking. He warns that "striving for perfection can get in the way during the early stages of the creative process" [10, p. 123].

Even though fifty years have passed from the time of Shakey, we are still at an early stage of figuring out how to make robots act intelligently. Perfection is consequently the wrong goal. Kelley advises to "fail early to succeed sooner" [27, p. 187]. Failing is of course not a goal in itself: "Failure sucks, but instructs" [10, p. 43]. We must 1) recognize that cognitive robotics is still in a prototype phase, 2) follow different paths, even though some (or many) of them may turn out as a dead end, and 3) as a community be honest and allow the members of the community to be honest in identifying such dead ends.

4 Reality is more than math

I have critized some of the technical assumption underlying current work in AI and robotics. The architecture and how modules are developed are only examples, there are many more such assumtions, which I leave to the reader to discover. But the technical choices are related to philosophical assumptions that go way beyond the AI and robotics communities, even beyond science as a whole. "The myth of objectivism has dominated Western culture, and in particular Western philosophy, from the Presocratics to the present day. The view that we have access to absolute and unconditional truths about the world is the cornerstone of the Western philosophical tradition." [17, p. 195]

Humans may be limited in recognizing such truths, but they are assumed to exist and the goal of science is supposed to figure them out. The ultimate tool for formulating such truths is mathematics. Algorithms are considered good if they are "mathematically clean" and find "the optimum" (or are at least close to "the optimum", hopefully with a known boundary).

From image recognition to movement planning, everybody unquestioningly publishes resultes of how far the solution is from some theoretical optimum. Have you ever worried whether your grocery shopping was optimal? What would that involve? Going the fastest way, finding the best parking space, moving in the supermarket on an optimal path, buying the products at the lowest price (or best quality? (how do you measure that?), choosing the shortest checkout line?). As humans we partially consider these aspects, but none of us would try to find a global optimum, and if just because only few of the criteria are measurable and comparable to alternatives.

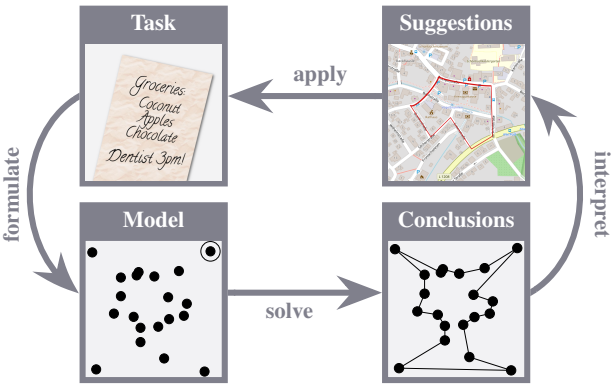


Fig. 2 General engineering approach: building a model of a real-world problem, solving the modeled problem, and reapplying it to the original task.

Objectivism is an assumption that seems to feel natural to most people. Lakoff, however, shows that it contradicts emprical findings: "Objectivist philosophy likes to view itself as having science on its side. In the case of biological categories, science is not on its side. [. . .] they do not accord with phonomena that are central to evolution—variation within species, adaptation to the environment, gradual change, gene pools, etc." [16, p. 195]

The objectivist assumptions are mirrored in the obsession for optimality in AI and computer science in general. Figure 2 shows the standard textbook diagram of engineering. To solve a real-world task, we build an abstract, simplified model of this task, then we solve the simplified task in the scope of the model, and apply the solution in the real world, hoping that our model was close enough to the original task.

Most of AI research, however, is only going on between the two lower boxes in Figure 2. Many researchers seem not even to be aware that they are working on simplified models. The only thing that counts is that the conclusions are optimal or close to the theoretical optimum. In fact, the notion of "optimum" is only defined in the lower two boxes. Reality has so many more parameters that are often unkown and unmeasurable that it is absurd to talk of optimality at all. Most AI communities, be it planning, reasoning, or machine learning, have spent decades to get optimal or close to optimal results for large problems in reasonable time. All of this assumes that the model is very close to the real-world task and that the theoretical optimum is also the best solu-

tion in reality. Empirical confirmation of this assumption is not sought.

Roboticists cannot so easily shut theirs eyes before reality. Even in simplified laboratory environments, robots have to cope with many aspects of reality. But the techniques that are used on those robots are taken over from AI. So we use techniques that have been developed with very idealized assumptions in environments where most of those assumptions do not hold. The result are the slow, clumsy robots that we are used to. We must replace our objectivist mindset when developing algorithms with more realistic assumptions.

Simon [25] coined the term "satisficing" for the human strategy of acting in a "good enough" way. This strategy could at first sight be taken as a shortcoming of human skills. But when we look at the world, for most realworld tasks, there simply is no optimum defined.

Gigerenzer and colleagues [5,4] have collected examples of real-world tasks where optimization does worse than simple rules of thumb. Such tasks are characterized by their lack of information and the multitude of interacting influencing factors—typical conditions for robotic applications.

Klein [13,14] is one of the few researchers that have moved outside the laboratory to look at how people make decisions in real life, especially under pressure, for example fire fighters. He found that such people hardly ever consider alternatives. They assemble pieces of memory to derive a new solution from known ones. Compare this to AI planning with a logical representation of the world and the goal to find an optimal solution. Any house would be in ashes before the fire fighters had even mentally formulated the problem, let alone solved it.

Katsikopoulos [9] describes how the field of Operations Research (OR) is developing decision methods for the real world and how it has overcome the urge to optimize theoretical problems. "The *Science of Better* (not of *Best*!) was and still remains a popular explication of what OR is." [9]

5 Shakey as a role model

In this article I have criticized some of the technical design decisions of Shakey. My intention is not to discredit Shakey or its developers. On the contrary, I think we should honor Shakey in that we see it as a wonderful starting point for autonomous robots from which we can learn. Without this learning, Shakey is just a suboptimal blueprint. Cognition is a really hard problem. To solve it we have to approach it from different angles instead of marching in single file.

My short overview of different architectural concepts demonstrates that researchers have creative ideas for advancing cognitive robotics. But these ideas need a scientific environment in which they can be explored with an equal level of recognition, not just a small subset that happens to match the philosophical mindset that western societies have been dragging along for more than two thousand years.

Change in a community must be triggered by individuals. Every reader has the power to start this transformation in the choice of research topics, in the mindset for reviewing proposals and papers, and in the way we educate the next generation of AI and robotics researchers.

Conflict of interest

The author declares that she has no conflict of interest.

References

1. Brooks, R.: Intelligence without representation. Artificial Intelligence **47**, 139–159 (1991)
2. Engelmore, R., Morgan, T. (eds.): Blackboard Systems. Addison-Wesley Publishing Company (1988)
3. Gharajedaghi, J.: Systems Thinking: Managing Chaos and Complexity, 3rd edn. Morgan Kaufmann (2011)
4. Gigerenzer, G., Brighton, H.: Homo heuristicus: Why biased minds make better inferences. Topics in Cognitive Science **1**, 107–143 (2009). DOI 10.1111/j.1756-8765.2008.01006.x
5. Gigerenzer, G., Todd, P.M., the ABC Research Group (eds.): Simple Heuristics That Make Us Smart. Oxford University Press (1999)
6. Haber, A., Sammut, C.: A cognitive architecture for autonomous robots. Advances in Cognitive Systems **2**, 257–275 (2013)
7. Hayes-Roth, B.: A blackboard architecture for control. Artificial Intelligence **26**(3), 251–321 (1985). DOI 10.1016/0004-3702(85)90063-3. URL `http://dx.doi.org/10.1016/0004-3702(85)90063-3`
8. Hayes-Roth, B., Hayes-Roth, F.: A cognitive model of planning. Cognitive Science **3**(4), 275–310 (1979)
9. Katsikopoulos, K.: Kirsch's, and everyone's, bind: How to build models for the wild? Cognitive Processing **20**(2), 269–272. URL `https://doi.org/10.1007/s10339-019-00914-1`
10. Kelley, D., Kelley, T.: Creative Confidence. Harpercollins Uk (2013)
11. Kirsch, A.: Lessons from human problem solving for cognitive systems research. Advances in Cognitive Systems **5**, 13–24 (2017). URL `http://www.cogsys.org/papers/ACSvol5/paper-5-4.pdf`
12. Kirsch, A.: A modular approach of decision-making in the context of robot navigation in domestic environments. In: 3rd Global Conference on Artificial Intelligence (GCAI) (2017)
13. Klein, G.: The fiction of optimization. In: G. Gigerenzer, R. Selten (eds.) Bounded rationality: The adaptive toolbox, pp. 103–122. MIT Press, Cambridge, MA (2002)
14. Klein, G.: Sources of Power: How People Make Decisions. The MIT Press. MIT Press (2017). URL `https://books.google.de/books?id=JW01DwAAQBAJ`
15. Kotseruba, I., Tsotsos, J.K.: A review of 40 years of cognitive architecture research: Core cognitive abilities and practical applications. arXiv:1610.08602 (2016)
16. Lakoff, G.: Women, Fire, and Dangerous Things: What Categories Reveal about the Mind. The University of Chicago Press (1987)
17. Lakoff, G., Johnson, M.: METAPHORS We Live By. The University of Chicago Press (1980)

18. Langley, P.: Progress and challenges in research on cognitive architectures. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17) (2017)
19. Lesser, V.R., Erman, L.D.: A retrospective view of the hearsay-ii architecture. In: IJCAI'77: Proceedings of the 5th international joint conference on Artificial intelligence (1977)
20. Maes, P.: Situated agents can have goals. Robotics and Autonomous Systems **6** (1990)
21. Matarić, M.: Behavior-based robotics as a tool for synthesis of artificial behavior and analysis of natural behavior. Trends in Cognitive Science **2**(3), 82–87 (1998)
22. Müller, A., Wünsche, H.J.: Object-related-navigation for mobile robots. In: Proceedings of IEEE Intelligent Vehicles Symposium (2012)
23. Nilsson, N.J.: Shakey the robot. Tech. Rep. 323, AI Center, SRI International (1984). URL `http://www.ai.sri.com/shakey/`
24. Scheutz, M., Harris, J., Schermerhorn, P.: Systematic integration of cognitive and robotic architectures. Advances in Cognitive Systems **2**, 277–296 (2013)
25. Simon, H.A.: Theories of bounded rationality. In: C. McGuire, R. Radner (eds.) Decision and Organization. North-Holland Publishing Company (1972)
26. Veloso, M., Carbonell, J., Pérez, A., Borrajo, D., Fink, E., Blythe, J.: Integrating planning and learning: The prodigy architecture. Journal of Experimental and Theoretical Artificial Intelligence **7**(1) (1995)
27. und Yves Pigneur, A.O., Bernada, G., Smith, A.: Value Proposition Design. Wiley (2014)