

Seminararbeit zum Thema:

John R. Searle:

„Das chinesische Zimmer“

Von Verena Haller, Andrea Pöham und Ingrid Hofmann

„Das Chinesische Zimmer“ - Ein Gedanken-Experiment zum Thema:

Künstliche Intelligenz oder vielleicht Maschinenintelligenz – Konkurrenz für den Menschen?

Als die künstliche Intelligenz vorankam, fingen Maschinen an auf Gebieten zu operieren, die als die Domäne der Menschen galt. Aus diesem Grund haben Menschen immer wieder versucht, die Einzigartigkeit ihrer Art unter Beweis zu stellen.

So überlegte der Philosoph **RENE DESCARTES** bereits vor 400 Jahren in seinem „Discours de la Methode“, wie man Maschinen mit menschlichen Äußerem von einem echten Menschen unterscheiden könnte.

Descartes nennt hier **zwei Methoden** der Unterscheidung zwischen Mensch und Maschine:

1) den Sprachtest und

2) den Handlungstest, wobei die den **Maschinen** fehlende Vernunft der Grund für das Nichtbestehen des Tests sei.

Descartes trennte auch die Seele als erkennendes Subjekt von der Gegenstands- und Körperwelt und betrachtete den entseelten Körper als technisch-konstruierten Apparat.

Im Verlauf dieser Entwicklung verändert sich die Wertschätzung der Maschine als Abbild, Vorbild oder Bild des Menschen. Descartes` Vergleich hat einen vorsichtigen neutralen Charakter und bezieht sich lediglich auf dem menschlichen Körper, nicht aber auf den Geist. Das frühe 19. Jahrhundert zeigte sich dem gegenüber ausgesprochen fasziniert von der Möglichkeit, der gesamte Mensch sei eine Maschine. Die Aufklärungszeit wiederum war durch eine explizite Maschinenkritik geprägt. Seit dem Entstehen der wissenschaftlichen Psychologie im 19. Jahrhundert schließlich sind Maschinenmodelle für den Menschen ein Teil dieser Wissenschaft, werden aber unterschiedlich bewertet.

Auch **ALAN TURING** beschrieb 1950 in einem Aufsatz ein Gedankenexperiment, das später als Turing-Test bekannt wurde, und das demselben Zweck wie Descartes Überlegungen, nämlich der Unterscheidung von Mensch und Maschine, diene.

Dieses Experiment basiert auf einem Gesellschaftsspiel, indem ein Spieler durch Fragen herausfinden sollte, ob sich in einem angrenzenden Zimmer ein Mann oder eine Frau befand. In Turings Test wurden „Mann oder Frau“ durch „Mensch oder Maschine“ ausgetauscht. Als Analogon der heutigen Zeit kann man sich einen Chat-Room vorstellen, wobei die Aufgabe des Fragestellers darin besteht, herauszufinden, ob sein

Gesprächspartner ein Mensch oder eine Maschine ist.

Turing folgerte unter der Annahme der CENTRAL-STATE-THEORIE, die davon ausgeht, dass der Sitz des Bewusstseins das Zentrale Nervensystem ist, das sich die geistige Aktivität des Menschen von einem Computer mit entsprechender Speicherkapazität nachbilden ließe.

Der TURING-TEST stellt sich die Frage:

Können Maschinen denken?

Im Sinne behavioristischer Theorie wurde danach eine Maschine „intelligent“ genannt, wenn eine Person, die über einen Terminal kommuniziert, durch Befragung nicht mehr feststellen kann, ob sich hinter dem Terminal ein Mensch oder ein Computer verbirgt.(Turing 1950).

Das kommt freilich darauf an, was man unter den Begriffen „Maschine“ und „Denken“ versteht. Beispielsweise könnte man die Ansicht vertreten, dass auch der Mensch eine Art von Maschine ist, **eben eine biochemische**. Und dann lautet die Antwort natürlich: **ja**.

Der bestandene Turing-Test heißt aber nicht, dass **Maschinen** im menschlichen Sinn denken können, sondern nur, dass sie ein Verhalten zeigen, das sich auf der Ebene textueller Interaktion von menschlichem Verhalten kaum mehr unterscheiden lässt.

Turing war der Meinung, dass sein Versuch ein guter Intelligenztest für einen Computer sei. Es ist jedoch fraglich, ob man die von Menschen einprogrammierten Antworten auf bestimmte Fragen als **Computer-Intelligenz** bezeichnen kann, und da sich die Kapazität des Computers stetig verdoppelt, kann einem Computer auch immer mehr Input gegeben werden. Ein solcher Test würde höchstens für die Intelligenz und Fleiß des Programmierers sprechen, sich auf jede erdenkliche Fragestellung eine Antwort auszudenken.

TURINGS Sicht, dass Denken ein Informationsverarbeitungsprozess ist, gehört in die Sichtweise des FUNKTIONALISMUS. Diese in der heutigen Philosophie des Geistes vertretene Richtung findet sich auch in der Psychologie besonders in der Kognitionswissenschaft wieder und versucht komplexe Tatbestände durch kausale Wirkungsbeziehungen zu erklären.

Turing, der englische Mathematiker vertritt die Meinung, jede Idee, jeder Gegenstand und jede Tätigkeit, die durch Sprache ausgedrückt werden könne, lasse sich im Prinzip in einer Kette von Nullen und Einsen verschlüsseln. Diese Zeichenfolgen könnten dann vom Programm einer Rechenmaschine umgewandelt und schließlich als neue Aussagen entschlüsselt werden.

Wittgenstein hingegen ist der Ansicht, dass die Kriterien für Sprache und Denken nicht in der Logik von Maschinen, Bändern und Codes, sondern in der tatsächlichen Praxis einer Sprach-gemeinschaft stecken.

Bedeutung sei letztlich in sozialer Praxis zu Hause, nicht in der Logik, und zum Denken gehöre eben sehr viel mehr als nur das Befolgen einer Reihe von Regeln.

Ob eine Maschine denken kann oder nicht, schrieb Turing im Jahr 1950, müssten wir auf der Ebene sprachlicher Äußerungen entscheiden. Wenn eine Testperson einem unbekanntem Gesprächspartner X via Bildschirm Fragen stellt und nicht entscheiden kann, ob die Antworten von einer Maschine oder einem Menschen kommen, so Turing, dann müsse man X eigentlich Denkfähigkeit zubilligen. Somit könnten auch Maschinen denken, sofern sie menschliche Kommunikation gut genug imitieren.

Das ist ohne Zweifel ein sehr pragmatischer Zugang. Denn erschließt sich Denkfähigkeit- oder allgemeine gesagt: *Geist* – nur über Äußerungen? Oder ist Geist nicht auch etwas Innerliches, etwas, das über rein maschinelle Prozesse hinausgeht?

Der amerikanische Philosoph John Searle sieht nun genau darin die Schwachstelle der Turingschen Argumentation. Im Jahre 1982 wurde der Turing-Test und seine Auffassung von maschineller Intelligenz durch Searles Gedankenexperiment „Das Chinesische Zimmer“ herausgefordert.

Sein Experiment, das in die Geistesgeschichte einging, beschrieb ein System, das den Turing-Test in Bezug auf das Verständnis der chinesischen Sprache bestehen würde.

Zur Person von John Searle

John R. Searle, geboren 1932 in Denver, Colorado und ist vermutlich der bekannteste zur Zeit lebende Philosoph. Seit 1959 ist Searle Professor für Philosophie an der University of California, Berkeley. Sein Werk basiert auf der Arbeit von J. L. Austin und wie schon erwähnt eben auf L. Wittgenstein. Seine Themen sind die Philosophie des Geistes, die Sprachphilosophie sowie die Metaphysik. Mit seinem Gedankenexperiment des „Chinesischen Zimmers „ hat John Searle die **Debatte um die starke KI** (auch AI) und die Frage, ob Programme wirklich in der Lage sind zu denken, stark belebt.

Searle, bringt also einen Stein ins Rollen...

Er veröffentlicht 1980 seinen Artikel „***Minds, brains and programs***“ in Behavioral and Brain Sciences.

Searle formuliert **seine These** so:

sowohl die Wahrheit des gesunden Menschenverstandes als auch die Wahrheit der Wissenschaft sind richtig. Er legt hier die Konsequenzen seiner Idee dar, insbesondere für das Leib-Seele-Problem, die künstliche Intelligenz, die kognitive Wissenschaft, für Fragen des Handelns und des freien Willens, sowie für die Philosophie der Sozialwissenschaften.

Im Artikel erläutert er, warum er die Argumente der starken KI nicht teilt und versucht diese mit seinem

mittlerweile berühmten Gedankenexperiment des Chinesischen- Zimmers zu widerlegen. Er bekräftigt, dass es ihm alleine um die starke KI geht und dass er die schwache KI durchaus als ein wertvolles Werkzeug anerkennt.

Kurz zur Auffassung der Forscher von Starker KI und Schwacher KI :

Grundsätzlich ist KI der Versuch, auf Computern intelligentes Verhalten nachzubilden.

1. **starke KI** – die richtig programmierten Prozesse im Computer sind geistige Prozesse, der Computer ist also intelligent
? Man kann Computern, die mit den „richtigen“ Programmen ausgestattet sind, Verstehen zusprechen
2. **schwache KI** – der Computer ist nur Instrument zur Untersuchung kognitiver Prozesse, der Computer simuliert nur Intelligenz

Searle geht von Roger Schanks Programm S.A.M. aus, um im darauf folgenden Abschnitt mit seinem Chinesischen Zimmer zu kontern. SAM muss in seinem Aufsatz stellvertretend für alle anderen Programme der starken KI wie z.B. ELIZA oder jede Art von Turing-Maschine herhalten.(S.A.M. ist ein Programm, das in der Lage ist, Fragen bezüglich Geschichten, die ihm zuvor in (schriftl. Form) erzählt wurden, zu beantworten).

Im Zentrum der Argumentation Searles stehen dabei immer wieder Begriffe wie Intentionalität und wie er es nennt „causal powers“. Leider versucht Searle nicht darzulegen, was er eigentlich unter diesen kausalen Kräften des Gehirns genau versteht.

Und wirbelt Staub auf

Searle bekommt teilweise Zuspruch jedoch auch viel Kritik und so verästelte sich die Diskussion über viele Jahre. Ende der Debatte ist keines in Sicht, und vielleicht ist das auch gut so, das Thema ist schließlich nicht unspannend.

Der Artikel der das Gedankenexperiment des Chines. Zimmers erläutert und der starken KI deutlich widerspricht, wird bis heute in der Kognitionswissenschaft heiß diskutiert und ist so aktuell wie zum Zeitpunkt seiner Erscheinung.

Searle wird aufgrund der Brisanz dieses Gedankenexperiments für die KW (Kognitionswissenschaft) und der damit verbundenen Kritik an der starken KW gerne auch als Provokateur bezeichnet.

Im obengenannten Artikel geht Searle von zwei Annahmen aus:

- 1) Intentionalität ist ein Produkt kausaler Prozesse im Gehirn. Diese Prozesse sind hinreichend oder

anders ausgedrückt: Intentionalität ist ein Feature (Merkmal) des Gehirns.

- 2) Ein Computerprogramm auszuführen reicht nicht aus, um Intentionalität zu erzeugen.

Als Schlussfolgerung ergibt sich daraus:

- 3) Das Gehirn erzeugt seine Intentionalität nicht, indem es ein Computerprogramm ausführt.(ergibt sich aus 1 und 2)
- 4) Jeder Mechanismus, der in der Lage ist Intentionalität zu erzeugen, muss die gleichen kausalen Kräfte besitzen wie unser Gehirn.(folgt aus 1)
- 5) Jeder Versuch Intentionalität künstlich zu erzeugen(starke KI) wird nur dann Erfolg haben, wenn er nicht nur Programme **designed**, sondern die kausalen Kräfte des Gehirns dupliziert.(folgt aus 2 und 4).

Zu John R. Searles Artikel mit dem Titel „Geist, Gehirn, Programm“ („Minds, Brains and Program“):

Zur Einleitung des Artikels:

In seinem Artikel unterscheidet Searle in der Einleitung zwischen schwacher und starker AI (= artifizielle Intelligenz).

„Nach der starken AI“, schreibt Searle, „ist der Computer aber nicht nur ein Instrument bei der Untersuchung des Geistes; vielmehr *ist* der recht programmierte Computer *selbst* ein Geist, in dem Sinne, dass Computern, die mit den richtigen Programmen ausgestattet sind, Verstehen und andere kognitive Zustände zugeschrieben werden können.“

Weil bei der starken AI der programmierte Computer kognitive Zustände aufweist, sind dort die Programme nicht bloss ein Instrumentarium, das uns die Überprüfung psychologischer Erklärungen erlaubt, wie es bei der schwachen AI der Fall ist; vielmehr stellen die Programme selbst Erklärungen dar.“

John Searles Aufsatz richtet sich gegen die eben ausgeführten Ansprüche der starken AI, speziell gegen den Anspruch, dass der recht programmierte Computer kognitive Zustände aufweist, und dass die Programme daher menschliches Erkennen erklären.

Searles Überlegungen knüpfen an die Arbeit von Roger Schank und seinen Kollegen an der Yale Universität an, weil er mit ihr vertrauter ist als mit anderen ähnlichen Arbeiten, wie er sagt. Trotzdem lassen sich seine Überlegungen, genauso gut auf Winograds SHRDLU oder Weizenbaums ELIZA, sowie auf jede Maschinensimulation menschlicher geistiger Phänomene anwenden, die den Turing-Test besteht.

Kurze Erläuterung von Schanks Programm:

Ziel des Programms ist die Simulation der menschlichen Fähigkeit, Geschichten zu verstehen.

Für die Fähigkeit menschlicher Wesen, Geschichten zu verstehen, ist kennzeichnend, dass sie imstande sind, Fragen zu einer Geschichte zu beantworten, auch wenn die Auskunft, die sie geben, nicht expliziter Inhalt der Geschichte war.

Nehmen wir zum Beispiel an, es wird uns die folgende Geschichte erzählt:

„Ein Mann ging in ein Restaurant und bestellte einen Hamburger. Als der Hamburger kam, war er total verbrannt und nur noch ein zusammengeschnittenes Krüstchen; der Mann stürmte wütend aus dem Lokal, ohne zu bezahlen bzw. ein Trinkgeld dazulassen.“

Wenn uns nun jemand fragt: „Hat der Mann den Hamburger gegessen?“, werden wir voraussichtlich antworten: „Nein, hat er nicht.“

Schanks Programme können Fragen in der Art und Weise beantworten, wie wir sie von menschlichen Wesen erwarten würden.

Verfechter der starken AI würden sagen, dass bei derartigen Frage-und-Antwort-Folgen:

1. die Maschine nicht bloss eine menschliche Fähigkeit simuliert, das heisst vortäuscht, sondern
2. dass sie die Geschichte versteht, und dass das, was die Maschine und ihr Programm tun, eine Erklärung der menschlichen Fähigkeit darstellt, eine Geschichte zu verstehen.

Für beide Behauptungen scheint Searle Schanks Arbeit nicht die mindeste Rechtfertigung zu geben.

Eine Möglichkeit der Überprüfung einer bestimmten Theorie des Geistes besteht laut Searle darin, sich die Frage zu stellen, wie es wäre, wenn der Geist tatsächlich auf der Grundlage der Prinzipien funktionierte, die die Theorie für alle Geister voraussetzt.

Searle folgert daraus: „Wenden wir nun diese Prüfmethode auf Schanks Programm an, indem wir folgendes Gedankenexperiment anstellen:“

Zum Gedankenexperiment „Das chinesische Zimmer“ selbst:

Nehmen wir an, ich, John Searle, bin in einem Raum eingesperrt, und man gibt mir einen Stoss Zettel mit chinesischer Schrift. Nehmen wir weiter an, dass ich, was tatsächlich der Fall ist, weder Chinesisch schreiben noch sprechen kann, und dass ich nicht einmal sicher bin, dass ich chinesische Schrift als chinesische Schrift erkennen und von zum Beispiel japanischer Schrift oder sinnlosem Gekritzeln unterscheiden kann.

Nehmen wir nun noch weiter an, dass man mir nach dem ersten Stoss Zettel mit chinesischer Schrift einen zweiten Stoss Zettel mit chinesischen Schriftzeichen gibt, zusammen mit einer Reihe von Anleitungen, diese auf Englisch, meiner Muttersprache, die mich anweisen, wie ich den zweiten Stoss zum ersten Stoss in Beziehung setzen soll. Diese Anleitungen ermöglichen es mir ausschließlich eine Reihe formaler Symbole zu einer anderen Reihe formaler Symbole in Beziehung zu setzen. Formal bedeutet, dass ich diese Symbole

ausschließlich anhand ihrer Form identifiziere.

Nehmen wir nun auch noch an, dass man mir einen dritten Stoss Zettel mit chinesischen Symbolen gibt, zusammen mit einigen Anleitungen, ebenfalls auf Englisch, die es mir ermöglichen, Teile dieses dritten Stosses mit den beiden anderen in Beziehung zu setzen. Diese Anleitungen weisen mich an, bestimmte Symbole in Antwort auf bestimmte Symbole, die mir mit dem dritten Stoss zugegangen sind, zurück zu geben.

Was ich nicht weiss, ist, dass diejenigen, die mir die Stösse gegeben haben, den ersten Stoss „Schrift“, den zweiten Stoss „Geschichte“ und den dritten Stoss „Fragen“ nennen. Desweiteren nennen sie die Symbole, die ich ihnen in Antwort auf den dritten Stoss zurück gebe „Antworten auf die Fragen“, und die Reihe von englischsprachigen Anleitungen, nennen sie „Programm“.

Nun stellen wir uns weiter vor, dass mir auch Geschichten in Englisch gegeben werden, zu denen ich auf die gleiche Weise Fragen beantworte, immer auf Englisch.

Nehmen wir nun noch abschließend an, dass ich nach einer gewissen Zeit den Anweisungen für das Hantieren mit den chinesischen Symbolen so gut zu folgen lerne, und die Programmierer so gut lernen, Programme zu schreiben, dass, vom Standpunkt eines Menschen aus betrachtet, der sich außerhalb des Raumes befindet, in dem ich eingeschlossen bin, meine Antworten auf die Fragen absolut ununterscheidbar sind von den Antworten, die einer geben würde, dessen Muttersprache Chinesisch wäre. Niemand, der nur meine Antworten liest, könnte erkennen, dass ich kein Wort Chinesisch spreche. Von außen gesehen sind also die Antworten auf die chinesischen Fragen gleich gut wie die auf die englischen. Jedoch was das Chinesische angeht, verhalte ich mich wie ein Computer. Ich führe kalkulatorische Operationen an formal spezifizierten Elementen aus.

Die starke AI behauptet, dass der programmierte Computer die Geschichten versteht und dass das Programm in gewissem Sinne menschliches Verstehen erklärt. Nun wollen wir diese Behauptungen im Lichte des Gedankenexperimentes überprüfen, meint Searle in weiterer Folge des Artikels:

1. Was die erste Behauptung angeht, liegt es doch bei unserem Beispiel mit dem „chinesischen Zimmer“ ganz klar auf der Hand, dass der Mann im „chinesischen Zimmer“ kein Wort von der chinesischen Geschichte versteht. Er hat Eingaben und Ausgaben, die sich von einem, der Chinesisch als Muttersprache spricht, nicht unterscheiden. Und er mag über jedes beliebige formale Programm verfügen, dennoch versteht er nichts.

Aus diesem Grund versteht auch Schanks Computer nichts von den Geschichten, egal in welcher Sprache sie ihm dargeboten werden bzw. er sie darbietet.

2. Was die zweite Behauptung angeht, wird deutlich, dass der Computer und sein Programm für das Verstehen keine hinreichenden Bedingungen schaffen. Der Computer und das Programm funktionieren, ohne dass es zu einem Verstehen kommt, sagt Searle.

Vertreter der starken AI gehen zumindest davon aus, dass Schanks Programm, auch wenn es nicht die ganze Wahrheit des Verstehens verkörpert, dennoch einen Teil davon verkörpert. Searle vermutet, dass das zwar empirisch denkbar ist, doch es bislang keinen Beweis dafür gibt, dass es sich tatsächlich so verhält, da ja das Beispiel mit dem „chinesischen Zimmer“ darauf hindeutet, wenn auch zweifellos nicht den Beweis dafür

erbringt, dass das Computerprogramm für das Verständnis der Geschichte belanglos ist.

Im Fall des Chinesischen hat der Mensch im „chinesischen Zimmer“ alles, was ihm die artifizielle Intelligenz mittels eines Programmes zu geben vermag, trotzdem versteht er nichts.

Im Falle des Englischen versteht er alles, und es gibt bislang keinen Grund anzunehmen, dass sein Verständnis das Mindeste mit Computerprogrammen zu tun hat, das heisst mit kalkulatorischen Operationen an rein formal spezifizierten Elementen. Solange sich das Programm als eine Reihe von kalkulatorischen Operationen an rein formal definierten Elementen bestimmt, legt das Beispiel mit dem „chinesischen Zimmer“ den Schluss nahe, dass diese Operationen keine hinreichenden Bedingungen für das Verstehen darstellen. Und es gibt auch keinen Grund anzunehmen, dass sie eine notwendige Voraussetzung oder auch nur einen wesentlichen Beitrag dazu leisten.

Der Punkt ist hier der, dass, egal welche formalen Prinzipien einem Computer eingegeben werden, diese für's Verstehen nicht ausreichend sind, da ein Mensch imstande ist, eben diese formalen Prinzipien zu befolgen, ohne dabei das Mindeste zu verstehen. Und es besteht auch kein Grund zur Annahme, dass jemand, der Englisch versteht, überhaupt mit irgend einem formalen Programm operiert, meint Searle.

Zu den Repliken:

Kritiker Searles weisen darauf hin, dass es verschiedene Grade von Verstehen gibt, ja sogar verschiedene Arten und Ebenen von Verstehen; dass es in vielen Fällen keine Tatsachen-, sondern eine Ermessensfrage ist, ob x von y verstanden wird. Auch darauf bezieht sich Searle in seinem Artikel.

Er meint: Wir legen Autos, Rechenmaschinen und sonstigen Kunsterzeugnissen metaphorisch bzw. analogisierend „Verstehen“ oder sonstige kognitive Prädikate nahe, doch solche Zuschreibungen beweisen nichts. Wir sagen z.B. „Die Tür *weiss*, wann sie aufzugehen hat, dank ihrer photoelektrischen Zelle.“ oder „Der Thermostat *nimmt* Temperaturen *wahr*“. Der Grund, warum wir diese Zuschreibungen machen, hat mit dem Umstand zu tun, dass wir auf Kunsterzeugnisse unsere eigene Intentionalität übertragen. Unsere Werkzeuge sind der verlängerte Arm unserer Zielsetzungen, und deshalb finden wir es natürlich, ihnen metaphorisch Intentionalität zuzuschreiben.

Im Übrigen bedeutet Verstehen sowohl das Vorhandensein bestimmter geistiger (intentionaler) Zustände, als auch das Wahrsein (die Gültigkeit) dieser Zustände.

Für die Zwecke dieser Erörterung sind wir nur mit dem Vorhandensein der Zustände befasst.

Intentionalität ist definitionsgemäß jene Eigenschaft bestimmter Geisteszustände, durch die diese sich auf Gegenstände oder Angelegenheiten in der Welt beziehen. Überzeugungen, Begierden und Absichten sind demnach intentionale Zustände.

In weiterer Folge des Artikels geht Searle auf verschiedene Entgegnungen, Repliken ein:

1. Die System-Replik (Berkeley)
2. Die Roboter-Replik (Yale)

3. Die Gehirnsimulator-Replik (Berkeley und MIT)
4. Die Kombinations-Replik (Berkeley und Stanford)
5. Die Replik mit dem Geist anderer (Yale)
6. Die Replik mit den vielen Behausungen (Berkeley)

Ad 1.:

Die System-Replik wendet ein, dass die Person in dem Raum die Geschichte zwar nicht versteht, dass sie jedoch Teil eines Systems ist, das die Geschichte durchaus versteht. Daraufhin erwidert Searle, dass die Person, auch wenn sie sich sämtliche Teile des Systems einverleibt, dennoch nicht Chinesisch versteht. Wenn die Person nicht verstehen kann, dann gibt es auch keinerlei Verstehensmöglichkeit für das System.

Während zum Beispiel das englische System, wenn wir einen Augenblick in diesem Jargon bleiben, weiss, dass zum Beispiel mit dem Wort „Hamburger“ ein Hamburger, etwas Essbares, gemeint ist, weiss das chinesische System bzw. Subsystem (es sind ja zwei Systeme, das englische und das chinesische, in ein und derselben Person) nur, dass auf ein bestimmtes Symbol ein bestimmtes Symbol folgt, nicht aber, was die Symbole bedeuten.

Searle führt weiter aus, dass die Systemreplik den Beweis einfach schuldig bleibt, einfach darauf insistiert, dass das System ein Verständnis des Chinesischen habe.

Ad 2.:

Die Roboterreplik meint: Angenommen wir stecken einen Computer in einen Roboter und dieser würde den Roboter in einer Weise funktionieren lassen, die Tätigkeiten wie Wahrnehmen, Gehen, Essen etc. sehr nahe kämen. Am Roboter würde eine Kamera angebracht sein, die ihn zum Sehen befähigen würde, und er hätte Arme und Beine, die ihn zum „Handeln“ befähigen würden. All dies würde von einem Computer-„Hirn“ gesteuert. Anders als Schanks Computer würde ein solcher Roboter echtes Verstehen und andere geistige Zustände aufweisen.

Searles Antwort auf die Roboterreplik ist, dass die Hinzufügung von Wahrnehmungs- und Bewegungsfähigkeiten in verstehensmäßiger und intentionaler Hinsicht an Schanks ursprünglichem Programm nichts ändert. Das wird erkannt, wenn das Gedankenexperiment des chinesischen Zimmers auf den Roboterfall angewendet wird; und zwar, wenn wir annehmen, dass statt eines Computers ein Mensch im Roboter ist. Dieser Mensch würde chinesische Symbole mit chinesischen Symbolen zusammenfügen und an die Außenwelt chinesische Symbole auswerfen. Nehmen wir an, dass einige der chinesischen Symbole durch eine am Roboter angebrachte Fernsehkamera zu der Person gelangen, andere, die die Person auswirft, die Funktion haben, Beine oder Arme des Roboters in Bewegung zu setzen. Alles was die Person im Roboter tut, ist ein Hantieren mit formalen Symbolen. In diesem Fall kennt der Roboter keinerlei intentionale Zustände.

Ad 3.:

Die Gehirnsimulatorreplik wendet ein: Angenommen wir entwerfen ein Programm, das die tatsächliche Abfolge der Neuronenaktivität in den Synapsen des Gehirns von jemandem simuliert, dessen Muttersprache Chinesisch ist, während dieser Geschichten auf Chinesisch hört und Fragen dazu beantwortet. Die Maschine nimmt als Eingabe chinesische Geschichten und Fragen über diese Geschichten auf, simuliert die formale Struktur wirklicher Chinesengehirne bei der Verarbeitung dieser Geschichten und wirft chinesische Antworten als Ausgabe aus. In einem solchen Fall müssen wir doch sagen können, dass die Maschine die Geschichte versteht.

Hier antwortet Searle folgendes: Selbst wenn wir der Funktionsweise des Gehirns so außerordentlich nahe kämen, würde auch das noch nicht genügen, Verstehen zu erzeugen.

Das Problem beim Gehirnsimulator ist, dass er vom Gehirn die falschen Sachen simuliert. Solange er nur die formale Struktur der Neuronenaktivierungen in den Synapsen simuliert, hat er nicht das am Gehirn simuliert, worauf es ankommt, nämlich seine Fähigkeit, intentionale Zustände hervorzubringen.

Ad 4.:

Die Kombinationsreplik besagt: Wenn wir alle drei vorangegangenen Repliken zusammenfassen, sind sie erheblich überzeugender. Stellen wir uns einen Roboter mit einem hirnförmigen Computer in der Höhlung unter seinem Schädel vor; stellen wir uns den Computer mit allen Synapsen eines menschlichen Gehirns programmiert vor; stellen wir uns das ganze Verhalten des Roboters als ununterscheidbar vom menschlichen Verhalten vor, und zum Schluss denken wir uns das Ganze als vereinigt System und nicht einfach als einen Computer. In diesem Fall würden wir doch dem System zweifellos Intentionalität zuschreiben.

Searle entgegnet: Wenn wir einen Roboter bauen könnten, dessen Verhalten im Grossen und Ganzen vom menschlichen Verhalten ununterscheidbar wäre, würden wir ihm Intentionalität zuschreiben. Dazu müssten wir nicht Kenntnis davon haben, dass sein Computergehirn eine formale Entsprechung des menschlichen Gehirns darstellte.

Searle sieht allerdings nicht, wie damit den Ansprüchen der starken AI geholfen wäre; denn der starken AI zufolge ist die Realisierung eines formalen Programms mit den richtigen Ein- und Ausgaben eine hinreichende Bedingung für Intentionalität, und um mit Newell zu sprechen, auf den sich die starke AI stützt, besteht das Wesentliche des Geistigen im Funktionieren eines physischen Symbolsystems; und wenn wir dem Roboter im obig angeführten Beispiel Intentionalität zuschreiben, hat das mit formalen Programmen nicht das Geringste zu tun.

Ad 5.:

Die Replik mit dem Geist anderer besagt: Woher wissen wir, dass andere Menschen Chinesisch oder was auch immer verstehen? Aus ihrem Verhalten. Nun kann der Computer die Verhaltenstests genauso gut bestehen, wie Menschen das können. Wenn wir den Menschen also Erkenntnisvermögen zuschreiben, müssen wir das im Prinzip auch Computern.

Hierauf meint Searle, dass es ihm in seinem Artikel nicht darum gehe, die Frage zu erörtern, woher wir wissen, dass andere Menschen kognitive Zustände haben. In den Kognitionswissenschaften ist die Wirklichkeit und Erkennbarkeit des Geistigen vorausgesetzt wie in den physikalischen Wissenschaften die Wirklichkeit und Erkennbarkeit des Physischen.

Ad 6.:

Die Replik mit den vielen Behauptungen besagt: Wie auch immer diese kausalen Prozesse aussehen, die für Intentionalität ausschlaggebend sind, irgendwann werden wir in der Lage sein, Apparate zu bauen, die über diese kausalen Prozesse verfügen, und das wird dann die artifizielle Intelligenz sein.

Diesbezüglich ist Searle der Meinung, dass damit das Projekt der starken AI trivialisiert wird, indem als artifizielle Intelligenz alles angesehen wird, das Erkenntnis künstlich produziert und erklärt. Er hingegen bezog sich mit seinem Gedankenexperiment vielmehr auf die These, dass geistige Prozesse am Modell des Rechners orientierte Prozesse sind, die an formal definierten Elementen ablaufen. Sein Ziel war es, diese These zu bestreiten. Wenn der Anspruch neu bestimmt wird, sodass er dieser These nicht mehr entspricht, meint er, greifen seine Einwände auch nicht mehr.

Der Beweis gegen die starke KI

Searle entwickelte vier Axiome und vier Schlussfolgerungen, um den formalen Gegenbeweis gegen die starke KI zu liefern.

Das erste Axiom: „Computerprogramme sind formal (syntaktisch)“

Ein Computerprogramm bearbeitet Informationen, in dem er sie in seine eigene Symbolsprache verschlüsselt und anschließend diese Information mit Hilfe genau festgelegter Regeln, dem Programm, manipuliert.

Genau dies ist der Grund, dass unsere heutigen Rechner so leistungsstark sind.

Interessant ist dabei auch, dass jede Information, die sich in einem solchen System codieren lässt und mit diesen expliziten Regeln gelöst werden kann, programmierbar ist. Diese Programme sind, wie auch Symbole, rein abstrakte Gebilde und deshalb in jedem physikalischen System darstellbar. Ein Programm hat also Syntax, aber keine Semantik.

Der Computer manipuliert das Symbol nur unabhängig von dessen Bedeutung.

Das zweite Axiom: „Dem menschlichen Denken liegen geistige Inhalte (Semantik) zu Grunde“

Menschliche Gedanken, Wahrnehmungen, Einsichten und ähnliches haben immer geistige Inhalte, die wir uns wiederum durch Kommunikation mit anderen mitteilen. Um Sprache zu verstehen braucht man immer Semantik, um die Wörter, die man hört oder von sich gibt, auch deuten zu können.

Ein Beispiel, in dem das gut klar wird: Wenn jemand an die letzten Wahlversprechen denkt, fallen wahrscheinlich jedem von uns ein paar Worte ein. Das geschieht nur, weil wir der deutschen Sprache mächtig sind und mit den Worten eine bestimmte Bedeutung verbinden.

Das dritte Axiom: „Syntax an sich ist weder konstitutiv noch hinreichend für Semantik“

Ein Symbol alleine, also Syntax, bedeutet noch nicht, dass man auch versteht, welche Bedeutung mit dem betreffenden Symbol in Zusammenhang gebracht werden soll. So impliziert das bloße Hantieren mit Symbolen – wie es Computerprogramme betreiben – nicht, dass auch die Semantik verstanden wird.

Ein Beispiel dazu wäre eine Vorrangregel. Ein Computerprogramm erkennt hier nur ein gleichseitiges Dreieck in den Farben rot und weiß, das auf dem Kopf steht. Es weiß aber nicht, dass dieses Zeichen im Straßenverkehr bedeutet, dass man dem Querverkehr den Vorrang geben muss.

Die erste Folgerung: „Ein Programm beinhaltet keinen Verstand und stellt auch keinen dar“

Es gibt einen klaren Unterschied zwischen formalen Elementen (Begriffen), die eine Bedeutung haben und solchen, die keine Bedeutung haben. Daraus folgt zwangsläufig, dass die Annahme, welche die starke KI vertritt, falsch ist.

Auf den ersten Blick ist die Folgerung korrekt und die drei Axiome scheinen wahr zu sein.

Allerdings haben Paul und Patricia Churchland dem etwas entgegen zu setzen: Übereinstimmend mit der Position Searles vertreten auch die Churchlands die Meinung, dass das chinesische Zimmer nicht wirklich Chinesisch verstehe. Kritik üben sie jedoch am 3.Axiom. Ihnen geht es nicht darum zu behaupten, dieses Axiom entspreche nicht der Wahrheit. Vielmehr kritisieren sie die Sicherheit, mit welcher Searle von der Richtigkeit dieses Axioms spricht. Sie bemängeln, dass das dritte Axiom noch nicht erwiesen ist. Wenn dem so wäre, dann wäre alle Anstrengung der KI, aus Syntax Semantik zu erzeugen, dahin. Schließlich nehme die KI ja gerade an, dass es möglich ist Semantik allein aus Syntax zu erzeugen. Vertreter der KI könnten ebenso als Axiom annehmen: Syntax *ist* konstitutiv und hinreichend für Semantik.

Sie werfen Searle vor, er hielte nur an seiner Argumentation fest, weil er sich nicht vorstellen könne, dass Syntax ausreiche um Semantik zu erzeugen. Die Churchlands vergleichen seine Haltung mit der Auffassung

eines bekannten irischen Bischofs aus dem 18. Jahrhundert, der sich nicht vorstellen konnte, dass Druckwellen in der Luft hinreichend für Laute sein könnten. Mit diesem Vergleich wollen die Churchlands sagen: *„Nur weil wir uns nicht vorstellen können, dass etwas der Fall ist, heißt noch lange nicht, dass es nicht tatsächlich der Fall ist. Und mit solch einem Beispiel haben wir es hier zu tun!“*

Das vierte Axiom: „Gehirne verursachen Geist“

Prozesse, die einen Computer als solchen definieren, sind gänzlich unabhängig von ihrer Hardware – ein gravierender Unterschied zum menschlichen Gehirn. Man weiß immer noch sehr wenig darüber, wie mentale Zustände im Gehirn erzeugt werden. Wo man es weiß, ist heute klar, dass dies durch neurobiologische Prozesse Zustände kommt.

Man nehme zum Beispiel einen Gedanken, den ein Mensch fasst. Dieser Gedanke entsteht, durch das Feuern spezifischer Neuronen in spezifische Zellverbände. Das Gehirn aktiviert also nicht einfach ein formales Muster oder Programm – was es zwar auch tut – sondern, erzeugt auch mentale Ereignisse mit Hilfe dieser neurobiologischen Abläufe. Ein Computerprogramm hingegen liefert jedoch nur eine Simulation von Verstand, verursacht aber keinen Verstand.

Um das Ganze klarer werden zu lassen, nehme man als Beispiel einen Verdauungsprozess her. Man kann mit einem Programm zwar gut simulieren, welche Vorgänge bei der Verdauung ablaufen, was aber nicht heißt, dass dieses Programm auch wirklich ein Stück Fleisch essen, verdauen und wieder ausscheiden kann.

Genauso verhält es sich auch mit dem menschlichen Gehirn.

Die zweite Folgerung: „Jedes andere System, das Geist hervorrufen kann, benötigt kausale Kräfte, die denen von Gehirnen (mindestens) äquivalent sind“

Diese Folgerung besagt im Prinzip nur, dass jedes System, das Verstand hervorruft über dieselben kausalen Kräfte verfügen muss wie das menschliche Gehirn.

Hierzu eignet sich als Beispiel der Vergleich zwischen Elektromotor und Benzinmotor. Ein Elektromotor, der ein Fahrzeug gleich schnell antreiben soll wie ein Benzinmotor benötigt auch eine äquivalente Menge an Kraft, damit er dieselbe Leistung erbringen kann.

Die dritte und vierte Folgerung: „Jedes Artefakt, das mentale Phänomene erzeugt, also jedes künstliche Gehirn, muss imstande sein, die spezifischen kausalen Kräfte von Gehirnen aufzubringen und dies ist nicht einfach durch Ausführen eines formalen Programms zu erreichen“

Es reicht also nicht aus, ein Computerprogramm ablaufen zu lassen, um menschliche Emotionen zu produzieren.

Was uns John R. Searle mit seinen Beweisen gegen die starke KI nicht zeigen wollte

John R. Searle fühlt sich von vielen missverstanden und weist deshalb extra darauf hin, was er nicht mit seinen Beweisen sagen wollte.

Das wichtigste dabei ist, dass er nicht versucht hat zu beweisen, dass ein Computer nicht Denken könne – er wollte lediglich die Annahme der starken KI, nämlich, dass das menschliche Gehirn ein Programm sei, widerlegen. Es ist nämlich durchaus möglich, bestimmte Arbeitsweisen des Gehirns rechnerisch zu simulieren. Nur, Denken und die reine Manipulation von Symbolen sind etwas Verschiedenes.

Auch wollte Searle nicht behaupten, dass nie eine denkende Maschine erfunden werden könnte bzw. dass nur biologische Systeme dazu in der Lage sind.

Mit seinem Modell des Chinesischen Zimmer hat er nur versucht zu zeigen, dass ein Programm keinen eigenen Gedanken hervorbringen kann, da es nur mit Symbolen hantiert.

Literaturliste:

Googlesuche:

- Ingrid Scharlau: „Mensch und Maschine“
- Dieter E. Zimmer: „Mr. Searle im chinesischen Zimmer“
- Marc Wilnauer: „Handout zum Kurzreferat über das chinesische Zimmer“
- Simon J. Büchner: „John R. Searle und das chinesische Zimmer“

Englischer Artikel von John R. Searle „Minds, brains and program“

„Einsicht ins Ich“ von Hofstädter und Dennett