How Artificial is Intelligence in AI?
Arguments for a Non-Discriminatory Turing test

Jack Birner

University of Trento, University College Maastricht
jack.birner@unitn.it

Abstract. Friedrich von Hayek's *The Sensory Order* (1952) presents a physicalist identity theory of the human mind. In a reaction to Karl Popper's criticism that such a "causal" theory of the mind cannot explain the descriptive and critical-argumentative functions of language, Hayek wrote a paper that was never published. It contains the description of a thought experiment of two communicating automata. This paper confirms the impression of the AI-like character of the structuralism and functionalism of Hayek's *Sensory Order*. In some important respects, what Hayek tries to do in his paper is similar to Turing's discussion of the question "can machines think?" Arguments will be given why according to a functionalist and physicalist identity theory of mind the distinction between artificial and "natural" intelligence cannot be upheld. According to such a theory, Turing tests are unnecessarily restrictive and discriminatory vis-à-vis machines. In the end, the question whether or not machines can think is not meaningless, as Turing thought. It can be replaced by the question if artificial minds are capable of consciousness. The Turing test , however, cannot give the answer.

Key words. theory of mind • physicalist identity theory • virtual machines • communication of symbolic description

1 Introduction

This paper is the consequence of the interest in the philosophy of Karl Popper that I share with Aaron Sloman. A couple of months ago he reacted to the announcement of a conference on Popper that bears my signature and this led to both of us reading some of the other's publications. We discovered that we had more interests in common. This happy chance meeting of minds led to my writing what you are now reading.[1] Popper is also one of the *dramatis personae* of this story, next to Friedrich von Hayek. Popper and Hayek became close intellectual and personal friends during and after the Second World War. In their published work they appear to agree on almost everything. Some aspects of their thought, however, convinced me that this could not be really true. And indeed, a closer look revealed that till the end of their lives they remained divided on several important issues. I have dealt with some of these, and with the influence – both positive and negative - they had on one another elsewhere.[2]

[1] Without Aaron's encouragement I would not have dreamt of sending a text to a workshop on AI. Let me hasten to add that his guilt stops here: I take full responsibility for everything that follows. I would also like to apologize in advance for not referring to authors who may have discussed the same or similar problems; these are my first steps in AI.

[2] Cp. Birner (2009) and (forthcoming).

2 Philosophy of mind

What I will take up here are their disagreements in the philosophy of mind. I do so first of all because Hayek's theory of mind and his defence against Popper's criticism have a strong AI flavour.[3] Second, there are some striking similarities between Hayek's work of the early 1950s and "Computing Machinery and Intelligence" (*CMI*) of 1950 by Alan Turing, the third main character of this tale. These parallels deserve more attention than has been given them.[4] In 1952 Hayek published *The sensory order: an inquiry into the foundations of theoretical psychology* (*SO*). The foundations mentioned in the title are a philosophy of mind that I will now summarize. Hayek tries to explain the human mind using only the laws of physics. He had adopted this explanatory programme from Moritz Schlick's *Allgemeine Erkenntnistheorie.* The ontological idea underlying it is that the mind does not have a separate existence from the brain. So Hayek's is a physicalist identity theory.

As the vehicle for his explanation he uses a neural-network model.[5] According to Hayek, mental processes consist in the continuous reorganization on many levels of a hierarchical system of relationships. That is why he speaks of an order of events. A neural network is one possible model of the mind. Hayek is a radical functionalist in the sense that he states that *any* physical configuration of elements and their relationships might embody mental processes. He introduces this idea thus:

> "That an order of events is something different from the properties of the individual events, and that the same order of events can be formed from elements of a very different individual character, can be illustrated from a great number of different fields. The same pattern of movements may be performed by a swarm of fireflies, a flock of birds, a number of toy balloons or perhaps a flight of aeroplanes; the same machine, a bicycle or a cotton gin, a lathe, a telephone exchange or an adding machine, can be constructed from a large variety of materials and yet remains the same kind of machine within which elements of different individual properties will perform the same functions. So long as the elements, whatever other properties they may possess, are capable of acting upon each other in the manner determining the structure of the machine, their other properties are irrelevant for our understanding of the machine." (*SO* 2.28)

Then he proposes a radically functionalist and structuralist hypothesis:

> "In the same sense the peculiar properties of the elementary neural events which are the terms of the mental order have nothing to do with that order

[3] Already hinted at in an afterthought to Birner 2009, where I wrote that Hayek missed the chance to be recognized as a pioneer in AI. This will be discussed below.
[4] But cp.Van den Hauwe (2011).
[5] As does Donald Hebb, the publication of whose *The Organization of Behavior* in 1949 almost kept Hayek from publishing his book. *SO* elaborates a manuscript that dates from 1920. For a discussion, cp. Birner (2014).

> itself. What we have called physical properties of those events are those properties which will appear if they are placed in a variety of experimental relations to different other kinds of events. The mental properties are those which they possess only as a part of the particular structure and which may be largely independent of the former. *It is at least conceivable that the particular kind of order which we call mind might be built up from any one of several kind of different elements – electrical, chemical, or what not; all that is required is that by the simple relationship of being able to evoke each other in a certain order they correspond to the structure we call mind.*" (*SO* 2.29, my italics)[6]

This sounds very AI-like. The link between Hayek's theory of mind and AI is even more apparent in the way Hayek developed his ideas after the publication of *SO*. That is the subject of the next section.

3 Popper's criticism

Upon publication of *SO* Hayek sent a copy to Popper. Although Popper was – as always - very polite in his reaction, he did not like it. Though Popper never writes this down, his main general objection to Hayek's theory of mind is that it is too inductivist. What he does write in a letter to Hayek (2 December 1952) is that he thinks his theory of the sensory order is deterministic. This implies, says Popper, that it is a sketch for a deterministic theory of the mind. Now Popper had just written a criticism (later published as Popper 1953) of this type of theory.[7] He argues that a deterministic theory of the mind cannot be true because it is impossible to have a deterministic theory of human language.

In his criticism, Popper uses a particular analysis of language. He considers it to be part of his solution to what he calls Compton's problem. Popper uses that name for what he considers to be a generalization of Descartes' formulation of the mind-body problem. Descartes asks how the immaterial mind can act upon the physical body. Popper wants to know how abstract entities such as the contents of ideas and theories can influence the physical world. He builds upon Karl Bühler's theory of the evolution of language. It says that the first function of language to emerge in human evolution is the *expression* of subjective states of consciousness. The next function to develop is *communication* (or *signaling*), followed by *description*. Popper adds a fourth function, *argumentation and criticism*. It presupposes the previous (or, as

[6] For a contemporary elaboration of this idea that seems to be very fruitful for understanding and measuring consciousness, cf. Tononi 2012.

[7] Apparently as a criticism of *SO*, of which he may have read the proofs. Cp. what Popper writes to Hayek (letter of 30 November 1953 – Klagenfurt Popper archives, folder 541.12, on file from microfilm of the Hoover Archives): "I was extremely pleased to hear that with "the challenge of my article on Language and the Body Mind Problem", I have done "a great service". I am really happy about this article. I have ??? ???? M... (?) on the problem, but although I think that I got somewhere, I don't know whether it is worth much. If you really can refute my views (?), it would, I think, be an achievement." (hand writing partially illegible).

Popper says, lower) functions. Not only has the need of humans to adapt to the environment given rise to new physical instruments, it has also produced their capacity to theorize. That is a consequence of the evolution of the higher functions of language: they serve to control the lower ones (Popper 1972: 240-41). Abstract contents of thought, meanings and the higher functions of language[8] have co-evolved. They help us control our environment "plastically" because they are adaptable. Popper proposes a dualistic and indeterministic theory of the mind and of the influence of the contents of consciousness on the world, which according to him can account for the higher linguistic functions – unlike physicalist and behaviourist theories:

> "When the radical physicalist and the radical behaviourist turn to the analysis of human language, they cannot get beyond the first two functions (see my [1953]). The physicalist will try to give a physical explanation - a causal explanation - of language phenomena. This is equivalent to interpreting language as expressive of the state of the speaker, and therefore as having the expressive function alone. The behaviourist, on the other hand, will concern himself also with the social aspect of language - but this will be taken, essentially, as the way in which speakers respond to one another's "verbal behavior." This amounts to seeing language as expression and communication.
>
> But the consequences of this are disastrous. For if language is seen as merely expression and communication, then one neglects all that is characteristic of human language in contradistinction to animal language: its ability to make true and false statements, and to produce valid and invalid arguments. This, in its turn, has the consequence that the physicalist is prevented from accounting from the difference between propaganda, verbal intimidation and rational arguments." (Popper and Eccles 1977: 58)[9]

Hayek took this criticism of Popper's very seriously.[10] He responded to it in "Within Systems and about Systems; A Statement of Some Problems of a Theory of Communication." That paper was never published. It was never finished, either. Later Hayek writes about it:

> [I]n the first few years after I had finished the text of the book [*SO*], I made an effort to complete its formulations of the theory in one respect. I had then endeavoured to elaborate the crucial concept of "systems within systems" but found it so excruciatingly difficult that in the end, I abandoned the

[8] All of these are inhabitants of what Popper in his later philosophy has called world-3.

[9] Popper & Eccles (1977) makes the same points that are made in the 1953 paper more forcefully.

[10] "With the challenge of your article on "Language and the Body Mind Problem" you have unwittingly done me a great service. Much has crystallized in my mind as a result of my inability fully to accept (?) the argument. I believe I can now (?) provide (?) a causal theory of description and intention, but of course only an "explanation of the principle" applicable to greatly simplified models and not sufficient to provide a full explanation either of human language or human intention. But sufficient to construct models possessing all the characteristics common to all instances of "description" and intention. I am still struggling with great (?) difficulties, but I believe I am getting somewhere." (Hayek to Popper, 30 October 1953, Popper Library, Klagenfurt, folder 541.12, on file, from microfilm Hoover archives, hand writing partially illegible).

> longish but unfinished paper that apparently nobody I tried it upon could understand". (Hayek 1982: 290)

In the paper Hayek follows a two-pronged defence strategy against Popper's criticism, one "negative," the other constructive or "positive". As to the former, Hayek states the purpose of the paper as

> "deriving from the study of certain kinds of causal systems conclusions concerning the character of our possible knowledge of mental processes. (…) [T]he main conclusion to which [the argument] will lead is that for any causal system there is a limit to the complexity of other systems for which the former can provide an analogon of a description or explanation, and that this limit necessarily excludes the possibility of a system ever describing or explaining itself. This means that, if the human mind were a causal system, we would necessarily experience in discussing it precisely those obstacles and difficulties which we do encounter and which are often regarded as proof that the human mind is not a causal system." (*Systems*: 1).

Put bluntly, this "negative" part of Hayek's reaction to Popper's criticism is of the heads-I-win-tails-you-lose type. The gut reaction of Popperian philosophers to such an argument would be to condemn it out of hand as an immunizing stratagem. Interestingly enough, Popper does not do so. I will briefly come back to this below. The average non-Popperian citizen of Academe might instead dismiss it as corny. That, however, would fail to do justice to Hayek. He gives two arguments for his conclusion. First, as he states in the next sentence, "[w]e shall find that to such a system the world must necessarily appear not as one but as two distinct realms which cannot be fully "reduced" to each other." (*ibid.)* The second argument invokes complexity. In a generalized form it says that an *explanans*, in order to be successful, has to be more complex than its *explanandum*. The argument is taken over from *SO*: "any apparatus of classification must possess a higher degree of complexity than is possessed by the objects which it classifies… therefore, … the human brain can never fully explain its own operations." (*SO*: 8.68).[11] This may be true or false but it certainly deserves closer examination. If it is true, then Hayek has demonstrated by a *reductio ad absurdum* that the mind cannot explain[12] itself (for it would have to be more complex than it is).

The complexity Hayek refers to, and which he does not explain in more detail, may consist of at least two circumstances. One has to do with problems of self-reference, the other with the impossibility of describing all the relevant initial conditions for explaining the human mind. Hayek does not mention or elaborate these aspects (which would deserve closer scrutiny). What he does instead is to work out, in subsequent publications, the methodological idea of in-principle explanations or explanations of the principle, which are all we can achieve in the case of complex

[11] For Hayek, who is a methodological instrumentalist, explanation is tantamount to classification. Cp. Birner (forthcoming).
[12] In the sense of classify, which is of course a view of explanation that is not shared by everyone (not by Popper, for instance).

phenomena.[13] Instead of rejecting this idea, that underlies Hayek's "explanatory impossibility theorem," as part of a move to make Hayek's naturalistic theory of mind immune to criticism, Popper takes it seriously enough to refer to it 25 years later.[14]

In the modern literature on the mind-body problem Hayek's argument is known as the explanatory gap (cf. Levine 1983 and 1999 and Chalmers 1999). In *SO* Hayek claims that his theory is less materialistic than dualistic theories because it does not assume the existence of a separate mind-substance: ''While our theory leads us to deny any ultimate dualism of the forces governing the realms of the mind and that of the physical world respectively, it forces us at the same time to recognize that for practical purposes we shall always have to adopt a dualistic view'' (*SO*, 8.46). This is because we cannot produce a complete description or explanation of the processes that constitute our mind and its relationships with the physical order without including a description of the subset of those same processes that do the describing and explaining, i.e., the mind itself. This again is because, as Hayek repeats in 8.44, his theory is not a double-aspect theory. The complete order of all neural processes, ''if we knew it in full, would ... not be another aspect of what we know as mind but would be mind itself.''

Since *SO* is an identity theory, rather than denying the possibility of reducing the sensory order to the physical order, it implies that there is no need to do so. In the physical order, events are similar or different to the extent that they produce similar or different external effects. In the sensory order, events are classified according to their sensory properties: ''to us mind must remain forever a realm of its own which we can know only through directly experiencing it, but which we shall never be able fully to explain or 'reduce' to something else'' (*SO* 8.98). Yet, the two ways of describing mental phenomena, in physical and in subjective terms, are two alternative ways of describing the same phenomena. For the practical purpose of describing the mind Hayek is a dualist in the sense that we humans with our human minds use different languages describing the mental and the physical. Ontologically, there is just one physical order.[15]

4 Hayek as a Pioneer of AI

[13] Cp. for instance Hayek 1967.

[14] "It has been suggested by F.A. von Hayek ([1952], p. 185) that it must be impossible for us ever to explain the functioning of the human brain in any detail since "any apparatus … must possess a structure of a higher degree of complexity that is possessed by the objects" which it is trying to explain." (Popper and Eccles 1977: 30).

[15] Cp. Levine 1999: 11: "Metaphysically speaking, there is nothing to explain. That is, we are dealing with a brute fact and there is no further source (beyond the fact itself) responsible for its obtaining. The fact that we still find a request for an explanation intelligible in this case shows that we still conceive of the relata in the identity claim as distinct properties, or, perhaps, the one thing as manifesting distinct properties. We can't seem to see the mental property as the same thing as its physical correlate. But though our inability to see this is indeed puzzling, it doesn't show, it can't show, that in fact they aren't the same thing. For what is the case cannot be guaranteed by how we conceive of it."

The constructive defence against Popper's criticism is undertaken in the second part of the paper. Hayek describes a thought experiment that is meant to demonstrate that a causal system *is* capable of one of the higher functions of language, description. By "system" he intends

> "a coherent structure of causally connected physical parts. The term system will thus be used here roughly in the sense in which it is used in von Bertalanffyi's "General System Theory (…) [By system I intend] a persistent structure of coherent material parts that are so connected that, although they can alter their relations to each other and the system thereby can assume various states, there will be a finite number of such states of which the system is capable, that these states can be transformed into each other through certain orderly sequences, and that the relations of the parts are interdependent in the sense that if a certain number of them are fixed, the rest is also determined." (*Systems*, pp. 4-5)

Hayek concentrates on the behaviour of a type of causal system that he calls "classifying system," for a fuller explanation of which he refers to *SO*.[16] After dealing, in the first part of the paper, with a series of preliminaries, Hayek is ready with

> "the setting up of the framework within which we wish to consider the main problem to which this paper is devoted. In the next section we shall take up the question how such a system can transmit to another similar system information about the environment so that the second system will as a result behave in some respects as if it had directly undergone those effects of the environment which in fact have affected only the first system, but have become the object of the "description" transmitted by that system to the second." (*Systems*: 18-9)

He introduces two automata[17] that communicate with one another by means of symbols. Since he uses them in a thought experiment, it is justified to consider them as virtual machines.[18] Hayek very ably concentrates on his main problem by excluding the different problem whether, or to what extent, the structure of the two systems have to be identical or similar in order to be able to interact with one another:[19] he assumes that they are identical. Hayek argues that the self-expressive or symptom and signaling functions of communication pose no problem for his thought experiment. Then he describes a situation in which the two systems are hunting a prey. S_1 can see the prey but S_2 cannot because it is hidden from it by an obstacle. The problem now is how S_1 can describe and communicate the description of the itinerary the prey is following to S_2. The manuscript breaks off in the middle of this attempt to fit the descriptive function of communication by means of symbols into the thought

[16] Hayek's description in *SO* of the human mind is that of a classifier system (a term he does not use).
[17] Hayek does not use that term but he refers to Von Neumann's theory of automata.
[18] Aaron Sloman's comment in correspondence.
[19] He addresses that problem elsewhere. For a discussion, cp. Birner (2009).

experiment, and in the framework of a causal theory of systems.[20] Apparently he did not succeed getting beyond the lowest two functions of communication.[21] This is precisely what Popper had said in his criticism.

5 Hayek and Turing

This section is dedicated to a (non exhaustive) comparison of the ideas in Hayek's *SO* and *Systems* with Turing's in *CMI*. The objective is to give additional arguments that Hayek's *SO* and even more so his *Systems* deserve a place in the AI literature: if Turing's *CMI* is about AI, then so are these texts of Hayek's.

5.1 What is the question?

In a comparison between Turing and Hayek we must not lose from sight that they address different problems – at least at first sight. In *CMI* Turing poses the question "Can machines think?" The problem Hayek wants to solve in *SO* is "What is consciousness?" This, at any rate, is my reconstruction; Hayek himself is much less sure and explicit in *SO*,[22] even though he writes: "it is the existence of a phenomenal world which is different from the physical world which constitutes the main problem" (*SO*, 1.84). This is part of the qualia problem. It is different from the question whether or not we humans can think; it is at best part of the latter problem. Nevertheless, the way Turing and Hayek elaborate their respective problems show some similarities that in my opinion make a comparison non futile.

Turing transforms his original question

> "into [a] more accurate form of [it:] I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9, to make them play the imitation game so well that an average interrogator will not have more than 70 per cent, chance of making the right identification after five minutes of questioning. The original question "Can machines think?" I believe to be too meaningless to deserve discussion." (*CMI*: 442).

[20] It breaks off in the middle of a word, "system". That suggests that part of the typescript has gone missing. I have repeatedly looked for the missing pages in the Hayek archives. A hand-written note by Hayek on the first of the 27 typewritten pages of the ms. reads: "seems incomplete." Added to Hayek's comment quoted in the third para. of section 3 above, this laconic note suggests that he has not looked very hard for possible missing pages, which may be very few in number.

[21] This is also suggested by the fact that years later Hayek writes to Popper that he feels "that some day you ought to come to like even my psychology" (letter of 30 May 1960, Hayek Archives, Hoover Institution on War, Revolution and Peace, box 44/2). This may be taken to imply that Hayek had not solved the problem of showing that causal systems are capable of communication descriptions to other causal systems, thus confirming Hayek's comments (Hayek 1982: 290) quoted above.

[22] This is highly uncharacteristic for Hayek, who in all his work follows a meticulously methodical approach. Cp. Birner (2013).

Now this reformulation comes much closer to the way in which Hayek elaborates the problem of *SO* in the second part of *Systems*. His thought experiment, which is meant to show that physical machines can express their internal states, signal, and communicate descriptions to one another, qualifies as an early exercise in AI. That exercise, moreover, is inspired by a physicalist identity theory of the human mind. Turing's "imitation game" is always interpreted as a procedure in which a human mind attempts to debunk a computer that tries to imitate another human mind. A generalized version of the game, one that is not based on the ontological assumption that a human mind and a computer (and/or its software – in the sequel I will delete this addition) are fundamentally different, would lose its purpose and become meaningless. If there are no fundamental differences between computers and human minds – as Hayek's physicalist identity theory asserts – a Turing test would only compare one kind of material realization of a mind with another. I will return to this in the Conclusion.

When Turing discusses the possible objection of the "Argument from Consciousness," *i.e.*, that machines can only be considered to be capable to think if they are capable of experiencing feelings and emotions, he deals with the same problem as Hayek in *SO*. Turing does not deny there is a problem, but he considers it as different from, and secondary to, the problem that he addresses:

> "I do not wish to give the impression that I think there is no mystery about consciousness. There is, for instance, something of a paradox connected with any attempt to localise it. But I do not think these mysteries necessarily need to be solved before we can answer the question with which we are concerned in this paper." (*CMI*: 447).

Now, according to Hume "Reason is, and ought only to be the slave of the passions." (Hume 1739: 415).[23] The very least we need for rational thought are motivations.[24] Hayek deals with this effectively by describing how intentions may be modeled in his thought experiment:

> "By intention we shall mean such a state of a system that, whenever its classifying apparatus represents a chain of actions as producing a result which at the same time the internal state of the system singles out as appropriate to that state, it will perform that chain of actions. And we shall define the result or class of results which in any such state will activate the chains of actions which will produce them as the goal or goals to which the intention is directed." (*Systems*: 17)

This is sufficient for the purpose of his thought experiment.

5.2 Functionalism

[23] Research in cognitive science shows that Hume was right.
[24] Aaron Sloman in correspondence.

In the above, I have described Hayek's functionalist approach to the mind. Compare this with what Turing writes:

> "The fact that Babbage's Analytical Engine was to be entirely mechanical will help us to rid ourselves of a superstition. Importance is often attached to the fact that modem digital computers are electrical, and that the nervous system also is electrical. Since Babbage's machine was not electrical, and since all digital computers are in a sense equivalent, we see that this use of electricity cannot be of theoretical importance. Of course electricity usually comes in where fast signalling is concerned, so that it is not surprising that we find it in both these connections. In the nervous system chemical phenomena are at least as important as electrical. In certain computers the storage system is mainly acoustic. The feature of using electricity is thus seen to be only a very superficial similarity. If we wish to find such similarities we should look rather for mathematical analogies of function." (*CMI*: 439)

This is identical to Hayek's mental functionalism and structuralism.

5.3 Machines as subjects of themselves

When, on p. 449, Turing writes about machines being their own subjects, he seems to have in mind a different problem than Hayek does when he addresses the question if causal systems can describe themselves – by which he means *fully* describe.

> "The claim that a machine cannot be the subject of its own thought can of course only be answered if it can be shown that the machine has *some* thought with *some* subject matter. Nevertheless, "the subject matter of a machine's operations" does seem to mean something, at least to the people who deal with it. If, for instance, the machine was trying to find a solution of the equation x^2-40a-11=0 one would be tempted to describe this equation as part of the machine's subject matter at that moment. In this sort of sense a machine undoubtedly can be its own subject matter. It may be used to help in making up its own programmes, or to predict the effect of alterations in its own structure. By observing the results of its own behaviour it can modify its own programmes so as to achieve some purpose more effectively. These are possibilities of the near future, rather than Utopian dreams." (*CMI*: 449).

This impression, however, may be mistaken. Compare the following passage:

> "The idea of a learning machine may appear paradoxical to some readers. How can the rules of operation of the machine change? They should describe completely how the machine will react whatever its history might be, whatever changes it might undergo. The rules are thus quite time-invariant. This is quite true. The explanation of the paradox is that the rules which get

> changed in the learning process are of a rather less pretentious kind, claiming only an ephemeral validity. The reader may draw a parallel with the Constitution of the United States." (*CMI*: 458)

This seems similar to the distinction Hayek makes, in para. 18, between changes within a causal system and changes of the system itself:

> "The concept of the state of a certain system must be carefully distinguished from the changes in a collection of elements which turn it into a different system. Different individual systems may be instances of the same kind of system (or possess the same structure) if they are capable of assuming the same states; and any one individual system remains in the same system only so long as it remains capable of assuming any one of the same set of states, but would become a different system in our sense. A full description of any system would have to include sufficient information to derive from it descriptions of all possible states of that system and of their relations to each other, such as the order in which it can pass through the various states and the conditions in which it will pass from one state into another. It will be noted that strictly speaking a change in the permanent nature of one of our systems such as would be produced by long term memory (the acquisition of new connections or linkages) being an irreversible change implies a change of the system rather than a mere change of the state of a given system." (*Systems*: 9-10)

The formulations are different, but Turing's and Hayek's ideas appear to be the same.

5.4 Hayek's fate

Some of Hayek's and Turing's central ideas are very similar or even identical. Yet Hayek has not been recognized as a pioneer of AI whereas Turing has. That might have been different if he had published *Systems*. The radically thorough systematic method that characterizes Hayek's approach to each and every problem he ever put on his research agenda[25] kept him from doing so; he had, after all, failed to complete what he considered to be the homework that Popper had assigned him with his criticism of *SO*. Had he published the paper, even without a satisfactory account of the communication of symbolic description between virtual machines, both Hayek and AI might have been spared a lost opportunity.

6 Conclusion: for a scientifically and morally sounder Turing test?

[25] For an explication of this methodical approach cp. Birner 2013. This is not the only case of Hayek's being the victim of his own ambitiousness and thoroughness. Cp. Birner 1994.

Perhaps the main defect of the Turing test as it is generally interpreted, is that it tests whether humans have the subjective impression that machine intelligence is human. As such, it may be of interest to psychology but hardly to AI. In addition, the Turing test is biased or at least not general (and hence unduly discriminatory in the scientific sense) in that it presupposes a particular type of theory of mind without making this explicit, one that excludes the physicalist identity position. In *CMI*, C, the interrogator, is a human being. In a scientifically sounder version of the Turing test the population of humans and machines should be randomly divided in testers and tested or judges and judged. But this would give rise to legitimate doubts as to what the test is really testing. Is it the capacity of mind-like entities to recognize similar mind-like entities?

There is no doubt that naturally evolved human minds and bodies are capable of much more complex tasks than artificially created mind-like systems and their physical implementations. This is not due to engineering problems in the realization of the latter but to the fact that human minds and bodies are the products of a very long evolutionary process. But we already know this without a Turing test.

Whether or not human judges in a Turing test can be fooled into thinking that machine intelligence is human also depends on whether or not these judges think that they share the same type of consciousness with the objects they judge. According to a radical physicalist identity theory of mind machines are capable of having consciousness and subjective feelings. If they don't,[26] this may be due to the fact that we humans happen to have a longer evolutionary history, in which we have learnt to have these impressions. Likewise, by interacting with humans, machines might learn to understand and explain why we have subjective feelings (as in Star Trek). They could even learn to have these impressions and sentiments themselves, particularly if these have survival value (which in an environment that includes interaction with human minds seems likely). The Turing test, however, is ill-suited for finding out whether or not artificially created mind-like machines have consciousness, or have consciousness that is similar to human minds. Giulio Tononi's Integrated Information Theory offers a much more sophisticated approach, one that even allows of measuring the degree of consciousness – at least in principle. In this perspective it also seems legitimate to ask if machines experience the same dualism as we humans do according to Hayek (*i.e.* we cannot speak of the realm of the mental without using subjective-psychological language;[27] see above, the last two paragraphs of section 3).

The possibility that machines have consciousness may even raise an additional, moral, objection to the traditional Turing test: it discriminates machines in favour of humans

[26] But how could we find out? This raises the same problems Hayek addressed in *Systems* without finding a solution.

[27] The non-reducibility of a subjectivist language to a physicalist one that Hayek argues for may be seen as a solution to what he considers to be a problem of complexity, *viz.* his explanatory impossibility theorem (as I have called it). That is because subjectivist language enables us to speak meaningfully about mental phenomena even in the absence of a complete reduction of them to an explanation in physical terms. Perhaps the idea can be generalized to the question if subjective language and/or impressions may serve to reduce complexity in general.

by assigning the role of judges only to the latter. Machines might feel discriminated against – if, I repeat, they are capable of moral feelings and other emotions at all.

So in the end, my arguments for a scientifically and morally sounder Turing test seems to lead to the conclusion that the Turing test does not serve any useful purpose at all. Turing's belief, quoted above, that "[t]he original question "Can machines think?" [is] too meaningless to deserve discussion" seems to me to be unfounded. Thinking involves things such as intentionality, description, explanation, understanding, creativity, having impressions and creativity. These are all features of consciousness. So Turing's question would reduce to the problem whether intelligent machines are capable of consciousness. That certainly is a difficult question, but it is hardly meaningless. As with so much research in AI, attempts to answer it have taught us more about human minds than about artificial ones, and is likely to continue to do so.

7 References

Birner J (forthcoming), "Generative mechanisms and decreasing abstraction", in Manzo (forthcoming)

Birner J (2014), F. A. Hayek's *The Sensory Order*: An Evolutionary Perspective? *Biological Theory* on-line first, DOI 10.1007/s13752-014-0189-4

Birner J (2013), "F.A. Hayek: the radical economist," http://econ.as.nyu.edu/docs/IO/28047/Birner.pdf

Birner J (2009), "From group selection to ecological niches. Popper's rethinking of evolution in the light of Hayek's theory of culture", in Parusnikova & Cohen 2009

Birner J (1994), "Introduction; Hayek's Grand Research Programme", in Birner & Van Zijp 1994

Birner J & van Zijp, R eds (1994), *Hayek, Co-ordination and Evolution; His Legacy in Philosophy, Politics, Economics, and the History of Ideas*, Routledge, London

Bunge, M ed. (1964), *The Critical Approach to Science and Philosophy. Essays in Honor of Karl R. Popper*, The Free Press, Glencoe

Chalmers, DJ (1999), "The Explanatory Gap. Introduction", in Hameroff et.al. (1999): 1-2

Hameroff SR, Kaszniak AW, Chalmers DJ eds (1999) *Towards a science of consciousness. The third Tucson discussions and debates.* MIT Press, Cambridge

Hayek FA (1952) *The Sensory Order. An Inquiry into the Foundations of Theoretical Psychology*. University of Chicago Press, Chicago, referred to as *SO*

Hayek FA (n.d.) "Within systems and about systems. A statement of some problems of a theory of communication." Hayek Archives folder 94/51, The Hoover Institution, referred to as *Systems*

Hayek FA (1964), "The Theory of Complex Phenomena", in Bunge (1964)

Hayek, FA 1967, *Studies in Philosophy, Politics and Economics*, University of Chicago Press, Chicago

Hayek FA (1982) "The Sensory Order After 25 Years", in Weimer and Palermo (1982)

Hume, D (1739), *A Treatise of Human Nature*, Selby-Bigge, LA ed., Claredon Press, Oxford, 1896
Levine, J 1999, "Conceivability, Identity, and the Explanatory Gap", in Chalmers et.al 1999: 3-13
Levine, J (1983), "Materialism and Qualia: The Explanatory Gap" *Pacific Philosophical Quarterly* 64:354-61
Manzo G ed. (forthcoming), *Paradoxes, Mechanisms, Consequences: Essays in Honor of Mohamed Cherkaoui.* Oxford: Bardwell Press
Parusnikova & Cohen RS eds. (2009), *Rethinking Popper*, Boston Studies in the Philosophy of Science 272, Berlin, Springer
Popper KR (1953) "Language and the Body-Mind Problem", in *Proceedings of the XIth Congress of Philosophy,* **7**: 101-7, Amsterdam, North Holland
Popper KR (1972) *Objective Knowledge. An Evolutionary Approach,* Oxford, Clarendon Press
Tononi, G 2012, "Integrated information theory of consciousness: an updated account" *Archives Italiennes de Biologie,* 150: 290-326
Turing, A (1950), "Computing Machinery and Intelligence," *Mind* LIX (236): 433-560, referred to as *CMI*
Van den Hauwe LMP (2011), "Hayek, Gödel, and the case for methodological dualism," *Journal of Economic Methodology* 18 (4): 387–407
Weimer WB and Palermo DS (1982) *Cognition and the Symbolic Process*, Hillsdale, N.J, Lawrence Erlbaum