Picha, Marek

The Gramophone : Lem against the Chinese Room

Pro-Fil. 2021, vol. 22, iss. Special issue, pp. 54-64

ISSN 1212-9097 (online)

Stable URL (DOI): https://doi.org/10.5817/pf21-3-2420
Stable URL (handle): https://hdl.handle.net/11222.digilib/144850

Pro-Fil (Special Issue) 2021: 54–64
https://doi.org/10.5817/pf21-3-2420

THE GRAMOPHONE: LEM AGAINST THE CHINESE ROOM

MAREK PICHA

Department of Philosophy, Faculty of Arts, Masaryk University, Brno, Czech Republic,
picha@phil.muni.cz

RESEARCH PAPER ▪ SUBMITTED: 15/10/2021 ▪ ACCEPTED: 19/11/2021

Abstract: The text focuses on Lem's rejection of the Chinese Room, a prominent challenge to the sufficiency of the Turing test. After outlining Lem's relationship to the Turing test, it offers an exposition of two of Lem's thought experiments, the Gramophone and the Jigsaw, whose critique is directly related to the critique of the Chinese Room. The text shows that Lem's key argument is to point out the computational naivety of the machines that feature in these experiments. The text concludes by presenting some of Lem's views on the nature of machine consciousness.

Keywords: Lem, Searle, Chinese Room, artificial intelligence, thought experiment

Stanisław Lem always felt comfortable among thinking machines and aliens. Sentient beings of various origins were the heroes, props or themes of all his novels and short stories, but he also addressed serious questions about the very nature of thinking in several non-fiction texts, especially in *Summa technologiae* (1964, further referred to as Summa), and in *Mystery of the Chinese Room* (1996, further referred to as Mystery).

Both of the mentioned texts contain passages devoted to the famous Turing Test (TT) and the implications of adopting TT as a mental criterion. Lem pays particular attention to *the formal critique of TT*, which is a set of considerations centered around the following argument:

1. One can succeed in TT by performing purely formal operations.
2. Purely formal operations can be performed without thinking.
3. Therefore, success in TT is not a sufficient condition for thinking.

Probably the most notorious case of formal critique of TT is the Chinese Room, thought experiment put forward by John Searle featuring formally operating system that perfectly mimics the language behavior of a native speaker (Searle 1980). Lem encountered the Chinese Room in Hofstadter and Dennett's anthology *The Mind's I*[1] and formulated an attitude toward that experiment that made him, in his own words, unpleasantly "towering over the crowd of sages" (M 89).

[1] This prestigious anthology also included three texts by Lem.

I present Lem's rejection of the Chinese Room. For Stanisław Lem was not only a direct inspiration to eminent philosophers of mind, not only a visionary anticipating later stalwarts of philosophical debate, but Lem's own arguments against the Chinese Room, the example so often taken up, were original and solid. After outlining Lem's relationship to TT, I offer an exposition of two of Lem's thought experiments, the Gramophone and the Jigsaw, whose critique is directly related to that of the Chinese Room – and show that Lem's key argument points out the constructional naivety of the machines featured in these experiments. I conclude by presenting some of Lem's views on the nature of machine consciousness.

Lem and the Turing Test

Lem reproduces TT in Summa very briefly.[2] He stresses that TT is a test of consciousness, not intelligence, and that it is a test with significant moral implications. The introduction of TT itself is brief, but concise: it is a conversational test of the distinguishability of machine and human behavior.

Lem adopts a metaphysical interpretation of TT, inferring the presence of conscious states from the indistinguishability of machine and human behavior: the machine behaves in the same manner as the human, therefore the machine has the same experiences as the human. The indistinguishability of behavior is a sufficient condition for ascribing consciousness to the machine since the machine truly *has consciousness*. The epistemic interpretation, on the other hand, is consistent with the intention with which Alan Turing put forward his test, namely, to find the ultimate publicly accessible criterion for the attribution of the mind. Indistinguishability of behavior is, on this epistemic interpretation, a sufficient condition for ascribing consciousness to a machine, since *no better condition can be found*. For Lem, however, TT is simply a test of conscious states: "If we cannot distinguish a machine from a man, we must admit that this machine behaves like a man or has consciousness" (S 113). Thus, Lem operates with a metaphysical interpretation in which the conversational faculty is a dependent and infallible manifestation of consciousness.

In Summa, Lem assumes no temporal, topical, or linguistic limits to the conversational test. He does not propose a satisfactory practical check of mental properties; he is concerned with affirming the relationship between conversational ability and consciousness. Lem's goal is a principled defense of TT as a sufficient mental criterion. That is why he focuses, in Summa, exclusively on the circumstances under which a machine succeeds in unrestricted TT. In Mystery, however, Lem's aim is different: he wants to deal with a specific critique of the sufficiency of TT. He thus alternates, somewhat messily, between considerations of three varieties of conversational tests: unrestricted, query-restricted, and topic-restricted.

For Lem, unrestricted TT is an acceptable indicator of consciousness. Success in unrestricted TT demonstrates the presence of pure mental states precisely in the means of the metaphysical interpretation of TT. Lem devotes a fair amount of attention to showing that a failure in unrestricted TT is not an acceptable indicator of the absence of consciousness. In Mystery (93), he gives an example of a cook who cannot answer a question about engines: "She says nothing because she has no idea, which has nothing to do with 'understanding' or 'not understanding'

[2] In the Czech translation, he refers to Turing's text as "Can a machine think?" (S 113). Lem became acquainted with TT in a Russian translation published in 1960 under this very title.

grammar, idiomaticity, or linguistic composition." Lem thus explicitly subscribes to the standard conception of TT as a sufficient but not necessary condition for the attribution of mental states.

Lem mentions the restricted conversational test in two passages in Mystery (92–93). First, he considers a test that would exclude certain questions beforehand, second, he reflects on a test in which the questions might relate exclusively to a short nursery rhyme. Of such attempts, he writes that there is "no analogy with the Turing test", explicitly rejecting a deeper connection between unrestricted TT and (query- and/or topic-) restricted conversational test. For Lem, the term 'restricted TT' would be a borderline case of contradiction in terms. TT is, by definition, unrestricted, and no form of restricted conversational test should have any significance for mental states attribution.

Success in restricted TT is inconsequential for Lem, since it can be achieved without the presence of conscious mental states. Equally inconsequential is failure. According to Lem, an unsuccessful conversational attempt is interpreted as "a mistake rather than a complete misunderstanding" (M 92). Lem is suggesting that we interpret the result of a restricted conversation based on our familiarity with the nature of the test itself. We opted for an informal behavioral procedure; therefore, we understand a conversational failure prima facie as a partial communicative failure of two thinking actors.

We can summarize that Lem attributes a heuristic value solely to success in unrestricted TT. Success in the restricted conversational test may be the result of purely formal procedures, whereas failure in any version of the test can be interpreted as ignorance or incompetence, not as an absence of the mind. As we will see later, Lem considers the Chinese Room to be an example of a restricted conversational test, which is why he repeats so many times the idea that such an example proves nothing and implies absolutely nothing.

The Gramophone

In Summa, Lem describes two kinds of machines that could attempt to pass an unrestricted TT, i.e., he considers two computational strategies to perfectly mimic human verbal behavior. The first strategy aims at creating an artificial duplicate of the human brain. Lem does not deal with this approach in detail; he sees this way of duplicating mental states as a metaphysically possible but constructively uninteresting solution that does not bring a deeper understanding of the nature of the mind. The second strategy is very different, it leads to the creation of a much more primitive machine:

> It is a Gramophone enlarged to the size of a planet (or the universe). It has very many, e.g., a hundred trillion recorded answers to all kinds of questions. When we ask a question, the machine does not 'understand' it, and only the form of the question, i.e., the sequence of the sound of our voice, sets in motion a relay that spins a record or tape with the recorded answer. (S 113)

Lem thus describes a machine akin to the puppets, the "magpies and parrots" that inspired Descartes to speculate on the detection of the mind (Descartes 1992, 42). The underlying system is a mechanically non-trivial but computationally simple combination of inputs and outputs – a purely formal system, since the procedure of assigning answers to questions does not consider the linguistic meaning of the terms used. However, such a formal system would in principle exhibit the external features of the mind and could converse quite well.

Later in Mystery, Lem comes up with another, less spectacular example. He considers a paper sheet with English inscription, turned face down and cut into pieces with unique shapes. Even if these pieces are shuffled, sooner or later the experimenter can reassemble the original sheet – solely due to the unique shapes of the pieces pinpointing their relative locations. On the reverse side of the reassembled sheet, there will of course be the original inscription in English (cf. M 90–91). Although the example with the jigsaw puzzle is not a direct variation of the Gramophone, it touches on the same deeper problem: the relationship between formal procedure and semantics. In both examples, the idea is to present a computationally simple manipulation with signs that leads to a linguistic behavior indistinguishable from that of an understanding human.

How do these examples relate to the Chinese Room? The Gramophone *is* the Chinese Room, only sixteen years earlier. Recall that the Chinese Room is about the ordering of formally identified characters into acceptable conversational output. The Gramophone is about the same thing, but with some variations: it is much larger than the room, operates with recorded sentences, and is not operated by a live actor. However, the size of the machine, the nature of the representation and the nature of the information processor do not affect the main principle of the experiments; both involve purely formal systems capable of passing unrestricted TT without the presence of real understanding.[3]

The Jigsaw is somewhat more modest in its original intent. It is not primarily a modification of the entire Chinese Room, but a simplification of its central concept. Lem tries to better approximate the concept of purely formal operation, which in the Chinese Room takes the form of manipulating symbols without available linguistic meanings. The part of the Chinese Room in which the actor combines representations based on their shape, the Jigsaw turns into combining pieces based on their shape, without the actor even knowing that these pieces have any representational function at all.

Lem writes that "originally Searle confused those to whom he was describing his experiment, for it was 'Chinese' etc." (M 94). Lem's modification is therefore intended to rid formal critique of TT of potentially misleading details. Lem seeks to show that the problematic step from formally (syntactically) identified representations to semantically interpreted output is entirely independent of whether the actor understands the output. He points out that performing a purely formal operation with symbols does not contradict the knowledge of meanings of those symbols, or else that the actor's understanding or non-understanding of the outputs plays no role at all in the experiment.

The Gramophone is a full-blooded precursor to the Chinese Room. The Jigsaw is presented by Lem much later as a simplification or purification of Searle's experiment. Both examples, however, are an original way of describing (and subsequently rejecting) the core of the formal critique of TT.

Conversational explosion

Although the Gramophone is more than thirty years apart from the Jigsaw, Lem criticizes these examples from the same position: both are naive by design.

[3] Moreover, Lem's Gramophone is completely identical to the hypothetical machine later known as the Blockhead, introduced in (Block 1981).

According to Lem, the Gramophone cannot in fact pass TT. Conversational outputs are generated by 'brute force', with the Gramophone assigning specific, pre-formulated answers to specific questions. The design of such a machine would have to account in advance for all possible questions (e.g., Why is there no processed cheese on the zero meridian?); it would have to include various variations of the same question (e.g., Why is there no cheese on the Greenwich meridian that can be spread on bread?); it would have to include prepared answers about the actual answers (e.g., I'm sorry I don't know the relationship between cheese and latitude.); and so on. The computational architecture of the Gramophone is simply not rich enough for success in TT.

Even though the similarity between Lem's example and Searle's experiment is evident, the above objection to the Gramophone cannot simply be transferred to the Chinese Room. For we know that the machines differ in the granularity of formally identified signs: whereas the Gramophone operates with the signs of sentences, the Chinese Room operates with the signs of words. Where the Gramophone needs a huge bank of pre-prepared phrases, the Chinese Room makes do with a limited bank of lexemes and ways of composing them into replicas. The computational architecture of the Chinese Room thus places less demand on the number of primitive signs.[4] The Chinese Room is in this sense richer and is thus much better equipped to handle TT. Lem's substantial criticism of the Chinese Room must be sought elsewhere, in a Cartesian inspiration.

Descartes assumed that there are tasks mechanically unsolvable: tasks testing general intellectual flexibility and tasks testing the ability to communicate (Descartes 1992, 41). Lem argues something very similar. According to Lem, the Gramophone could not actually pass TT, since the purely formal pairing of questions with pre-prepared answers is not sufficiently conversationally flexible. Lem supposes that Gramophone could be defeated in TT with a sufficiently developed strategy (S 114). He describes a sequence of steps where we first test the opponent's ability to understand the joke, then the ability to recall the previously told joke, and finally the ability to recapitulate the previously told joke. The Gramophone necessarily unmasks itself as a machine during one of these steps. This time, the same objection can be raised against the Chinese Room: the computational architecture of the machine assigning Chinese symbols is not flexible enough for the machine to successfully pretend to understand in an unrestricted conversational test.

Lem's argument against the formal critique of TT here is reminiscent of Dennett's later argument against radical skepticism. Dennett considers the famous thought experiment of the Brain in a Vat and points out the computational demands of creating a perfect fictional world: "[T]here are too many possibilities to store. In short, our evil scientists will be swamped by combinatorial explosion as soon as they give you any genuine exploratory powers in this imaginary world" (Dennett 1991, 5). Lem anticipates the same implications for the conversation test much earlier. Genuine freedom in conversation with the machine allows for the testing of such intellectual abilities, the strictly formal processing of which soon leads to a combinatorial explosion of possible answers, i.e., a conversational explosion.

[4] In addition, such a system of representations is both productive and systematic, see (Fodor 1975).

This is how Lem indirectly addresses the above objection to the Chinese Room in Mystery:

> After all, it is possible that among the questions will be ones that 'reveal a fundamental misunderstanding' of all the texts of the Chinese Room, which I see as permissible because John Searle and his respondents thought it eccentric (in my eyes) that the so-called strong artificial intelligence, or the hypothesis that a machine would pass the Turing test, would be dealt with negatively as the 'mystery of the Chinese Room' – the machine that represents the room understands nothing, and yet answers the questions as if it did. (M 92)

The passage deserves clarification. First, it is noticeable that Lem misidentifies the target of his criticism. For in his seminal text, Searle does not define strong artificial intelligence as a level of machine sophistication, but as a research program. According to Searle, strong artificial intelligence is based on the thesis that success in TT implies the presence of real understanding, i.e., strong artificial intelligence builds on a metaphysical interpretation of TT (Searle 1980). Thus, it is not a claim about the ability of a machine to pass TT, but a claim about the consequences that would follow from such an ability.[5] Strictly speaking, Lem does not deal negatively with strong artificial intelligence, but only with the assumption that machines of a certain type are in principle capable of passing TT.

More important, however, is the first section, where Lem sets out to counter traditional objections to the Chinese Room. He takes issue with the uncritical acceptance of the supposition that the Chinese Room can pass TT. As in the case of the Gramophone, he suggests the existence of a strategy that exposes a machine with such a naive construction. Once we explore non-trivial intellectual abilities, such as interpretation of figurative meaning or summarization, the Chinese Room's architecture proves inadequate.

Lem's critique of The Chinese Room can be summarized as a trio of related arguments. The first argument is based on the limitations of the architecture of the Chinese Room:

Argument from conversational explosion

1. Passing TT presupposes mastery of the conversational explosion.
2. The Chinese Room cannot master the conversational explosion.
3. Therefore, the Chinese Room cannot pass TT.

The conclusion of this argument enters as a premise into two separate inferences. In one of them, Lem reaches the same result as Searle, namely that it is possible, even likely, that the Chinese Room will not understand the conversation.

Argument from conversational failure

1. The Chinese Room cannot pass TT.
2. A system that cannot pass TT may not understand.
3. Therefore, the Chinese Room may not understand.

[5] Alternatively, 'strong intelligence' for a time did indeed denote the degree of advancement of a machine, specifically the fact that such a machine has consciousness and truly understands. However, as far as I know, 'strong intelligence' has never referred to the actual ability of a machine to pass TT.

Conversational failure also means that Searle's experiment fails to demonstrate what its author claims it should demonstrate, namely the insufficiency of TT and the hopelessness of a strong artificial intelligence program.

Argument to the irrelevance of the Chinese Room

1. The Chinese Room cannot pass TT.
2. Strong artificial intelligence says nothing about systems not passing TT.
3. Therefore, the Chinese Room is an irrelevant rebuttal of strong artificial intelligence.

This is a novel way to challenge the Chinese Room. For Lem is not rejecting a partial inference step within reasoning over a hypothetical situation, he is rejecting the acceptability of the hypothetical situation itself. Staying with the most familiar objections to the Chinese Room, the system and the robotic replies, we see that both accept the initial premise that the machine will pass TT. The replies then diverge in what they require for understanding beyond the computational architecture described by Searle: a proper (systemic) perspective and/or a proper (robotic) implementation. Lem, however, rejects the significance of the thought experiment from the outset. Not because he is a priori suspicious of the epistemic value of fictional scenarios, but because the Chinese Room, like the Gramophone, is built on inconsistent assumptions arising from its constructional naivety.

Why does Lem consider the Jigsaw to be naive? He thinks it is a poorly designed test. While the Jigsaw does not contain the same technical hurdle as the Gramophone, it fundamentally misses its research goal. Recall that the Jigsaw is a simplification of the formal critique of TT, where pure syntactic operations lead to meaningful outputs. According to Lem, it is naive to expect that relevant mental states should emerge in the actor in such a scenario. The Jigsaw thus heads in the same direction as the system reply to the Chinese Room. According to the system reply, Searle's scenario incorrectly expects that reasoning should occur at the information processor; the Jigsaw is then meant to illustrate the pointlessness of such an expectation, since the perspective of the formally operating actor is not relevant to understanding the hidden text on the flipside of the puzzle. Unlike the proponents of the system reply, however, Lem believes that in his Jigsaw and in the Chinese Room, the relevant point of view simply does not exist: "'Understanding' is not involved in the tests at all, and thus we cannot speak of a present 'consciousness'" (M 94). Both the Jigsaw and the Chinese Room assume machines participating in restricted conversational tests, and the design of these machines allows them to handle only a limited class of tasks. It would be a mistake, Lem thinks, to expect consciousness to appear somewhere in such machines.

Lem and machine consciousness

What machines can succeed in TT? In Summa, Lem divides thinking machines into artificial brains and 'ordinary' machines (S 113–144, quotation marks original). He then describes an artificial brain as a machine that is "as complex as a human brain", inside which we find "a vast number of circuits connected in the way that neurons are connected in the brain, then its memory blocks, etc." It is impossible to say for sure whether Lem understands such an artificial brain in the same way that the brain simulator reply to the Chinese Room understands it, i.e., as a realization of the sub-symbolic computational architecture of the human brain in an artificial substrate. The connection of circuits corresponding to neuronal synapses would suggest such an interpretation, but Lem's reference to memory blocks again points rather to a classical computational architecture. In any case, Lem understands such an artificial brain to be

comparable to the human brain in every relevant respect. Its computational architecture is not naive and allows the machine to pass TT, therefore an artificial duplicate of human brain should have conscious states.

'Ordinary' thinking machine is not a copy of the human brain. It is a machine capable of handling the conversational explosion based on classical architecture and symbolic programming. Lem describes the creation of such a machine not as duplicating neuronal functions and connections into inorganic material, but as the gradual refinement of a computer program. Lem envisions a computationally naive machine (the Gramophone) that is exposed to various conversational strategies in an imitation game. After each failure, the machine in question is augmented with the ability to cope with the task at hand. The machine thus learns to master more and more conversational tasks, defeating more and more unmasking strategies. Some conversational tasks have already been discussed in the previous chapter:

A. to respond with laughter to a joke
B. to recall a joke
C. to recapitulate a joke

However, there are other tasks to be found in Lem's texts:

D. to deduce
E. to induce
F. to compare
G. to "capture the 'essence' of differently formulated identical contents" (all in S 114)
H. to paraphrase a story (M 13)
I. to exhibit the instinct for self-preservation (M 14)
J. to consider semantic polymorphism
K. to infer probabilistically (both in M 85)
L. to provide discursively appropriate responses (M 91)
M. to be linguistically performative (M 163)

Lem thinks that there is a finite number of these tasks and that by improving the conversational abilities of the machine we can reach a point where the machine succeeds in TT. Such a machine will then have mental states in accord with the metaphysical interpretation of TT. Let us reiterate the crucial importance of these conversational tasks in the critique of the Chinese Room and the Gramophone: these machines are not capable of solving the tasks, so they cannot succeed in TT, and are thus not relevant to considerations of machine thinking. The computational architecture of these machines is not flexible, rich, responsive, and complex enough.

Lem bases his belief in the possibility of 'ordinary' thinking machines on the thesis that a machine with a sufficiently flexible, rich, responsive, and complex architecture can be constructed by classical methods. However, this thesis itself is based on the implicit assumption that all the above conversational tasks are solvable by purely formal means. At the same time, however, it seems that a machine, that is supposed to handle all the above conversational tasks, must have semantics. Namely, tasks C, G, and J, which require content recognition and identification of the dispensable parts of utterances, are explicitly described as tasks related to semantic objects, i.e., to the essence and content of the message. Can the requirement for formality of operations on signs be satisfied together with the requirement for access to the meanings of those signs?

From the examples in Summa, we can infer that, according to Lem, somewhere in the process of designing the Gramophone, semantics emerges. Conversational tasks that require the manipulation of the content of the message are then understood as formal operations on established semantic objects. Lem was not alone in this position; the same thesis would later be formulated by John Haugeland as "the formalists' motto: If you take care of the syntax, the semantics will take care of itself" (Haugeland 1985, 106). Unfortunately, Lem is very opaque on this point. In fact, he simply states the emergence of semantics in the machine: "The designer, offended in his pride, perfects the machine and builds in such a memory that it can recapitulate what has been said' and 'finally, after a long series of refinements, he places in the machine [...] the ability to capture the 'essences' of differently formulated identical contents".[6] Thus, the key problem of strong artificial intelligence is not satisfactorily solved here; Lem's birth of semantics in the machine just takes place with the wave of a designer's magic wand.

The idea of a gradually improving thinking machine leads Lem to another scenario, which again foreshadowed a deeper interest of the scientific community. He contemplates machines that, in the process of refinement, find themselves between borderline models, i.e., between the initial Gramophone with 'zero consciousness' and the final thinking machine with 'full consciousness' (S 116). When does consciousness appear in such a sequence? Although Lem does not answer this question, he uses the given idea to support a thesis about the incremental nature of consciousness. According to Lem, the machines in this sequence differ in the degree of consciousness: "The disconnection of the individual elements ('neurons') of the machine causes only slight quantitative changes ('fading') in consciousness, just as a progressive disease process or a surgeon's knife does in a living brain" (S 116). Consciousness has degrees and the thinking system may gradually lose the ability to experience. Lem thus anticipates philosophical considerations of the alteration of phenomenal states when neurons are replaced by artificial duplicates.[7]

I should emphasize that Lem's critique of the Chinese Room is independent of his conception of consciousness. Even if it turned out that consciousness emerges in leaps and bounds in the machine, the Chinese Room would remain an irrelevant constructionally naive thought experiment. It is also noteworthy that Lem's critique is independent even of the magical step of semantic emergence. For Lem agrees with Searle's conclusion that understanding does not emerges in the Chinese Room, and he bases his critique not on a competing philosophy of mind but on the computational deficiency of the hypothetical machine.

Conclusion

Stanisław Lem disagreed with the Chinese Room even before it was possible. In Summa, he presented a machine that corresponded in all essential parameters to the later means of formal critique of TT, and then rejected that machine as computationally naive. Thirty years later, in Mystery, he set himself squarely against Searle's scenario and offered a stark, clean variation of it. He showed that the Chinese Room suffered from methodological as well as structural flaws.

Lem's contribution to the debate on the nature of artificial intelligence cannot be overstated. Of the ideas presented above, the following deserve wider attention:

[6] Both examples in (S 114). In (M 14), Lem describes at some length the way in which the task of paraphrasing a message can be handled formally. However, paraphrasing is a fundamentally different type of task from summarizing (recapitulating) in terms of semantic requirements.
[7] E.g. (Searle 1992, 66) or (Chalmers 1995).

The rejection of naive ideas about TT. A fundamental error of formal critique is the failure to appreciate the strategic possibilities of the judge in TT. The Chinese Room and similar thought experiments emphasize those aspects of conversational exchange that involve easily algorithmizable procedures, such as the production of well-formed sentences.

The danger of conversational explosion. Advanced testing strategies within TT require capabilities on the machine side that cannot be implemented in the Chinese Room. Success in TT requires a computational design that allows operating on the semantics of symbols.

The constructive dilemma of the Chinese Room. Either the Chinese Room can succeed in TT, then the room must understand, and the strong artificial intelligence program is not compromised – or the Chinese Room cannot succeed in TT, then it is not a relevant counterexample, and the strong artificial intelligence program is not compromised.

Syntax is sufficient for semantics. A purely formal operation at the lowest computational level of the machine does not preclude the emergence of semantic objects and reference to those objects at a higher computational level of the machine.

Exposing the Chinese Room. If we purge Searle's experiment of misleading details, it turns out that he expects consciousness in the wrong place. There is no right place in the Chinese Room.

At several points in Mystery, his late work, Lem reveals a significant shift in views. He was no longer convinced that the duplication of linguistic behavior alone was sufficient to attribute a mind.[8] This does not mean, however, that he discounted his criticism of the Chinese Room. Perhaps his own goals intersected with the ultimate goals of formal critique towards the end of Lem's life, and they were united by doubts about the sufficiency of TT, but Searle's thought experiment remained irrelevant, naive, and misguided.

References

S – Summa technologiae
M – Mystery of the Chinese Room

Bibliography

Block, N. (1981): Psychologism and Behaviorism. *The Philosophical Review*, 90(1): 5–43.

Dennett, D. C. (1991): *Consciousness Explained*. Little, Brown and Co.

Descartes, R. (1992): *Rozprava o metodě*. Nakladatelství Svoboda.

Fodor, J. A. (1975): *The Language of Thought*. Harvard University Press.

Haugeland, J. (1985): *Artificial Intelligence: The Very Idea*. MIT Press.

Hofstadter, D., Dennett, D. (1981): The Mind's I. Bentam Books.

[8] See (M 17) and (M 138).

Chalmers, D. (1995): Absent Qualia, Fading Qualia, Dancing Qualia. In: Metzinger, T. (ed.): *Conscious Experience*. Imprint Academic.

Lem, S. (1995): Summa technologiae. Magnet-Press.

Lem, S. (1999): Tajemství čínského pokoje. Mladá fronta.

Searle, J. R. (1980): Minds, Brains and Programs. *Behavioral and Brain Sciences*, 3(3): 417–457.

Searle, J. (1992): The Rediscovery of the Mind. MIT Press.

Turing. A. (1950): Computing Machinery and Intelligence. *Mind, New Series*, 59(236): 433–460.