



## The Three Minds Argument

Jamie Cullen  
Artificial Intelligence Laboratory  
The University of New South Wales  
[jsc@cse.unsw.edu.au](mailto:jsc@cse.unsw.edu.au)

*Journal of Evolution and Technology* - Vol. 20 Issue 1 – June 2009 - pgs 51-60  
<http://jetpress.org/v20/cullen.htm>

### Abstract

Searle (1980, 2002, 2006) has long maintained a position that non-biologically based machines (including robots and digital computers), no matter how intelligently they may appear to behave, cannot achieve “intentionality” or “consciousness,” have a “mind,” and so forth. Standard replies to Searle’s argument, as commonly cited by researchers in Artificial Intelligence and related communities, are sometimes considered unsatisfactory by readers outside of such fields. One possible reason for this is that the Chinese Room Argument makes a strong appeal to some people’s intuitions regarding “understanding” and necessary conditions for consciousness. Rather than contradict any such intuitions or conditions, I present what in my view is an independent and largely compatible intuition: If Searle’s argument is sound, then surely a human placed under similar testing conditions as a non-biological machine should succeed where a machine would allegedly fail. The outcome is a new rebuttal to the Chinese Room that is ideologically independent of one’s views on the necessary and sufficient conditions for having a “mind.”

### 1 Introduction

Searle’s Chinese Room Argument (CRA) claims to examine and reject the assertion that:

[The] appropriately programmed computer really is a mind, in the sense that computers given the right programs can literally be said to *understand* and have other cognitive states. (Searle 1980.)

The CRA has been re-described with many variations over the years, and is perhaps the most frequently cited argument against the possibility of “Artificial General Intelligence” (AGI) and related notions.<sup>1</sup> While many in the Artificial Intelligence (AI) community may readily dismiss Searle’s claims, perhaps

citing well-known replies, such as the Robot Reply or the Systems Reply (described later), many of the better-known replies have counter arguments provided by Searle (such as in his original 1980 paper). Regardless of whether or not one accepts such counter arguments, after more than twenty five years of intense debate, the CRA apparently still refuses to die.

In my opinion, a significant amount of this continued debate is caused by two inter-related factors: (a) Some opponents of the CRA are possibly not aware of the key belief underlying Searle's argument; and (b) A strong appeal is frequently made to a "commonsense" intuition that sometimes misleads people into incorrectly accepting the CRA.

The dual purpose of this paper is to (a) draw explicit attention to the earlier mentioned underlying belief, and (b) provide an alternative appeal to a commonsense intuition that makes the sleight of hand underlying the Chinese Room more readily apparent, whilst neither affirming nor contradicting the underlying belief. I hope that the argument presented here is found to be intuitive by people both inside and outside of the Artificial Intelligence research community.

I will conclude the paper by discussing the implications of the presented argument, and by re-examining the roles of the various possible participants in the CRA. I will then draw some conclusions regarding the structure of the CRA, and the relevance (or lack thereof) of "intentionality" and related philosophical topics that are commonly raised in connection with the argument.

## **2 The Chinese Room**

### **2.1 Overview**

The essence of the CRA is that Searle imagines himself being in a room into which Chinese symbols are fed. He does not speak Chinese but has a set of rules to follow (which I will refer to herein as the "rule-book"). By following this rule-book he is theoretically assumed to be able to transform one set of Chinese symbols into another well enough for a conversation to occur between a human Chinese speaker outside the room, and "something in the room." Depending on the stance one takes in debating the matter, this "something" might be considered as the rule-book itself, a system combining Searle, the rule-book, and the room as a whole, and other potential variations. In the literature surrounding the CRA, as well as in Searle's original paper, the rule-book might also be replaced by a digital computer, a robot, or some other non-biological artifact. I shall refer collectively to the rule-book and such variants on the rule-book as "artifacts under examination" (or more briefly as "artifacts").

While it is beyond the scope of this paper to attempt a full summary of the debate surrounding the Chinese Room (such a summary would be extremely difficult to accomplish in a normal journal-sized article), I will provide a brief description of two of the more well-known replies: the Systems Reply and the Robot Reply (Cole 2004).<sup>2</sup>

In the Systems Reply it is argued that, while the person in the room (Searle) might not understand Chinese, the system as a whole understands. The fact that Searle himself does not understand Chinese is irrelevant in the same sense that an individual section of a human brain (let's assume a part of the brain not typically associated with linguistic ability) might not understand Chinese when considered in isolation. However, the brain when considered as a whole system, does "understand." Searle's reply to this line of reasoning is to ask us to imagine that he (in the room) now memorizes the rule-book and keeps track of everything in his own head (Searle 1980). He argues that, since the only key component of the system is effectively himself, and since he still would not "understand" Chinese in any conventional sense (Searle is apparently considering his own English-speaking personal viewpoint in this case), the fact that the system as a whole appears to "understand" Chinese proves nothing.

In the Robot Reply we are asked to imagine that the “rule-book” (computer program) is placed into a robot that can interact in a sensorimotor sense with its environment. This might allow a connection of the symbols with real sensorimotor experience. Searle’s reply is to ask us to still consider the person in the room, but to ask us to imagine that some of the inputs to the room are received directly from a camera mounted on a robot, and that some of the outputs are used to actuate the robot’s limbs. He then argues that Searle (in the room) still does not know what those symbols “mean,” regardless of their source. He is simply following the rules of the rule-book (Searle 1980).

## 2.2 A Two-Participant Turing Test in Chinese

We might consider the Chinese Room Argument to be a kind of “Two Participant Turing Test.” In the original “Turing Test” (also known as the Imitation Game), we are asked to imagine a game of three participants: an Interrogator, a Human player, and a Machine player. The three participants are all physically separated, cannot see one another, and may communicate only through a textual communication facility (e.g. by “typing” textual messages to one another). The Interrogator’s task in the game is to attempt to correctly distinguish the Human from the Machine using only the communication in the game (Turing 1950). To form the Chinese Room variant of the Turing Test, we remove the Human player, and change the language of testing to Chinese. The Chinese Speaker outside takes on the role of Turing’s original “Interrogator,” and the “Something in the Room” takes on the role of the Machine. The Chinese Room Argument then looks inside the Machine player to attempt to show that the Machine does not really “understand,” even if it can fool the Interrogator that it is human over a channel of linguistic communication. Searle then takes on various roles within the Machine in an attempt to show that regardless of an allegedly successful conversation between Machine and Interrogator, he still does not “understand” the conversation. His underlying (perhaps implicit) assumption is that if he (in the room) does not “understand,” or cannot identify exactly “where” or “how” “understanding” is taking place, then “understanding” must not be taking place.

In Searle’s words:

... formal symbol manipulations by themselves don’t have any intentionality; they are quite meaningless; they aren’t even symbol manipulations; since the symbols don’t symbolize anything ... The aim of the Chinese room example was to try to show this by showing that as soon as we put something into the system that really does have intentionality (a man), and we program him with the formal program, you can see that the formal program carries no additional intentionality. *It adds nothing, for example, to a man’s ability to understand Chinese.* (Searle 1980; emphasis mine.)

## 3 Presenting the underlying ideology: *Élan vitale* in the Chinese Room

Searle’s argument is sometimes over-broadened (perhaps misconstrued) as a rebuttal of the concept of man-made intelligence in general. That is, the argument is seen as an attack on Artificial Intelligence in the broader sense of something artificially constructed by human beings, rather than more familiar artifacts such as a robots, digital computers, analog computers, Universal Turing Machines, etc. However, Searle’s argument does not seem so broad as to include all kinds of man-made artifacts. Indeed, he states:

Assuming it is possible to produce artificially a machine with a nervous system, neurons with axons and dendrites, and all the rest of it, sufficiently like ours, again the answer to the question (Could an artifact, a man-made machine, think?) seems to be obviously, yes. (Searle 1980.)

More recently he has stated: “There is no question that machines can think, because human and animal brains are precisely such machines” (Preston and Bishop 2002).

Where Searle's views apparently diverge from those of many AI and AGI researchers is that he appears to believe that there is something physical, special and yet to be discovered that can't be simulated or duplicated by certain kinds of artifact. Indeed, at a recent conference, Searle expressed disappointment in the apparent lack of neuroscience research applied towards the search for "Neural Correlates of Consciousness" (NCC). He told the same audience that an important goal for neuroscientists should be to search for this hypothesized NCC (Searle 2006). In Searle's words:

Could we produce consciousness artificially in some medium other than carbon-based organic compounds? The short answer to that question is that we just do not know at present. Since we do not know how the brain does it, we do not know what sorts of chemical devices are necessary for its production. Perhaps it will turn out that consciousness can only be produced using electrochemical phenomena, but does not require organic compounds. It might be that certain types of organic compounds in certain conditions are causally sufficient without it being the case that they are causally necessary. At present we just do not know. (Preston and Bishop 2002.)

In my opinion, Searle's quest for a biological causal agent of consciousness or NCC can be viewed as a modern re-expression of an older idea, not widely held in high esteem these days by the scientific community: *Élan vitale*.

*Élan vitale* is a term attributed to the philosopher Henri Bergson from his book *Creative Evolution*, first published in the early 20th century (Bergson 1911). *Élan vitale* (normally translated from French into English as "vital force" or "vital impetus") is a hypothetical explanation for the development of organisms and evolution, which Bergson linked closely to consciousness.<sup>3</sup> It is essentially an elusive mystical force that acts on inanimate matter and gives rise to life and consciousness. The overall notion is very similar to the more modern notion of an NCC. In this paper I will refer collectively to all ideas that postulate a mysterious something in biological matter that gives rise to "consciousness," "minds," "life" and so forth (including the idea of an NCC) as *Élan vitale*.

#### 4 An ideological spectrum

I find it useful to visualize *Élan vitale* as lying on an ideological spectrum of ideas regarding necessary and sufficient conditions for possessing a "mind." On the left side we might place more liberal and inclusive notions regarding "minds," particularly those related to structure or strictly information-based ideas. Walking from left to right we progressively add stricter requirements, moving towards more anthropocentric notions of mental phenomena. As the spectrum is deliberately ideological in nature there will naturally be strong disagreements between proponents of any one ideology about which requirements are necessary, sufficient, or even real. At the extreme left, we might imagine that suitably arranged matter shaped into an appropriate structure could produce "consciousness" or a "mind." Near the left end of the spectrum, we might place ideas such as the Physical Symbol System hypothesis (Newell and Simon 1963). In these and related ideas, "consciousness" is probably considered as an illusory notion or perhaps an emergent property of appropriately organized matter.

Walking towards the right we might add embodiment (both physical and virtual) to our list: i.e., a computer might not be capable of having a "mind," but a computer hooked up to sensorimotor apparatus, such as an epigenetic robot, or virtually embodied in a simulator, might fit the bill. Slightly further to the right, we might remove the virtual requirement, and allow only physically embodied machines. At the extreme right of the spectrum we might place the dualistic notion that only a human being with a "soul" may possess a "mind." Shortly before Cartesian Dualism seems an appropriate position for the notion of *Élan vitale* or the NCC. Of course, like most analogies that attempt to place a complex set of ideas on a linear spectrum, this one is possibly an over simplification. However, I find this mental picture to be a

useful reminder that, in discussing the Chinese Room, we are often talking about opposing ideologies that are presently very difficult (if not impossible) to prove one way or the other, regardless of which one (if any) we might personally subscribe to.

## **5 Embracing the underlying ideology: Applying the Chinese Room to a human being**

Part of the difficulty in artifact-related replies (which include many of the popular replies to the CRA, such as the Robot and Systems Replies), is that Searle's counter replies always seem to come back to his underlying belief in *Élan vitale*, which is normally grounded in a different position on the ideological spectrum to his opponent's argument. As many of us have likely experienced, arguments in which people take opposing positions, but for which incontrovertible evidence is not readily available either way, often make for an intense debate with no obvious resolution in sight. This appears to be the case with the Chinese Room debate.

While non-biological artifact-based arguments may never convince Searle to abandon his apparent belief in *Élan vitale*, one may choose to frame the debate in a different way in order to see more clearly if his argument (rather than any specific ideology) really does hold up. As Searle's argument appears to hinge on whether or not Searle (in the room) "understands" the Chinese conversation taking place, the question I ask is relatively simple: What would happen if we applied the Chinese Room test criteria to a human being? Would Searle in the room be able to understand the conversation?

Such a framing of the debate removes the ambiguity and confusion that may result from taking about non-biological artifacts, or getting caught up in discussions about what may/may not be capable of "intentionality," "minds," "consciousness," etc. Rather than going around and around on an ideological Ferris Wheel debating what can/can not have "consciousness," we instead choose to test something (actually "someone" in this case) who, by definition, must have it, and see what the test says about them and their ability to have a mind.

Let us imagine what might happen if we placed a human being under the same test conditions that Searle chooses to test an artifact. From there we can examine whether Searle in the room might be able to understand Chinese, and consequently whether or not the Chinese Room is *structurally* valid from this perspective. If the Chinese Room test cannot be passed by a human being, then what hope might an artifact have?

## **6 The Three Minds Argument (TMA)**

Imagine a modified form of the CRA. In this particular thought experiment, we still have a Chinese Speaker outside the room (whom we shall name "Li"). Searle is also in the room as before. As with the original experiment, Searle in the room speaks no Chinese. However, instead of an artifact to be tested (rule-book, robot, computer, etc.), we shall imagine a Chinese person called "Nao" who is in the room with Searle. Unfortunately, Nao had a severe accident which damaged his body beyond repair. Through some miracle of modern medicine and cybernetics, however, his brain is able to live on in a jar. For the purposes of this experiment, we will assume that Nao is a living conscious human being during the time of the test. Naturally, Nao still possesses a real human mind, despite his immense physical handicap. In the modern world, which frequently confers equal rights for disabled people (not to mention any emerging views on cyborg or potential transhumanist rights), there are probably few people in the academic community not willing to consider Nao as a human being, rather than sub-human or somehow no longer capable of "consciousness," having a "mind," etc.

To expand on the story above, we might further imagine Searle in the room as the full-time caregiver for Nao (a tough job, but presumably far better than being stuck in a jar). Searle's job is both to input the

symbols into Nao's jar, and to interpret the outputs from the jar as written symbols and put them outside the room. As this is a thought experiment, let us imagine Nao has sensors on his jar such that he can receive symbols in Chinese via Searle through the pathways that used to correspond to sensory input (perhaps the visual pathways) and understand them. Nao also has a similar ability for outputting written symbols, say using some neural connections formerly used for motor action. These simple sensors and actuators take on a similar role to the pencil and paper in the rule-book metaphor, in that they facilitate the updating of state and output of conversational symbols from the brain in the jar. Searle's job is still to pass the symbols back and forward from the CRA to the person outside the room (Li).

For the sake of argument, we might even imagine that Searle also has the job of maintaining the life support system of Nao's jar. He repeatedly pumps a foot pedal to power the mechanism that is keeping Nao alive, among some other life support maintaining duties (flow of oxygenated blood to the brain, etc.). That is, without Searle around, Nao would die, and he or she wouldn't be able to continue a conversation with Li. It should be apparent that Nao is entirely dependent on Searle in order to continue in the conversation with Li. One might imagine the room as a kind of hospital room instead of a nondescript "Chinese Room." Note that in the original CRA, Searle must follow the rules of the book with no deviation on his part allowed. Likewise, in the hospital room, Searle is bound (say, by hospital codes of conduct) to perform these duties with no deviation from the rules allowed.<sup>4</sup>

Note that Searle's job in the original CRA is arguably a "mindless" task: he does not attempt to understand the Chinese Symbols being passed to him, but simply follows a set of rules laid out in the rule-book. Likewise, by pumping the foot pedal, and passing symbols to and from Nao, Searle is performing a similarly "mindless" task that is, nevertheless, necessary for Nao to continue communicating with Li (and stay alive).

The intent behind the TMA is that structurally the CRA and the TMA are as close as possible. I suggest that we should not get too hung up on the exact details of the above description. The essence of the TMA is that we wish to guarantee that the entity being tested must already have "intentionality," "consciousness," and whatever other properties and attributes one might decide to assign to a human being's brain (and central nervous system, if you like). So instead of two minds and an artifact being tested for "mind" capability, we have three minds already by definition: Li, Nao, and Searle.

Now that we have three "guaranteed" minds in the experiment, a reasonable question to ask is this: *Would Searle in the room understand the conversation that takes place between Nao and Li?*

The obvious answer appears to be "No." Li and Nao would converse (albeit awkwardly) in Chinese, while Searle would in a sense facilitate, but still not understand the conversation occurring in a language foreign to him.

## **7 So who is communicating with whom?**

The intuition that Searle still would not understand Chinese, even if a human was used in the room in place of an artifact, is a key component of the Three Minds Argument. It seems intuitive that Searle in the room would not understand the conversation as he normally does not speak Chinese, and is actually not part of the direct conversation. However, Searle's argument for an artifact not being able to have a "mind," etc. hinges upon precisely the same point. It seems that in both the TMA and CRA, Searle in the room merely facilitates the conversation, rather than participates.

As Pinker (1997) points out, we often view human beings in different ways in different contexts. For example, in the context of a serious moral dilemma, it makes sense to view a human being as a living, thinking entity with emotions and a mind. In the context of an airplane engineer designing a plane, the

passengers are viewed as weight, little more. In the context of using a wrench, the human is little more than a torque applying tool. I posit that in the context of the CRA, Searle can be (perhaps rather ironically) viewed as little more than a “mindless” processor or simple homunculus. Searle’s own linguistic ability is mostly irrelevant in the context in which he is being used in the experiment, other than perhaps in being able to read the rules. As in the original rule-book formulation of the CRA, his task is simply to pass symbols to and from the “rule-book” (artifact) and to follow the rules as given to him without any choice on his part to deviate from them.

Searle in the room is still a human being; however, his role in the conversation is more that of a conduit of information, rather than a conversational participant. In my view, such a reframing of the situation draws attention to the underlying sleight of hand in the Chinese Room: assuming any communication is actually possible, it would be occurring between Li and Nao, not Li and Searle. As Searle observes, it is perhaps intuitive that in his role in the Chinese Room he would not “understand” the conversation. However, this is because he does not speak Chinese, and is not actively participating in the conversation. In my view, the “commonsense” intuition that Searle in the room would not “understand” is actually a valid and correct intuition. However, Searle’s inference as to the reason behind this lack of “understanding” is not. When we start with something that Searle presumably must concede has “intentionality” (a person), Searle in the room still cannot understand the conversation. The properties that the artifact from the original Chinese Room might possess become largely irrelevant when viewed in this context: the question to ask in the CRA is not related to “intentionality,” it is about who is talking to whom.

## **8 Possible Objections**

### **8.1 Missing Embodiment Reply**

One concern that might be raised is that Nao appears to have lost conventional human sensorimotor embodiment himself. It might then be argued that Nao has now consequently also “lost” some abilities such as “intentionality.” In response, we could imagine a variant of the TMA in which Nao’s life support system is actually a giant robot in the room. Inside the robot is Searle sitting in the cranial cavity with Nao’s brain, all connected up in some appropriate way. We now have a situation somewhat similar to Searle’s response to the original Robot Reply: Searle acts (perhaps somewhat ironically) as a kind of “homunculus” inside Nao’s robot body, passing symbols to and from Nao’s robotic sensory apparatus to his brain interface. He still does not understand the conversation, even though Nao has a real “mind.”

Note that this particular reformulation of the TMA is only one of many possible variations on the idea. I suggest that one may consider modifying the equipment any way one likes to allow Searle a larger/smaller role in the functioning of Nao’s existence, as long as the reader remains convinced that Nao has a “mind.” I suspect that in all such cases, the conversation will still be between Nao and Li, with Searle unable to understand it.

### **8.2 Intuition Pump Reply**

In the pejorative sense of the term, “intuition pumps” are philosophical arguments that essentially mislead us into accepting an idea by making an appeal to our “commonsense” intuitions, while helping us to conveniently overlook important details. Indeed, the concept of the intuition pump has been used before to criticize the Chinese Room (Dennett 1991). It might also be argued that the Three Minds Argument is an intuition pump as it too makes a strong appeal to intuition. While I have no objection to the claim that the Three Minds Argument makes an appeal to intuition, not all appeals to intuition are necessarily incorrect or misleading: just because something may be intuitive, doesn’t mean it is wrong. A claim that an argument is an intuition pump clearly requires a strong supporting argument pointing out where and how the argument might be misleading, if such a criticism is to hold up to scrutiny.

### 8.3 Isn't this just the Systems Reply?

In the Systems Reply Searle in the room as is essentially considered as a component in a larger machine: while Searle in the room does not “understand,” the system as a whole does. The key distinction between the Systems Reply and the Three Minds Argument is that the Systems Reply is described with the assumption that a non-biological artifact is being tested. For someone ideologically orientated towards the right hand side of the described “ideological spectrum,” these arguments may still appear as “obviously” flawed, due to the source material simply not having the “right” properties for consciousness. For a person more inclined towards the left side of the spectrum, such a concern seems more likely to be a non-issue. The Three Minds Argument attempts to side step such concerns by simply not relying on a non-biological artifact. If the test is to be a fair one, it should be structurally valid for people as well as for machines.

### 8.4 Didn't you put the “intentionality” back in?

Part of Searle's original argument is that he believes that human beings have “intentionality.” As a computer can't have “intentionality” (by Searle's reasoning), the test cannot be passed. Now that we've effectively “put the intentionality back in” by replacing the artifact with a human being, Searle might simply agree that the test can now be passed, as the system now has “intentionality” (or some other quality that we might believe that a human being possesses that an artifact does not). This response misses the point of my argument: In my view, “intentionality” and related issues are “red herrings” in the Chinese Room. The key is not what special properties the “mind” being tested might possess, it is about Searle in the room and his “understanding” of the conversation. The Three Minds Argument attempts to show that Searle still doesn't understand the conversation irrespective of what properties the entity under examination might possess. When we can see that even a person cannot pass Searle's test, the structural flaw in the Chinese Room becomes clear.

## 9 Conclusion

Searle (1980, 2002, 2006) has long maintained a position that non-biologically based machines (including robots and digital computers), no matter how intelligently they may appear to behave, cannot achieve “intentionality” or “consciousness,” or have a “mind.” His position is apparently rooted in an underlying belief that presently undiscovered biological mechanisms give rise to such phenomena.

I suggest that Searle's search for “Neural Correlates of Consciousness” may be characterized as a modern reformulation of a significantly older idea, normally going by the name of *Élan vitale*. The Physical Symbol System Hypothesis and *Élan vitale* can be imagined as occupying roughly opposed positions on a spectrum of ideologies related to necessary and sufficient conditions for “minds.” In his original CRA paper, Searle quipped of the Artificial Intelligence Research Community that it would be “difficult to see how anyone not in the grip of an ideology would subscribe to such an idea” (Searle 1980). However, as discussed throughout this paper, Searle's viewpoint appears to be grounded in a similar, if opposing, ideological position.

In the Three Minds Argument, I attempt to side-step the problem of arguing for or against specific ideologies regarding the “mind” and take a different approach. The man-made artifact described in the CRA is replaced by a human being (who is clearly capable of “intentionality” or any other cognitive properties one might believe that human beings have). I then argued that Searle in the room performs an analogous role, but apparently still does not understand Chinese. In a sense, the Three Minds Argument is a form of *reductio ad absurdum* for the Chinese Room. We start with someone we are convinced has a “mind,” and work backwards, until we see that Searle in the room still does not understand Chinese. If a

human being can't pass the test, then the test must be logically flawed, assuming one accepts the structural equivalence of the two thought experiments.

The overall intent of this paper is neither to critique nor to question any specific beliefs about consciousness, minds, and so forth, but instead to demonstrate that the Chinese Room does not prove anything about such things. In the meantime, the question of whether or not Artificial General Intelligence is possible remains both open and provocative.

### **Acknowledgments**

Thanks to Alan Blair and Peter Slezak for their feedback on earlier drafts of this paper.

### **Notes**

1. I have deliberately avoided using Searle's term "Strong Artificial Intelligence" here, as I believe that confusion around what this term signifies has muddied the Chinese Room debate. I have instead opted for a more modern term, which I believe better reflects the type of research at which Searle's argument is normally directed today, irrespective of Searle's original intent.

2. Readers may wish to consult (Cole 2004) and (Preston and Bishop 2002) for relatively recent summaries and more in-depth discussions of the various replies.

3. *Élan vitale* has also been described as a possible inspiration for the fantasy concept of "The Force," as represented in the 1977 movie *Star Wars* (incidentally, released a few years before publication of the CRA).

4. I suspect that most reformulations of the CRA (Robot version, etc.) can be amended to a three minds version. As the artifact is supposed to be a complete "intelligent" machine of some kind, it seems perfectly reasonable to replace it with a human brain and see if the test still makes sense. If such a substitution cannot be made, then I believe the onus is on the author of such an experiment to show exactly why such a variation of his test should be considered, and how it is equivalent to the original Chinese Room thought experiment.

### **References**

Bergson, H. 1911. *Creative evolution*. Translated by A. Mitchell. Henry Holt and Company. New York.

Dennett, D. 1991. *Consciousness explained*. Penguin Books. London.

Cole, D. 2004. The Chinese Room Argument. *The Stanford Encyclopedia of Philosophy* (Summer 2004 Edition), ed. E. Zalta. <http://plato.stanford.edu/archives/sum2004/entries/chinese-room/>

Newell, A., and H. Simon. 1963. GPS: A program that simulates human thought. In *Computers and thought*, ed. E. Feigenbaum and J. Feldman. McGraw-Hill. New York.

Pinker, S. 1997. *How the mind works*. Penguin Books. London.

Preston, J. and M. Bishop. 2002. *Views into the Chinese Room: New essays on Searle and Artificial Intelligence*. Oxford University Press. Oxford.

Searle, J. 1980. Minds, brains, and programs. *The Behavioral and Brain Sciences* 3 (3): 417-57.

Searle, J. 2006. Talk given at Towards a Science of Consciousness (conference). *Philosopher's Zone*, ABC National Podcast. <http://www.abc.net.au/rn/philosopherszone/index/date2006.htm> (Retrieved: 2<sup>nd</sup> May 2009).

Turing, A. 1950. Computing machinery and intelligence. *Mind* 59: 433–60.