Christian Kaernbach

No Virtual Mind In the Chinese Room

Abstract: *The Chinese room thought experiment of John Searle militates against strong artificial intelligence, illustrating his claim that syntactical knowledge by itself is neither constitutive nor sufficient for semantic understanding as found in human minds. This thought experiment was put to a behavioural test, concerning the syntax of a finite algebraic field. Input, rules and output were presented with letters instead of numbers. The set of rules was first presented as a table but finally internalized by the participants. Quite in line with Searle's argument, uninformed participants mastered the syntax but did not explicitly report semantic knowledge. In order to test the virtual mind reply to the Chinese room argument, the reaction time pattern of the participants was compared to that of an informed control group. The correlation was quite high but could be traced back to memory load and response priming, i.e. to syntactical factors. No trace of tacit semantic knowledge of the task could be found in the experimental group.*

Introduction

Computer scientists have amazed the public for decades with the vision of computers that think, computers that are self-aware and argue with us about their consciousness (Turing, 1950; Epstein, 1992). Would we have the right to turn them off? Could they be held responsible for their acts? Should they have the right to vote? For the last quarter of a century John Searle (1980) has thrown serious doubts on these claims. He argued that a computer would never understand the meaning of the symbols it is dealing with. While it might be able to manipulate them appropriately, i.e. while it might master their syntax, it would never know their semantics.

In his famous Chinese room thought experiment Searle imagined a monolingual English speaker, locked in a room, together with a large set of rules comprising lists of Chinese symbols. While not giving the meaning of the Chinese symbols these rules would enable the person to treat incoming Chinese symbols

Correspondence: Christian Kaernbach, Institut für Psychologie, Karl-Franzens-Universität Graz, Schubertstr. 51a, 8010 Graz, Austria. www.*kaernbach.de*

Journal of Consciousness Studies, **12**, No. 11, 2005, pp. 31–42

in such a way as to sort out some other Chinese symbols as a kind of response. A native Chinese speaker outside the room might judge the output to be an appropriate and meaningful response to the input. The person inside the room, however, would not understand the meaning of the symbols. This person would correspond to the CPU of a computer dealing seemingly intelligently with input and output symbols.

Searle considers in his article several replies offered when he had the occasion to present his thought experiment to a number of workers in that field. The best known is the system reply, suggesting that the person locked in the room is part of a whole system, and the system understands the meaning of the symbols it is handling. Searle rejoins that the person might internalize the rules and the look-up tables and do all the necessary operations internally. The system would then be in the person, and if the person would not understand Chinese, neither would the system because it would be part of the person.

The virtual mind (or multiple personality) extension of the system reply (see e.g. Weiss, 1990, Donald Perlis in Hayes *et al*., 1992, and MacLennan, 2001) maintains that if the person internalizes everything necessary to do the operations internally the person would actually understand Chinese with a separate part of the personality that does not communicate with the English speaking part of the personality. Let us for sake of clarity call these two parts of the personality Dr. Jekyll and Mr. Hyde.[1] Mr. Hyde would handle the symbols, and understand their meaning, not being able to convey this to Dr. Jekyll who is reporting to us. If internalizing the rules would indeed induce a kind of multiple-personality state we could not rely on the explicit report by Dr. Jekyll stating no knowledge of Chinese. Tacit knowledge, however, whether syntactical or semantic, should show up in reaction time patterns of the person performing along these rules, as it is the Chinese speaking part of the person, Mr. Hyde, that is responsible for the observable behaviour.

In order to test the virtual mind / multiple personality extension of the system reply to the Chinese room thought experiment, the present study implemented a behavioural test of this thought experiment, looking for signs of explicit as well as of tacit knowledge. The next section discusses the scenario that will be simulated. Then the selected scenario is introduced in detail, for the experimental group as well as for an informed control group. The next two sections present the results of this experiment. The final section discusses the contribution of this experiment to the debate.

The Scenario

Searle chose the Chinese chatterbot as a parody of Turing's (1950) vision of a computer that could do small talk. Descartes (1637) was right to suppose that small talk — as simple as it seems to us who do it every day — represents a hurdle that is very difficult to surmount by machines. The Loebner prize, a competition awarding 100,000 US$ to the first program passing their version of the Turing test, has up to now — more than fifty years after Turing's vision — not

[1] The analogy is not perfectly correct, as with Stevenson these two know of each other.

inspired any results that could fool the observers. This represents a major obstacle to implementing the original Chinese room scenario as a real experiment: There is no acknowledged set of rules that would implement a chatterbot. On the other hand, arithmetic — for humans much more of a proof of intelligence than small talk — has been successfully implemented in mechanical computing devices as early as in the middle of the seventeenth century.

Chatterbots and the Turing test are no longer the touchstone of research on intelligent systems. Neil Bishop, the organizer of the 2002 Loebner competition, defines two classes of research in this field (cited from Sundman, 2003): 'Decision Sciences (DS) and the human mimicry side called Mimetics Sciences (MS).' Only in mimetic sciences (the smaller branch of the two) would the capability of small talk represent an important advance. However, intelligent systems that don't try to mimic human behaviour bear the same problem of possibly not understanding the semantic meaning of the symbols they are dealing with. The argument of Searle holds also for DS systems. Puccetti (1980) has suggested transfering Searle's thought experiment to a 'chess room' and concludes that a chess computer does not understand the game it is playing. The same question arises for a pocket calculator. Does it understand what it is doing? *Does it calculate?* Or is it just manipulating symbols that don't mean anything to it? Is its name a misnomer? In analogy to the Chinese room thought experiment, a person sitting in an 'algebraic room' might not understand the meaning of the symbols being manipulated, even if to the outside observer the response behaviour would perfectly make sense. In contrast to the original chatterbot scenario, the set of rules to make a calculator operate is agreed upon. The thought experiment, transferred from the Chinese room to the algebraic room, can therefore realistically be considered for a behavioural test.

Methods

In order to get a manageable set of rules, the arithmetic was constrained to calculations in a finite algebraic field. This field is also known as Z modulo N. Consider first the special case that N equals 10. This corresponds to the output of the final digit of a common calculator. For some calculations, the result corresponds to that of normal arithmetic (e.g., 3+1=4 and 3×1=3), but 'overflow' calculations that reach or exceed N are reduced to their modulo N value, ignoring the overflow (e.g., 3+8=1 and 3×8=4).

If exploiting commutativity, A(op)B = B(op)A, the list of rules comes down to N×(N+1)/2 results for addition, and the same number of results for multiplication. For the experiment, N was set to 5, giving a total of 30 rules (see Table 1a). Rules and task were presented in an obscured manner, using letters and brackets instead of digits and operators (see Table 1b). A person can easily internalize these rules. Would that person develop overt understanding of doing arithmetic? Would a part of the person develop that understanding? Would this understanding show up in the reaction time pattern?

The experimental group (EG) comprised 12 participants. The task of the participants was to answer to a stimulus with a key press. The stimulus consisted of

(a)

A+B	**0**	**1**	**2**	**3**	**4**
0	0	1	2	3	4
1		2	3	4	0
2			4	0	1
3				1	2
4					3

AxB	**0**	**1**	**2**	**3**	**4**
0	0	0	0	0	0
1		1	2	3	4
2			4	1	3
3				4	2
4					1

(b)

XY	**K**	**L**	**J**	**D**	**F**
K	K	L	J	D	F
L		J	D	F	K
J			F	K	L
D				L	J
F					D

[XY]	**K**	**L**	**J**	**D**	**F**
K	K	K	K	K	K
L		L	J	D	F
J			F	L	D
D				F	J
F					L

(c)

```
JL→D[J]
FJ→L[D]
FL→K[F]
DK→D[K]
KL→L[K]
KK→K[K]
DF→J[J]
JJ→F[F]
LD→F[D]
DD→L[F]
FF→D[L]
KF→F[K]
JK→J[K]
LL→J[L]
JD→K[L]
```

Table 1. Addition and multiplication matrices of a finite algebraic field of size 5 (Z modulo 5)

(a) Left column. Normal representation with digits. Only the upper right-hand half of the matrices needs to be filled due to commutativity. (b) Obscured representation of the same algebraic field with letters. (c) Format of table as presented to participants (see Appendix A for explanation).

two letters taken from a set of five possible letters, sometimes with, sometimes without brackets. Depending on the letters and on the presence of the brackets, only one of five possible response keys was correct.

EG participants were not informed on the translation of letters to digits. Instead, they solved the task by looking up a table containing 15 cryptic lines such as ‘JL→D[J]’ appearing on the left side of the screen (see Table 1c and Appendix A). To these participants, the assignment of stimuli to responses could appear to be completely arbitrary. They were told that the experiment was about the dynamics of learning. As explained above, however, the table of rules represented algebraic tasks, translated to letters.

EG participants performed 60 blocks of 50 trials each in sessions of 5 or 10 blocks during a period of at maximum 10 days. The table of rules was mostly present on the computer screen during the initial blocks but less so later. During the final 10 blocks, it was not present. Participants had to have internalized the rules in order to successfully pass these last 10 blocks.

The control group (CG) comprised 12 participants of approximately the same age. CG participants started with 23 blocks of training to do calculations in Z modulo 5 with digit keys. After that, they underwent a second training (7 blocks) for a specific translation of letters to digits. After these two training sessions, CG participants performed 25 blocks of calculations. During these blocks they performed exactly the same task as EG participants, responding to pairs of letters with or without brackets by pressing one of five letter keys. The only difference was that they answered by doing calculus, while EG participants relied on an internalized table of rules that needed not to correspond to anything useful or meaningful. For methodological details see Appendix A.

Results I: Ask Dr. Jekyll

After each session of 10 blocks, as well as after the final block, EG participants were interviewed. They were asked to describe the task in their own words. In a total of 72 interviews, none of them used words like ‘calculate’, ‘arithmetic’, ‘add’, ‘multiply’ or the like.

During these interviews they were encouraged to mention any regularities that could underlie the table of rules. Many participants came up with mnemonic phrases for some of the rules. But did they notice that some letters (numbers) played a special role in the response scheme? Four participants were even in the final interview not aware of any regularity. The regularity most often reported by the others concerned the special role of the letter that stood for 0. This regularity was reported overtly by seven participants. An eighth participant reported that this letter ‘went faster’, without having noticed that there was a regularity in the response schema concerning this letter. This is a nice example of the transition of tacit to explicit knowledge, stuck midway at the end of the experiment for this participant. The regularity concerning the letter that stood for 1 was reported by two participants, as was the regularity $3+4=3\times4$. The regularity $2+2=2\times2$ was reported by three participants. Note, however, that these regularities constitute syntactical knowledge. They do not tell anything about the meaning of the symbols. Especially, the understanding of the number quality of the symbols is not needed in order to notice that some of them have a simpler response scheme than others.

The final interview, at the end of the experiment, comprised two more questions than the regular after-session interview. First participants were asked whether they thought that associativity holds for the operation with or without brackets. Participants could answer on a scale ranging from 1 (always) to 5 (never). The average across all participants was 2.75 for addition and 2.9 for multiplication, indicating that participants had no idea whether associativity would hold. Only one participant reported ‘always’. He did not report doing math during the regular interviews, and was not sure that there was a translation of letters to numbers (see next point).

The last question of the final interview openly addressed the possibility that the letters stood for numbers. This is a very sensitive test of explicit knowledge, and is usually not applied because such a direct question might cause explicit knowledge that was not explicit beforehand. Participants were told that there were two experimental groups, a group A where the letters stood for numbers and the table of rules was actually an addition and multiplication table, and a group B where there was no such possible translation. Note that in reality there was no such group B in the experimental design. EG participants (i.e. of group A) were asked whether they felt they had belonged to group A or group B. They could answer on a five-point scale, with 1 meaning 'sure group A' and 5 meaning 'sure group B', and 3 meaning 'no idea'. The first two participants were asked this question in parallel, and they started immediately to check whether their table of rules would confirm to the mathematical laws. They both were stunned that it did and reported 'sure group A'. For the next ten participants the experimenter took care that they did not test this hypothesis but simply rated their confidence to belong to group A or to group B. The average rating across these ten participants was 3.2, indicating that the participants had on average no idea. Only one of the last ten participants reported 'sure group A'. She was not the same that reported associativity always to hold. She did not report doing math during the regular interview.

In summary, the EG participants developed no overt understanding of doing math. Most of them detected some syntactical regularities in the table of rules but that did not induce any mathematical understanding of the symbols they were handling. This is a nice illustration of the Chinese room argument of Searle, where the operators master the syntax of the symbols they are manipulating but don't understand their meaning.

As the participants had internalized the table of rules, the system reply does not apply: the system was part of them. It could, however, be argued that the math-understanding 'system in the participants' had actually developed an understanding of the symbols but that this system (Mr. Hyde) was different from that part of the participants that could communicate with the experimenter during the interview (Dr. Jekyll). This virtual mind argument sounds like beyond proof. How could we hope to learn something about the state of mind of a part of the personality that is not able to communicate to us? The situation is, however, less hopeless than it seems. It is Mr. Hyde who punches the keys, so there is a way to learn something about his state of mind by analysing the reaction time patterns of EG participants and compare them to those of CG participants.

Results II: Check Mr. Hyde

Many skills in everyday life are acquired without being taught, and more often than not one is not able to explicitly state the knowledge underlying these skills. This knowledge is then often called 'tacit knowledge'. In order to study the acquirement of tacit knowledge in the laboratory, implicit learning paradigms have been designed. For instance, in artificial grammar paradigms (Reber, 1967), participants could after some training tell which construction was syntactically correct and which was not, without being able to explicitly state the rules of that grammar. The initial study on sequence learning by Nissen and Bullemer (1987) demonstrated tacit knowledge in Korsakoff patients by analysing their reaction time pattern. There is a contentious debate going on in psychology on methodology and interpretation of results in implicit learning and tacit knowledge experiments (see e.g. Berry, 1997; Stadler and Frensch, 1998). A major question is how to define when a process is non-explicit. Have the right questions been asked during the interview? Did the participants use a different type of explicit knowledge than that expected by the experimenter? The problem is to verify beyond doubt that the acquired knowledge has not become explicit.

As a matter of fact, explicit semantic knowledge can be ruled out on firm grounds in the present study.[2] For the following analysis, however, it is not relevant whether semantic knowledge did get explicit or not. While the technique is inspired by implicit learning paradigms, the goal is different: it is not to show that there is tacit knowledge in spite of no explicit reports. It is to reveal all kinds of knowledge, whatever the state of consciousness of that knowledge.

During the final ten blocks the reaction time (RT) was on average 1.4 s and the error rate was on average 3.6%. During these blocks RT varied as a function of the 30 tasks (cf. Table 1). Figure 1a shows the average RT for each task for the EG versus for the CG in a two-dimensional plot. It gets obvious from Figure 1a that there is a clear-cut correlation of the RT profile of the two groups ($r = 0.81$). This correlation is highly significant: the probability p for an alpha error (false positive) is very small ($p < 10^{-7}$), so the correlation can't be caused by random variations.[3] Would this correlation of the RT profiles signify that Mr. Hyde had an understanding of the different tasks similar to the understanding of the control group?

[2] Participants did not report overtly any meaningful understanding of the symbols they were dealing with. Moreover, even when confronted with the possibility of a meaningful interpretation of the symbols they were undecided. Direct questions like this one run the risk of inducing conscious knowledge. In the present study, this question and its negative result were possible because the semantics of the symbols was far fetched for the participants. In classical implicit learning paradigms the knowledge under question has to be of moderate difficulty: not too easy so as not to induce explicit knowledge, and not too difficult in order to get some learning during the experiment. This trade-off is the difficult part of implicit learning paradigms. In contrast to this, the design of the present study was such that explicit semantic knowledge was hard to acquire. This is mainly due to the fact that this knowledge was not necessary and would not even be helpful in order to solve the task fast and accurately. Common implicit learning paradigms deal with the gradual acquiring of (possibly tacit) syntactical knowledge. In the present study, the syntax was given explicitly. The participants had an absolutely sufficient explanation for their performance: They had internalized the syntax. There was no need for a semantic interpretation. The algebraic room is only a rough cartoon of the Chinese room, and it is obvious that the much more complex program needed to implement a decent Chinese chatterbot would present even more difficulties for the acquirement of explicit semantic knowledge. — Please note that even in the case that some of the participants had acquired explicit semantic knowledge the argument of Searle would still hold. Humans have unlike computers free capacity to think about what they are doing. As long as there is a single participant with perfect syntactical knowledge but without semantic understanding, this participant would constitute the analogy to the computer in Searle's metaphor. For the present study, however, it is quite advantageous that none of the participants showed explicit semantic knowledge, as this allows to average reaction times over homogenous groups of participants.

[3] A precondition for testing a correlation coefficient for significance is that the two variables are normally distributed. This was tested with a χ^2 test with 10 bins ([$-\infty$,-1.6], [-1.6, -1.2], ... [1.2, 1.6], [1.6,∞]). Both RT distributions were not significantly different from a normal distribution: $\chi^2_9(CG) = 9.8$ and $\chi^2_9(EG) = 8.8$, both values are well beyond $\chi^2_{9:0.05} = 16.9$. This comes a bit as a surprise, as there is no reason why the RT distributions should be normally distributed over task type. Similar χ^2 tests have been performed for the RT distributions after removal of the effect of the predictor variable X, both for the complete groups as well as for the subgroups (see correlation coefficients reported later in this section). Only in one case was the distribution highly significantly different from a normal distribution: the distribution of the RT for the complete EG after removal of the effect of X (see y-axis in Fig. 1b). In this case ($r = -.07$), however, there is no question of significance.

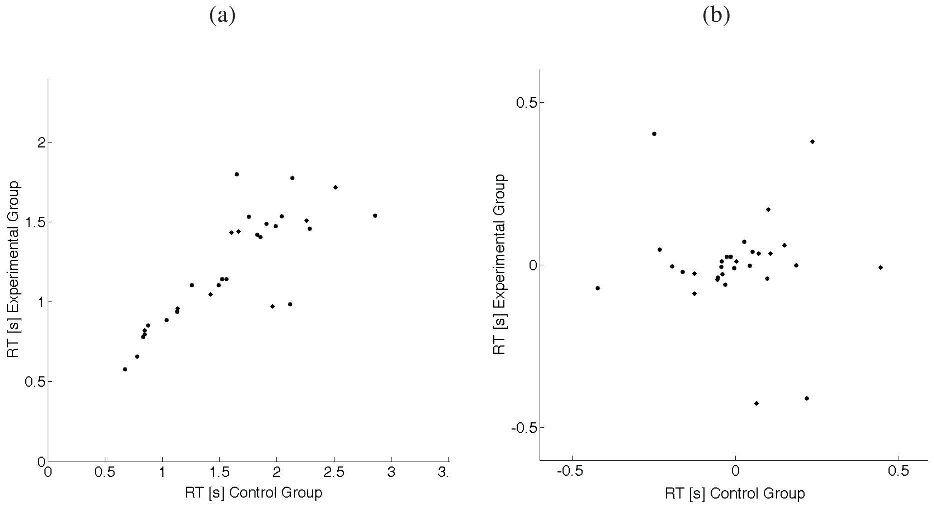

Figure 1. Comparison of RT profiles of experimental group and control group (a) before and (b) after removing the effect of memory load and response priming.

Two factors could be identified as the factors responsible for the RT variability in both groups: memory load and response priming (see Appendix A). RT variability was predicted with a combined variable X, expressing both memory load and response priming. After removing the effect of these factors, the partial correlation coefficient between groups is not significantly different from zero ($r = -0.07$). Figure 1 shows the RT data before and after removing the effect of the predictor variable X. So it seems that there are no traces of commonalties to be found in the RT patterns of the two groups beyond the effects of memory load and response priming. These effects are, however, not linked to a semantic understanding of the symbols being manipulated, but inherent to the syntax. The failure to find traces of common processing beyond syntactical factors is not due to a lack of test power: Subgroup comparisons (N=6/N=6) demonstrate that even after removal of these effects there is still enough variability in the RT profiles to indicate common processing within groups (partial correlations CG: $r = 0.44$, $p<0.0075$, EG: $r=0.70$, $p<10^{-5}$).

In summary, the RT profiles of the experimental group did not show any commonalties with those of the control group that could be attributed to hidden semantic knowledge. Such commonalties as there were could be attributed to effects of memory load and response priming — effects that are explicable without resorting to semantic concepts.

Discussion

Searle's Chinese room argument has been enlivening the debate on artificial intelligence for more than two decades now. While some propose to close the Chinese room (Weiss, 1990) — as if it ever had been open — many others contribute to the ongoing debate. Stevan Harnad even calls Searle's paper the 'most influential target article' of *Behavioral and Brain Sciences* (Harnad, 2001). This is not the place to review or even balance the contributions to this debate. The present article focuses on the system reply and its extension, the virtual mind hypothesis. Nevertheless, a general remark may be permitted: Many of the arguments put forth in this debate are quite speculative in nature, musing about what might or might not happen in the hypothetical situation conceived by Searle. However, the Chinese room situation is not beyond realization. It is true that real implementations of the Chinese room will in the foreseeable future most likely not cover small talk. They will treat smaller problem spaces instead. But the question of whether semantic understanding arises from (or is identical to) mastering the syntax remains the same.

The present study tested the hypothesis that a person impersonating the hypothetical occupant of the Chinese room and internalizing all the rules might harbour a virtual mind that is understanding Chinese. Decent chatterbots have not been programmed up to now, so there is no set of rules agreed upon. In order to perform a real experiment, the Chinese room scenario was transferred to a scenario with a well-defined and manageable set of rules, dealing with the numbers of a finite algebraic field. As expected, participants did not report explicitly doing arithmetic, even if presented with this possibility. The reaction time pattern showed a clear-cut correlation to the reaction time pattern of a control group doing arithmetic. But this correlation was due to factors such as memory load and response priming — factors that have nothing to do with the semantics of numbers. In other words, the analysis of the reaction time patterns did not reveal any trace of semantic knowledge in the experimental group.

What could have been expected? Was there any hypothetical possibility for semantic knowledge to show up in this paradigm? Reaction time patterns often reveal semantic — as opposed to syntactical — knowledge. In the present study it would have been conceivable that the reaction time profiles of the experimental and the control group correlate due to the ‘overflow’ feature of a task, a feature not obvious from the table but present in the mind of the control group due to the semantic understanding of the symbols. Another possible candidate could have been the priming of the ‘2×A’ task by a prior ‘A+A’ task (beyond simple response priming as in any other two tasks with the same result), given that multiplication is defined by multiple addition. The results have been cross-checked for any such effects. However, the only factors that played on both reaction time profiles could be attributed to memory load and response priming — features inherent in the table of rules, i.e. to syntactical features.

The failure to find traces of semantic knowledge does not disprove the virtual mind hypothesis. Two loopholes remain: On the one hand, one could claim that virtual minds are so very different from human minds that no traces of them are to be expected when comparing their behaviour to that of humans. This would, however, not only safeguard the hypothesis against this experiment but against tests of any kind: it would make the virtual-mind hypothesis a matter of faith. If, on the other hand, virtual minds are akin to human minds, so that behavioural commonalities are to be expected, it could well be that I looked in the wrong place. The present study serves as a proof of principle, opening the door to experimental tests of a fascinating thought experiment.

Acknowledgments

I am grateful to Uli Walther for his advice on finite algebraic fields, to Erdmute Sommerfeld for early discussions of the method, to Carsten Bogler for his help with the experiments, to Stevan Harnad and John Searle for helpful discussions of the results and implications, and to Stephen Grossberg, Bob Kentridge, Peter König, and William Yost for helpful comments on the manuscript. I am indebted to Sabine Windmann for asking the question 'Does a calculator calculate?' It has inspired this study.

Appendix A: Methodological Details

In total, 24 psychology students took part in the experiment. Both the experimental and the control group (EG and CG) comprised 10 female and 2 male students. All participants but one (EG) reported right-handedness.

The five possible stimulus letters as well as the five possible response keys were D, F, J, K, L. The positions of these letters on a standard computer keyboard correspond to the rest positions of the index and middle finger of the left hand, and of index to ring finger of the right hand.

For EG participants, there was initially a table of rules present on the computer screen. This table (see Table 1c) contained lines such as 'JL→D[J]'. This indicated that whenever the two stimulus letters J and L (in whatever order) appeared the answer would be D if the two letters were given without brackets, and J if they were given with brackets. These 15 lines coded for all possible combinations of letters, so that in total 50 stimulus situations could be answered correctly with the help of this table. The translation from letters to digits, the meaning of the presence or absence of brackets (addition versus multiplication), and the order of the 15 lines varied between participants. They were, however, constant between different sessions of the same participant so that the participant would always have exactly the same table of rules.

A trial started with an 'empty screen' (more precisely: empty stimulus area, the table of rules always stayed on screen) for 500 ms, the presentation of an exclamation mark in the center of the screen for another 500 ms, and another 'empty screen' for 1000 ms. Then came the presentation of the possibly bracketed two-letter stimulus. Participants were instructed to respond as fast as possible but to avoid errors. In case of a wrong response, the appropriate line of the table of rules was displayed above the stimulus, and the participant had to press the correct key. During the first 20 blocks, the table of rules was present during the first 40 trials of each block. During the next 30 blocks, the table of rules was present during the first 25 trials of each blocks. During the final 10 blocks, the table of rules was not present. A session of 10 blocks lasted about 50 minutes.

CG participants were initially presented with pairs of digits, either with or without brackets. They were explained which of these two variants meant addition and which meant multiplication. They had no table to help them but had to give the correct answer by doing arithmetic, pressing the correct digit key. In case of a wrong response the correct result was displayed above the stimulus, and the participant had to press the correct key.

Then they were trained in a specific translation of numbers to letters. For each participant of the experimental group there was one participant of the control group who used exactly the same translation of numbers to letters and the same 'bracket' syntax for addition versus multiplication. Participants of the control group were trained for both directions of translation. During this training, the 50-trial blocks were halved. During the first 25 trials of each block participants were shown a digit and had to respond with the correct letter, and during the last 25 trials of each block it was vice versa. A table of this translation (5 entries) was present on the screen during this task. As the instruction was to react as fast as possible, the translation table was soon ignored by the participants.

Finally CG participants performed the same task as EG participants, pressing a letter key in response to a letter stimulus. The trial structure was the same for both groups, with the exception that there was no table of rules present on the screen of CG participants, and that for CG the feedback in case of errors consisted in displaying the correct key instead of a line of the table of rules.

Appendix B: Memory Load and Response Priming

Variance in the RT data due to memory load M and response priming P was predicted with a combined variable X=M+P. For both groups, some of the stimulus pairs induce more, other less memory load. Tasks such as 0×0=0 will activate only one memory representation (M=1), including the representation of the result of the parallel task 0+0=0. Other tasks have a higher memory load. The highest memory load is M=4 for tasks such as 2+3=0, 2×3=1.

Response priming can play in favour or against the correct response. Letters present in the stimulus prime the corresponding response keys. Of the five possible stimulus letters, the one standing for zero (zero letter, ZL) plays a special role. It occurs more often as a response than any other key (14 times versus 9 times for the other keys). It was therefore assumed that ZL exerts more priming effect than the other letters; the exact amount of priming is of no importance to the following analysis. A given pair of stimulus letters primes one or two responses. On the other hand, this pair is associated with two possible responses. The set of the primed and of the associated responses might or might not be equal. If the two associated responses are unequal, and if the irrelevant one of them is more primed than the relevant one, the priming effect P is set to +0.3, indicating that the task is difficult because the primed response must be inhibited. If, on the other hand, the relevant response is primed and the irrelevant response is not, or to a lower degree, P is set to –0.3, reflecting a facilitation. The absolute value of the priming effect (0.3) was determined by optimizing the correlation between the RT data and predictor X. Table A.1 lists the values of M, P, and X for all 30 tasks. The variable X correlates significantly with the RT data for both groups (EG: r=0.84, CG: r=0.94).

T	CR	SR	M	P	X	T	CR	SR	M	P	X	T	CR	SR	M	P	X
0+0	$\underline{\underline{0}}$		1	–0.3	0.7	0+2	$\underline{2}$	$\underline{\underline{0}}$	2	+0.3	2.3	3×4	2		3		3
0×0	$\underline{\underline{0}}$		1	–0.3	0.7	0+3	$\underline{3}$	$\underline{\underline{0}}$	2	+0.3	2.3	4+4	3	1	3		3
0×1	$\underline{\underline{0}}$	$\underline{1}$	2	–0.3	1.7	0+4	$\underline{4}$	$\underline{\underline{0}}$	2	+0.3	2.3	4×4	1	3	3		3
0×2	$\underline{\underline{0}}$	$\underline{2}$	2	–0.3	1.7	1+1	2	$\underline{1}$	2	+0.3	2.3	1+2	3	$\underline{2}$	3	+0.3	3.3
0×3	$\underline{\underline{0}}$	$\underline{3}$	2	–0.3	1.7	1×2	$\underline{2}$	3	3	–0.3	2.7	1+3	4	$\underline{3}$	3	+0.3	3.3
0×4	$\underline{\underline{0}}$	$\underline{4}$	2	–0.3	1.7	1×3	$\underline{3}$	4	3	–0.3	2.7	1+4	0	$\underline{4}$	3	+0.3	3.3
1×1	$\underline{1}$	2	2	–0.3	1.7	1×4	$\underline{4}$	0	3	–0.3	2.7	2+3	0	1	4		4
2+2	4		2		2	3+3	1	4	3		3	2×3	1	0	4		4
2×2	4		2		2	3×3	4	1	3		3	2+4	1	3	4		4
0+1	$\underline{1}$	$\underline{\underline{0}}$	2	+0.3	2.3	3+4	2		3		3	2×4	3	1	4		4

Table A1. Value of reaction time predictor X=M+P as a function of task T, correct response CR, and to be suppressed response SR (if different from CR). Primed responses are single underlined (weak prime) or double underlined (strong prime).

References

Berry D. (1997), *How Implicit is Implicit Learning?* (Oxford: Oxford University Press).

Epstein, R. (1992), 'The quest for the thinking computer', *AI Magazine*, **13***,* pp. 81–95.

Descartes, R. (1637), 'Discourse on method', Trans. John Cottingham, Robert Stoothoff and Dugald Murdoch, in *The Philosophical Writings of Descartes*, Vol. I, pp. 109–51 (New York: Cambridge University Press).

Harnad, S. (2001), 'Mind, machines and Searle II: What's wrong and right about Searle's Chinese room argument?' in *Essays on Searle's Chinese room Argument,* ed. M. Bishop and J. Preston (Oxford: Oxford University Press).

Hayes, P., Harnad, S., Perlis, D. & Block, N. (1992), 'Virtual symposium on virtual mind', *Minds and Machines,* **2**(3), pp. 217–38.

MacLennan, B.J. (2001), 'Grounding analog computers', *Psycoloquy*, **12**(052).

Nissen, M.J., Bullemer, P.T. (1987), 'Attentional requirements for learning: Evidence from performance measures', *Cognitive Psychology,* **19**, pp. 1–32.

Puccetti, R. (1980), 'The chess room: further demythologizing of strong AI', *Behavioral and Brain Sciences* **3**, pp. 441–2.

Reber, A.S. (1967), 'Implicit learning of an artificial grammar', *Journal of Verbal Learning and Verbal Behavior*, **6**, pp. 855–63.

Searle, J. (1980), 'Minds, brains, and programs', *Behavioral and Brain Sciences*, **3**, pp. 417–24.

Stadler, M. & Frensch, P. (1998), *Handbook of Implicit Learning* (Thousand Oaks, CA: SAGE Publications).

Sundman, J. (2003), 'Artificial stupidity', http://archive.salon.com/tech/feature/2003/02/26/loebner_part_one/

Turing, A. (1950), 'Computing machinery and intelligence', *Mind*, **LIX**(236), pp. 433–60.

Weiss, T. (1990), 'Closing the Chinese room', *Ratio* (New Series) **III**(2), pp. 165–81.