

Premier semestre 2011/12

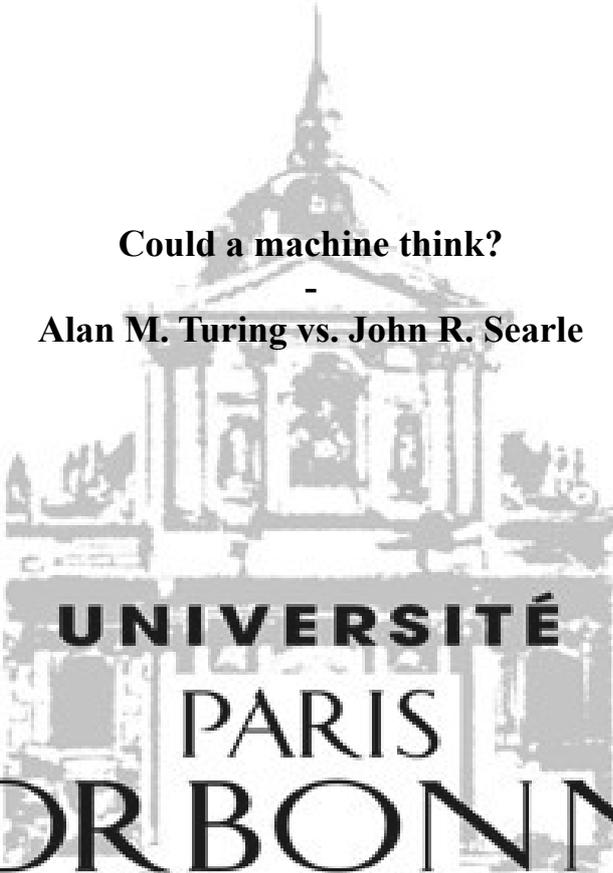
Séminaire de deuxième cycle:  
**Philosophie des sciences cognitives**

Directeur du séminaire: Dr. Pascal Ludwig

**Could a machine think?**

-

**Alan M. Turing vs. John R. Searle**



**UNIVERSITÉ  
PARIS  
SORBONNE**

---

**P A R I S I V**

Mario Günther  
51 rue Saint Fargeau  
75020 Paris  
0140303767  
[Mario.Guenther1@gmx.de](mailto:Mario.Guenther1@gmx.de)

12005618  
Philosophie (étudiant ERASMUS)

Paris, en Janvier 2012

## **Directory**

1 Preface.....	3
2 The Symbol Manipulation Approach to Intelligence.....	3
3 The Turing Test.....	6
4 The Chinese Room Argument.....	8
5 The Systems Reply.....	10
6 Conclusion.....	13
References.....	15

## 1 Preface

“Could a machine think?” asks John R. Searle in his paper *Minds, Brains, and Programs*. He answers that “only a machine could think<sup>1</sup>, and only very special kinds of machines, namely brains.”<sup>2</sup> The subject of this paper is the analysis of the aforementioned question through presentation of the symbol manipulation approach to intelligence and Searle's well-known criticism to this approach, namely the Chinese room argument. The examination of these issues leads to the systems reply of the Chinese room argument and tries to illustrate that Searle's response to the systems reply does not detract from the symbol manipulation approach.

The symbol manipulation approach to intelligence includes the physical symbol systems hypothesis. According to this hypothesis intelligent acts can be described and explained solely by symbol manipulation. Alan M. Turing provides the symbol manipulation approach with his concept of a universal Turing machine and a test which delivers a criterion for intelligence. Both, Turing and Searle choose the (methodological) comparison between a human and a machine to exemplify their (converse) points of view on the matter.

## 2 The Symbol Manipulation Approach to Intelligence

Based on the work of Alan M. Turing (inter alia the paper *Computing Machinery and Intelligence*) Allen Newell and Herbert Simon formulated the physical symbol system hypothesis (PSSH): “A physical symbol system<sup>3</sup> has the necessary and sufficient means for general intelligent action.”<sup>4</sup> This claim implies both that human thinking is a form of symbol manipulation (because a symbol system is necessary for intelligence) and that machines can be intelligent (because a symbol system is sufficient for intelligence).

According to the physical symbol system hypothesis (human) intellectual performance can be explained and described by symbol manipulation.<sup>5</sup> Symbol manipulation is an elaborate process which is reconciled with two different levels of description: On the one hand with the materialistic view about how a process is physically realised and on the other hand with

---

1 For the purposes of this paper thinking is identical to an intelligent act (and intelligence is a capacity for various intelligent activities).

2 Searle, John R. 1980. “Minds, Brains, and Programs”, *The Behavioral and Brain Sciences*.

3 A physical symbol system (also called a formal system) takes physical patterns (symbols), combining them into structures (expressions) and manipulating them (using processes) to produce new expressions.

4 Newell, Allen; Simon, Herbert A. 1976. *Computer Science as Empirical Inquiry: Symbols and Search*, "Communications of the ACM", *Communications of the ACM* 19 (3).

5 cf. Hoffmann, Achim 1993. *Komplexität einer künstlichen Intelligenz*, Berlin.

the description of the performance of the process by rule-governed operations on symbols. The PSSH dictates that symbol manipulation and thinking are processes of a fundamentally alike type. They are both physically realised and governed by rules and representations.<sup>6</sup>

A computer is a physical apparatus. Therefore, the performance of a computer can be described by the causal structure of its properties. The (physical) states of a computer are distinguished in terms of physical description. Its state transition is thus determined by physical laws. However, it is possible to prescind from these physical properties and to specify a functional description of the (physical) machine.<sup>7</sup> An appropriate state-transition diagram (Turing calls it “table of instructions”<sup>8</sup>) provides such a functional description of a computer. To explain which calculation a computer executes and which regularities a specific, particularly programmed computer features, one must refer on the objects of the domain in which the intended interpretation of calculations fall.<sup>9</sup>

It is explained in the same way, why and how persons do what they do. The Explanation, for instance why a chess player moves a specific piece on a certain square is determined by the role of the chess piece in the game and the particular objectives and intentions which the chess player pursues. Little attention, however, would be paid to the specific neurophysiological state of the player's brain or the geometrical properties of the moved chess piece in explaining the actions of the player.<sup>10</sup>

Philosopher and cognitive scientist Zenon W. Pylyshyn is also an advocate of the PSSH. He distinguishes fundamentally between two kinds of descriptions: (1) The description of what a computer does through description of his internal states and (2) the description of where the action of the computer refers to.<sup>11</sup> The former type of description relates to the intrinsic properties of the machine, whilst the latter is concerned with an entirely other domain, a domain independent of the physical construction of the apparatus.<sup>12</sup>

A parallel can be drawn between these fundamentally different processes and the human brain, which, on the one hand, can be described by its constitutive and neurophysiological properties, but also by the intentional description of human thinking. The different descriptions feature a problem, namely, of how it is possible that persons show a certain

---

6 cf. Hoffmann: *Komplexität einer künstlichen Intelligenz*, Berlin.

7 cf. Hoffmann: *Komplexität einer künstlichen Intelligenz*, Berlin.

8 Turing, Alan M. 1950. “Computing Machinery and Intelligence”, *Mind* 59/433-460.

9 cf. Hoffmann: *Komplexität einer künstlichen Intelligenz*, Berlin.

10 cf. Hoffmann: *Komplexität einer künstlichen Intelligenz*, Berlin.

11 cf. Pylyshyn, Zenon. W. 1984. *Computation and Cognition: Toward a foundation for cognitive science*. MIT Press/Bradford Books.

12 cf. Hoffmann: *Komplexität einer künstlichen Intelligenz*, Berlin; cf. Pylyshyn: *Computation and Cognition: Toward a foundation for cognitive science*.

comportment because of particular aims, desires and assumptions, whereas at the same time their behaviour is determined by physical laws and their neurophysiological states.<sup>13</sup>

In analogy, the state transitions of a machine depend simultaneously on the physical laws and the abstract properties of integers. This is only possible because both, the integers as well as the rules describing the relations between the integers, are represented in the machine as symbolic expressions and programmes. The physical realisation of these representations is finally responsible for that the machine processes indeed in accordance with the represented rules.<sup>14</sup>

Computers as physical symbol systems only apply rules because of the formal properties of the represented expressions. Consequently, as stressed by Searle, they can progress only syntactically and cannot differentiate between the interpretation of the expressions and the expressions themselves. Therefore it is imperative that all pertinent properties of the interpretation domain have a counterpart in the used expressions. In other words, there must be an isomorphic relation between the semantic level and the syntactic level.<sup>15</sup> In short, all semantic differences have to be retrieved on the syntactical level of the used expressions.<sup>16</sup>

The problem of intentional actions is reduced by the separation of the semantic and syntactic aspects of cognitive processes to name an appropriate mechanism which operates on meaningless symbols and in doing so, executes meaningful processes, e.g. calculating. This also means, that the computer works entirely independent of the interpretation of its symbols. In this sense, B. C. Smith writes in the prologue of *Reflection and semantics in a procedural language*, that an “internal process [...] is required to react only to the “form” or “shape” of these mental representations [i.e. symbols], without regard to what they mean or represent – this is the substance of the claim that computation involves formal symbol manipulation”<sup>17</sup>.

The syntactic nature of a computer's processes shows in this way how a relation between symbol manipulation and laws of causality is possible – at least in principle. Moreover, the explanation of cognitive laws necessitates not (at least not necessarily) the explanation of neurophysiological processes in the brain because it is possible to explain computations

---

13 cf. Hoffmann: *Komplexität einer künstlichen Intelligenz*, Berlin.

14 cf. Hoffmann: *Komplexität einer künstlichen Intelligenz*, Berlin.

15 cf. Block, Ned J. 1995. “The Mind as the Software of the Brain”. In Smith, E. E. and Osherson D. N., editors, *Thinking: An Invitation to Cognitive Science*, Cambridge: MIT.

16 cf. Hoffmann: *Komplexität einer künstlichen Intelligenz*, Berlin.

17 Smith, B.C. 1985. Prologue to “Reflection and semantics in a procedural language”. In R.J. Brachmann and H.J. Levesque, editors, *Readings in knowledge representations*. Morgan Kaufmann, Los Altos, CA.

without simultaneously considering the physical realisation of a computer.<sup>18</sup>

The paper presented a sketch of the symbol manipulation approach to intelligence. In the next section the investigation displays Alan M. Turing's criterion that – according to him – legitimates the attribution of intelligence to computers. In order to understand Turing's criterion, his concept of “computer” is in need of a more precise explanation.

### 3 The Turing Test

Alan M. Turing proposes in his paper *Computing Machinery and Intelligence* to consider the question “Can machines think?”<sup>19</sup>. In his approach, he replaces the question by another which is “expressed in relatively unambiguous words”<sup>20</sup> to avoid the vagueness of the ordinary concepts of “machine” and “think”. He attempts to describe the new form of the problem in terms of a game (“imitation game”<sup>21</sup>). Nowadays, the game is known as the Turing test.

The game's setting can be described as follows: An interrogating person, sitting in a closed room, communicates by teleprinter with a machine in a second room and a person in a third room. The object of the game for the interrogator is to determine which of the other two is the person and which is the machine. According to Turing, the original question “Can machines think?” can be replaced by a question, which is similar to the following: Is the interrogator able to distinguish between the machine and the person? In explicit terms, the criterion determining that a machine can think is the interrogator's inability to tell the difference.<sup>22</sup>

Turing restricts the kind of machines which are permitted to take part in the game to digital computers. He explicates the idea behind a digital computer by exemplifying an analogical human computer. A digital computer consists of three parts: (1) a store of information, (2) an executive unit and (3) a control. On the other hand, one can suppose that the human computer has no authority to deviate from his given rules and he has also a supply of paper on which to carry out his calculations.<sup>23</sup> The store corresponds to the human computer's paper, “wether this is the paper on which he does his calculations or that on which his book

---

18 cf. Hoffmann: *Komplexität einer künstlichen Intelligenz*.

19 Turing: “Computing Machinery and Intelligence”.

20 Turing: “Computing Machinery and Intelligence”.

21 Turing: “Computing Machinery and Intelligence”.

22 cf. Turing: “Computing Machinery and Intelligence”; cf. Block: “The Mind as the Software of the Brain”.

23 cf. Turing: “Computing Machinery and Intelligence”.

of rules is printed. In so far the human computer does calculations in his head, a part of the store will correspond to his memory.”<sup>24</sup>

The executive unit (also called central processing unit) carries out the specific operations involved in a calculation. As part of the digital computer's store the “book of rules” is called “table of instructions”. The control monitors and ensures that these instructions are followed correctly and in the right order. The book of rules is, of course, a convenient fiction. Real human computers just remember what they have got to do. If one wants to make a digital computer mimic the behaviour of the human computer in some operation, one has to ask him how it is done and then translate the answer into the form of an instruction table (state-transition diagram). To put the appropriate instruction table into the digital computer in the way that it will do X means to programme a machine to execute operation X.<sup>25</sup>

Digital computers are intended to execute any operations which could be done by a human computer. Turing points out that the level of similarity between a human and a digital computer is the level of “mathematical analogies of function”<sup>26</sup>, whereas the feature of using electricity is a superficial (and not necessary) similarity. Hence, (it's noteworthy for the issue of this paper, that) Turing holds a view of functionalism. This functionalism implicates that certain functions (of thinking) can be implemented (or realised) in a variety of different materials. As consequence of this thesis of multiple realisability and independency of (physical) manifestation respectively, he feels – in metaphorical opposition to Searle – that there is “little point in trying to make “a thinking machine” more human by dressing it up in [...] artificial [but humanlike] flesh.”<sup>27</sup>

Digital computers are discrete-state machines. Such a machine moves by sudden clicks from one (quite) definite state to another. Further, a discrete-state machine can only take up a finite number of possible states and is deterministic in the Laplacian sense. Given the table of a discrete-state machine it is possible to predict what it will do and there is – according to Turing – no reason why “this calculation should not be carried out by means of a digital computer”<sup>28</sup>. Provided the calculation could be executed sufficiently quickly, the digital computer could mimic the behaviour of any discrete-state machine. In this sense, digital computers are universal machines. Digital computers which are suitably programmed for each case and have the aforementioned property of universality enable the

---

24 Turing: “Computing Machinery and Intelligence”.

25 cf. Turing: “Computing Machinery and Intelligence”.

26 Turing: “Computing Machinery and Intelligence”.

27 Turing: “Computing Machinery and Intelligence”.

28 Turing: “Computing Machinery and Intelligence”.

execution of all computing processes without designing (various) new machines. As a consequence of this, all digital computers are in this sense equivalent.<sup>29</sup>

The concept of a universal machine (also called universal Turing machine), developed by Turing in 1937<sup>30</sup>, is generally considered as the theoretical model of present-day computers. The questions “Are there imaginable digital computers which would do well in the imitation game?”<sup>31</sup> or “Could a machine think?”<sup>32</sup> seem to be different from Turing's initial question “Can machines think?”<sup>33</sup>. But with respect to the concept of a universal machine (especially the universality property), these questions are equivalent to this one: Can one particular digital computer (i.e. a Turing machine) be made to play satisfactorily the imitation game?<sup>34</sup>

Since the attribution of thought to the computer is warranted by passing the Turing test, Turing hazards no metaphysical speculations as to what thought is – proposing no definition or conjecture as to the essential nature thereof. This is perhaps one indicator of why Turing's sharp-cut criterion for intelligence is not indisputable. As mentioned earlier, Searle criticises the adequacy of the Turing test.<sup>35</sup> He doubts whether a Turing machine is the appropriate model to describe and explain cognitive processes. He exemplifies his Chinese room argument in a similar way than Turing his universal Turing machine by describing a particular “human computer”.

#### 4 The Chinese Room Argument

The Chinese room argument is a thought experiment of John R. Searle. The argument parodies the Turing test and therefore also the symbol manipulation approach to intelligence. The argument is directed at what Searle defines as strong artificial intelligence (AI). It counters to the claim of strong AI that computers can or someday could think. The argument is based on two premisses: (1) brains cause minds and (2) syntax doesn't suffice for semantics.<sup>36</sup> For strong AI “the appropriately programmed computer really *is* a mind in the sense that computers given the right programs can be literally said to *understand* and

---

29 Turing: “Computing Machinery and Intelligence”.

30 cf. Hoffmann: *Komplexität einer künstlichen Intelligenz*, Berlin.

31 Turing: “Computing Machinery and Intelligence”.

32 Searle: “Minds, Brains, and Programs”.

33 Turing: “Computing Machinery and Intelligence”.

34 cf. Turing: “Computing Machinery and Intelligence”.

35 cf. Searle: “Minds, Brains, and Programs”.

36 cf. Searle: “Minds, Brains, and Programs”.

have other cognitive states.”<sup>37</sup>

Searle seizes Turing's analogy between a digital computer and a human computer. The person in the Chinese room is similar to Turing's human computer, he behaves like a computer. The setting of the Chinese room can be described as follows: A monolingual English speaker is locked in a closed room and is given (like Turing's human computer) a book of rules in English for correlating formal symbols on an initial stack of paper with symbols on a second stack of paper. According to Searle “formal” and “syntactic” respectively means that one “can identify the symbols entirely by their shapes.”<sup>38</sup> Further symbols on a third stack of paper and more instructions in English enable the person in the room to correlate elements of the third stack with elements of the first two stacks and to return certain sorts of symbols with certain sorts of shapes in response.<sup>39</sup>

The people giving the English speaker the symbols call – unknown to him – the first stack a “story”, the second stack a “script” and they call the third stack “questions”. They call the symbols the person hands them back “answers to the questions” and call the book of rules “the programme” (table of instructions). Searle contends that, if the person gets good enough at following the instructions for manipulating the symbols and the programmers get good enough at writing the programmes, than – from the external point of view, i.e. “from the point of view of somebody outside the room”<sup>40</sup> (e.g. Turing's interrogator) – the “answers” to the “questions” are “absolutely indistinguishable from those of native Chinese speakers”<sup>41</sup>.

Producing answers by manipulating uninterpreted formal symbols means according to Searle “as far as the Chinese is concerned”<sup>42</sup>, that the person in the room just behaves like a computer. Searle writes that it is “quite obvious” that the person in the Chinese room doesn't understand a word of the Chinese stories. The person has inputs and outputs that are indistinguishable from those of native Chinese speakers and he can have any formal programme, but he still understands nothing.<sup>43</sup> “For the same reasons”, Searle concludes, a “computer understands nothing of any stories”<sup>44</sup>, since in the Chinese case, the computer is a person and in the other cases where the computer is not a person, the computer has

---

37 Searle: "Minds, Brains, and Programs".

38 Searle: "Minds, Brains, and Programs"; cf. Smith: Prologue to “Reflection and semantics in a procedural language”.

39 cf. Searle: "Minds, Brains, and Programs".

40 Searle: "Minds, Brains, and Programs".

41 Searle: "Minds, Brains, and Programs".

42 Searle: "Minds, Brains, and Programs".

43 cf. Searle: "Minds, Brains, and Programs".

44 Searle: "Minds, Brains, and Programs".

nothing more than the person has in the other case where the person understands nothing. Furthermore, Searle argues that the programme doesn't explain (human) understanding. The Computer and its programme do not provide sufficient conditions for understanding, since the computer and the programme are functioning and there is no understanding. Searle criticises the PSSH, according to which the English speaker would, when he understands a story in English, "doing exactly the same – or perhaps more of the same"<sup>45</sup> as what he was doing in manipulating the Chinese symbols.

A programme is defined in terms of computational operations on purely formal elements. The Chinese room argument suggests that these elements by themselves "have no interesting connection with understanding"<sup>46</sup>. In this way Searle argues further that the formal principles of a computer will not be sufficient for understanding, "since a human will be able to follow the formal principles without understanding anything."<sup>47</sup> In short, formal symbols are certainly not sufficient conditions for understanding. Moreover, he states that no reason has been given to suppose that, when someone understands English, he is operating with any formal programme at all.

According to Searle, "understanding" implies the possession of mental, i.e. intentional states. Intentionality is the feature of certain (internal) mental states by which they are directed at objects and states of affairs in the world.<sup>48</sup> He holds that intentionality is constitutive for intelligence.<sup>49</sup> As a consequence of this premise, a computer can't be intelligent since the symbols it processes are meaningless (i.e. lack semantics); and it doesn't matter how intelligent-seeming a computer behaves and which kind of programming makes it behave that way. The computer's internal states and processes, being purely syntactic, lack semantics. It doesn't have intentional (i.e. for Searle meaningful) states.

## 5 The Systems Reply

Searle presents replies towards the Chinese room argument in his own article to invalidate them afterwards. One of them is the so called systems reply: The "individual person who is locked in the room does not understand the story"<sup>50</sup>, but "he is merely part of a whole

---

45 Searle: "Minds, Brains, and Programs".

46 Searle: "Minds, Brains, and Programs".

47 Searle: "Minds, Brains, and Programs".

48 cf. Searle: "Minds, Brains, and Programs".

49 cf. Hoffmann: *Komplexität einer künstlichen Intelligenz*.

50 Searle: "Minds, Brains, and Programs".

system, and the system does understand the story.”<sup>51</sup>

Searle responds by a modification of the original thought experiment: Imagine the person internalises the whole system by memorizing the rules and script and doing the lookups and other operations in his head. The person just uses the own body to listen to noises (“Chinese utterances”) and produce noises (“Chinese responses”). Searle persists on his viewpoint that the person “understands nothing of the Chinese, and [...] neither does the system, because *there isn't anything in the system that isn't in him*”<sup>52</sup>. If he doesn't understand, then there is no way the system could understand because the system is just a part of him”<sup>53</sup>.

Moreover, according to Searle, the systems reply would lead to the absurd consequence that “mind is everywhere”<sup>54</sup>. For instance, “there is a level of description at which my stomach does information processing”<sup>55</sup>. If a certain sort of input, output and a programme in between would be sufficient for an understanding system, then, Searle insists, “there is no principled way to distinguish the motivation for saying the Chinese subsystem understands from saying that the stomach understands.”<sup>56</sup>

From the point of view of the person in the Chinese room there is no information in either input – neither the food nor the Chinese. According to Searle, the information of the Chinese is solely in the eyes of the programmers and there is no reason why they shouldn't treat the input and output of the person's digestive organs as information, viz. physical patterns with meaning. But he maintains also that the mental-nonmental distinction can't be just in the eye of the beholder, it has to be intrinsic to the systems.<sup>57</sup>

About this point Searle writes: “We can't, on the one hand, say that anything is a digital computer if we can assign a syntax to it, and then suppose there is a factual question intrinsic to its physical operation whether or not a natural system such as the brain is a digital computer.”<sup>58</sup> That means that the property of being mental is attributed to a system only if it has the appropriate intrinsic physical property or properties. This claim contradicts the functionalist's thesis of multiple realisability.

Furthermore, it's just implausible for Searle that “while a person doesn't understand Chinese, somehow the conjunction of that person and bits of paper might understand

---

51 Searle: "Minds, Brains, and Programs".

52 I set the words in Italics.

53 Searle: "Minds, Brains, and Programs".

54 Searle: "Minds, Brains, and Programs".

55 Searle: "Minds, Brains, and Programs".

56 Searle: "Minds, Brains, and Programs".

57 cf. Searle: "Minds, Brains, and Programs".

58 Searle, John R. 1992. *The Rediscovery of the Mind*, Cambridge: MIT.

Chinese.”<sup>59</sup> Scenarios like the stomach-information processing, the conjunction of the person with a stack of paper and the realisation of a computer using a roll of toilet paper and a pile of small stones<sup>60</sup> aim to provoke intuitions: A stomach is not intelligent, paper does nothing contribute to intelligence and small stones are not the “right kind of stuff”<sup>61</sup> (cf. Turing's “artificial flesh”) for possessing mental states.

The inability of toilet paper and stones to have mental states and the Functionalist's claim that programmes might be implemented in such stuff yields the conclusion that computation (the “right” programming) does not suffice for thought. Searle concludes similarly that what “matters about brain operations is not the formal shadow cast by the sequences of synapses but rather the actual properties of the synapses.”<sup>62</sup>

Therefore, as seen, Searle rejects the Turing test for understanding Chinese. He argues that the programme of a real Chinese “understander” is not sufficient for understanding Chinese, no symbol-manipulation theory of Chinese understanding (or any other intentional state) is correct about “what makes something a Chinese understander.”<sup>63</sup> Thus the conclusion of Searle's argument is that the fundamental idea of thought as symbol processing is wrong, even if it allows us to build a machine that can duplicate the symbol processing of a person and thereby duplicate a person's behaviour.

According to the systems reply, the whole system does understand Chinese even though the person who is acting as the central processing unit (CPU<sup>64</sup>) does not. Philosopher Ned Block argues that, if the whole system understands Chinese, we should not expect the CPU to understand Chinese. The effect of Searle's internalisation move is an attempt to destroy the analogy between looking inside the computer and looking inside the Chinese room. If one looks inside the computer, one sees many chips in addition to the CPU. But if one looks inside the internalised Chinese room, all one sees is a person. Noteworthy is, that although the non-CPU components are no longer easy to see, they aren't gone, they are still there, they are only internalised.<sup>65</sup>

The systems reply indicates that it is invalid to reason from “John does not understand Chinese” to “The system of which John is a part does not understand Chinese.”<sup>66</sup> Further,

---

59 Searle: “Minds, Brains, and Programs”.

60 cf. Searle: “Minds, Brains, and Programs”.

61 Searle: “Minds, Brains, and Programs”.

62 Searle: “Minds, Brains, and Programs”.

63 Block: “The Mind as the Software of the Brain”.

64 The CPU of a computer is a device with a finite number of states whose activity is determined solely by its current state and input. Since the person is acting as the CPU, the output will be determined by the input and the “state”.

65 cf. Block: “The Mind as the Software of the Brain”.

66 cf. Block: “The Mind as the Software of the Brain”.

the systems reply looks upon the Chinese room as an English system implementing a Chinese system. The person is aware of the thoughts of the English system. In virtue of doing the symbol manipulation task, the person is also implementing a real intelligent Chinese-speaking system. The person's body, according to the systems reply, houses “two genuinely distinct intelligent systems.”<sup>67</sup> This means that the Chinese system thinks, but though the person implements this thinking he is not aware of it.

The English system implements the Chinese system because the English system is following the instructions of a programme in English to “make Chinese noises and not the other way around.”<sup>68</sup> In effect, Searle uses the fact that the person is not aware of the Chinese system's thoughts as an argument that it has no thoughts. But this argument is invalid. There are real counterexamples, e.g. multiple personalities are often cases in which one personality is unaware of the others. In fact, Searle does not refute the systems reply. His argument is disabled since it is invalid. He has no argument against the systems reply – only his intuitions.

## 6 Conclusion

Turing himself writes that he has “no very convincing arguments of a positive nature to support [...] [his] view.”<sup>69</sup> The PSSH is and remains until today – as the name indicates – a hypothesis. But the Chinese room argument with respect to the systems reply can't serve as a destructive argument against the symbol manipulation approach to intelligence.

In fact, if Searle's argument merely discountenances theoretic or metaphysical identification of thought with computation, the behavioural evidence – and consequently Turing's point – remains unaffected. Since computers seem to think, the conclusion that the essential nonidentity of thinking with computation would seem to warrant is, that whatever else thinking essentially is, is possessed by computers (hypothetically) too.

In this last section, the paper presents a serious problem of the Turing test and indicates to the meanings and the relation of “intelligence” and “intentionality” in order to unveil a question for further investigation to clarify the question of whether a machine could think (at least in a different way than humans).

The Turing test is an attempt to define intelligence. Behaviourists define the mental activities and abilities in terms of behavioural dispositions, i.e. the tendency to emit certain

---

67 Block: “The Mind as the Software of the Brain”.

68 Block: “The Mind as the Software of the Brain”.

69 Turing, “Computing Machinery and Intelligence”.

behaviours in response to certain stimuli<sup>70</sup> since something can be intelligent even if it has never had the chance to manifest its intelligence. It is important to behaviourists that the stimuli and the behaviour be specified in non-mentalistic terms. Therefore, intelligence could not be defined in terms of the disposition to give responses to the questions of the interrogator, since that would mean to define a mental notion in terms of another mental notion.

According to the Turing test, a computer is intelligent if and only if the interrogator can't tell the difference between the utterances given by the computer and the answers given by a person. That means, that anything is deemed intelligent, if it can pass the Turing test, viz. the Turing test is a sufficient condition for intelligence. Turing proposed to replace the concept of intelligence with the concept of passing the Turing test. His definition circumvents the problem of specifying non-mentalistically the behavioural dispositions that are characteristic of intelligence by bringing in the discrimination behaviour of a human interrogator – in this context – better called a judge.<sup>71</sup>

Nonetheless, Ned Block spots a gap in Turing's proposal. Passing the Turing test depends on the discrimination behaviour of the human judge. But what kind of properties or dispositions should the judge have? What kind of properties determine the discrimination behaviour? One way of dealing with this problem of characterisation of the judge consists of including a specification of the mental qualities or dispositions of the judge in the description of the test. However, this will ruin the test as way of defining the concept of intelligence in non-mentalistic terms.

According to Block, the problem with the Turing test is not (as it is for Searle) that despite a machine's ability to pass the test, this machine can remain unintelligent nonetheless; the problem for theoretical purposes is that the Turing test focuses on performance rather than on competence. Performance is evidence for competence indeed, but “the core of our understanding of the mind lies with mental competence, not behavioural performance.”<sup>72</sup> Therefore, the Turing test is somewhat inadequate in its definition of intelligence.

Nonetheless, the Turing test avoids the problem of solipsism. According to the extreme form of the solipsist view, the only way by which one could be sure that a man thinks is to be that particular man. Likewise according to this very secure view to know that a machine thinks is to be the machine and to feel oneself thinking.<sup>73</sup> It's not, Searle counters, “how I

---

70 cf. Block: “The Mind as the Software of the Brain”.

71 cf. Block: “The Mind as the Software of the Brain”.

72 Block: “The Mind as the Software of the Brain”.

73 cf. Turing: “Computing Machinery and Intelligence”.

know that other people have cognitive states, but rather what it is that I am attributing when I attribute cognitive states to them.”<sup>74</sup>

What makes something to have cognitive states indicates to the relation between intelligence and intentionality. According to Searle, intelligence requires intentionality. As a conclusion, if a intentional (cognitive) state is meaningful, then the corresponding intelligent act(s) is/are also meaningful. Searle calls this kind of intelligent acts “understanding”. He attributes this predicate to the English speaker in the Chinese room, though not to the whole system speaking Chinese in his thought-experiment.

But maybe intelligence does not necessarily imply intentionality. In this sense, an “intelligent act” is something different than “understanding”. To understand would imply to think, but not vice versa. There is even a reason to represent this view. Every intentional state has an intention and an intentional content. However, there are bodily sensations, e.g. the experience of orgasm, that are genuinely mental states but have no intentional content. Therefore, while intentionality is an important feature of many mental states, but it's not the essence of the mental.<sup>75</sup>

In order to manifest the possibility of “intelligence” without “intentionality” it is necessary to define the terms more precisely than in this paper. Only one important difference should be alluded namely that an intelligent system must have certain intelligent capacities to do certain operations. In other words, the criterion for an intelligent system is that it can do something, i.e. has the capacity to do something. In this sense, intelligence is necessarily future-oriented.

In contrast, the criterion of an intentional system could be in part its causal history. This controversial thesis if for instance hold by Block: A System “must have a history that makes its states represent the world, i.e. have aboutness.”<sup>76</sup> In this sense intentionality has a past-oriented requirement. The precise enough question for a further investigation arises whether a system can satisfy the future-oriented needs for intelligence, while missing the past-oriented requirement of intentionality. It is noteworthy in this context, that the symbol manipulation approach to intelligence – which tries to explain intelligent processes by reducing them to unintelligent mechanical processes – does not explain intentionality.<sup>77</sup>

---

74 Searle: "Minds, Brains, and Programs".

75 cf. Block: “The Mind as the Software of the Brain”.

76 Block: “The Mind as the Software of the Brain”.

77 cf. Block: “The Mind as the Software of the Brain”.

## References

Block, Ned J. 1995. "The Mind as the Software of the Brain". In Smith, E. E. and Osherson D. N., editors, <i>Thinking: An Invitation to Cognitive Science</i> , Cambridge: MIT.
Hoffmann, Achim 1993. <i>Komplexität einer künstlichen Intelligenz</i> , Berlin.
Newell, Allen; Simon, Herbert A. 1976. "Computer Science as Empirical Inquiry: Symbols and Search", <i>Communications of the ACM</i> 19 3/113-126.
Pylyshyn, Zenon. W. 1984. <i>Computation and Cognition: Toward a foundation for cognitive science</i> . MIT Press/Bradford Books.
Searle, John R. 1980. "Minds, Brains, and Programs", <i>The Behavioral and Brain Sciences</i> 3/417-457.
Searle, John R. 1992. <i>The Rediscovery of the Mind</i> , Cambridge: MIT.
Smith, B.C. 1985. Prologue to "Reflection and semantics in a procedural language". In R.J. Brachmann and H.J. Levesque, editors, <i>Readings in knowledge representations</i> . Morgan Kaufmann, Los Altos, CA.
Turing, Alan M. 1950. "Computing Machinery and Intelligence", <i>Mind</i> 59/433-460.