

Das chinesische Zimmer

In dem Gedankenexperiment vom chinesischen Zimmer geht es um unser Verständnis vom menschlichen Geist. Es behandelt die Frage, ob sich geistige Zustände allein in Analogie zu einem Computer verstehen lassen: Handelt es sich bei ihnen um Zustände, die in computeranalogen Verarbeitungsprozessen entstehen und die durch ihre Stellung in solchen Verarbeitungsprozessen vollständig charakterisiert werden können? John Searle (*1932) hat 1980 das Gedankenexperiment des chinesischen Zimmers entwickelt, um ein entsprechendes Verständnis des menschlichen Geistes zu widerlegen. Es geht ihm darum, zu zeigen, dass eine Stellung in einem computeranalogen Verarbeitungsprozess nur die *strukturelle Form* eines bestimmten Zustands verständlich macht, aber nicht seinen *Inhalt*. Searles These lautet, dass wir geistige Zustände gar nicht verstehen

können, wenn wir sie nicht als Zustände verstehen, die einen bestimmten Inhalt haben. Genau dies mache die Analogie zu einem Computer nicht verständlich. Das Gedankenexperiment vom chinesischen Zimmer (in Searles Buch *Minds, Brains and Science*, 1983) ist somit ein Gedankenexperiment, das auf die Verteidigung einer bestimmten These zielt. Das Szenario des Gedankenexperiments ist dafür entwickelt worden, für diese bestimmte These zu argumentieren. Dennoch hat das Gedankenexperiment Bedeutung für unser Verständnis des menschlichen Geistes im allgemeinen und auch für das Verständnis von Sprache und ihrer Bedeutung im besonderen.

»[...] stellen Sie sich vor, Sie wären in ein Zimmer eingesperrt, in dem mehrere Körbe mit Chinesischen Symbolen stehen. Und stellen Sie sich vor, daß Sie (wie ich) kein Wort Chinesisch verstehen, daß Ihnen allerdings ein auf Deutsch verfaßtes Regelwerk für die Handhabung dieser chinesischen Symbole gegeben worden wäre. Die Regeln geben rein formal – nur mit Rückgriff auf die Syntax und nicht auf die Semantik der Symbole – an, was mit den Symbolen gemacht werden soll. Eine solche Regel mag lauten: ›Nimm ein Kritzel-Kratzel-Zeichen aus Korb 1 und lege es neben ein Schnörkel-Schnarkel-Zeichen aus Korb 2«. Nehmen wir nun an, daß irgendwelche anderen chinesischen Symbole in das Zimmer gereicht werden, und daß Ihnen noch zusätzliche Regeln dafür gegeben werden, welche chinesischen Symbole jeweils aus dem Zimmer herauszureichen sind. Die hereingereichten Symbole werden von den Leuten draußen ›Fragen‹ genannt, und die Symbole, die Sie dann aus dem Zimmer herausreichen, ›Antworten‹ – aber dies geschieht ohne Ihr Wissen. Nehmen wir außerdem an, daß die Programme so trefflich und Ihre Ausführung so brav sind, daß Ihre Antworten sich schon bald nicht mehr von denen eines chinesischen Muttersprachlers unterscheiden lassen. [...] In so einer Lage, wie ich sie gerade beschrieben habe, könnten Sie einfach dadurch, was Sie mit

den formalen Symbolen anstellen, kein bißchen Chinesisch lernen.

Der Witz der Geschichte ist nun schlicht folgender: weil Sie ein formales Computerprogramm ausführen, verhalten Sie sich aus der Sicht eines Außenstehenden so, als verstünden Sie Chinesisch – und dennoch verstehen Sie nicht ein Wort Chinesisch. Wenn aber die Ausführung eines passenden Computerprogramms in *Ihrem* Fall nicht ausreicht, um Chinesisch zu verstehen, dann reicht das auch bei keinem anderen digitalen Computer aus.«

(John Searle: *Geist, Hirn und Wissenschaft*, übers. von Harvey P. Gavagai, Frankfurt a. M.: Suhrkamp, 1984, S. 31 f.)

Das Szenario

Das kontrafaktische Szenario ist so gebaut, dass es eine Position im Inneren eines Computers als einer Symbolverarbeitungs-maschine nachbildet. Aus dieser Position stellt sich der Computer folgendermaßen dar: Er wird immer wieder mit bestimmten Eingaben gespeist. Diese Eingaben werden von dem Computer in bestimmter Weise verarbeitet. Dies schlägt sich in bestimmten Ausgaben des Computers nieder. Die Verarbeitung basiert dabei auf einem Computerprogramm. Das Computerprogramm ist so gestaltet, dass es jeweils auf bestimmte Eingaben hin bestimmte Prozesse in Gang setzt, die zu den Ausgaben führen. Das kontrafaktische Szenario des chinesischen Zimmers illustriert nun, wie es wäre, wenn ein Mensch die Arbeit des Computers ausführte. Dieser Mensch müsste sich mit all den Verarbeitungsregeln vertraut machen können, die die Programme des Computers ausmachen. Zu diesem Zweck wird ihm ein Regelbuch gegeben. Das Regelbuch gibt vor, was der Mensch bei einer bestimmten Eingabe zu tun hat. Dazu ist es erforderlich, dass er die Eingabe erkennt und dass er die entsprechenden Regeln versteht und anwenden kann. Es ist aber nicht erforder-

lich, dass er die Eingabe versteht. Genauso wenig ist es erforderlich, dass er die Ausgabe versteht, die er auf die Eingabe hin auf Basis der Regeln generiert. Wer in einem Computer – in dem chinesischen Zimmer – sitzt, muss nur wissen, welche Zeichen er aufgrund der Regeln zu greifen und nach draußen zu reichen hat.

Zur Interpretation

Das Gedankenexperiment vom chinesischen Zimmer verfolgt das Ziel, eine bestimmte These zu verteidigen. Verteidigt werden soll die These, dass geistige Zustände sich nicht als solche verstehen lassen, die aufgrund einer computeranalogen Struktur das sind, was sie sind. Entscheidend für dieses Ergebnis ist die These, dass geistige Zustände einen bestimmten Gehalt haben. Das Durchspielen des Szenarios vom chinesischen Zimmer soll nun deutlich machen, dass die Symbolverarbeitungsregeln diese bestimmten Gehalte nicht verständlich machen. Die Symbolverarbeitung kann auch jemand leisten, der weder die Eingaben noch die Ausgaben versteht.

Diese Überlegungen lassen sich gut mit den Begriffen der Syntax und der Semantik fassen: Syntax ist die Lehre von der rechten Formung und Kombination von Zeichen. Semantik ist die Lehre der Bedeutung von Zeichen. Die These des Gedankenexperiments kann man in diesen Begriffen folgendermaßen artikulieren: Die Semantik geistiger Zustände lässt sich nicht (vollständig) durch die syntaktischen Strukturen erklären, in denen diese Zustände stehen. Das Szenario des Gedankenexperiments hat so gesehen einen illustrativen Charakter. Die These, die mittels seiner gestützt werden soll, lässt sich auch unabhängig von ihm argumentativ verteidigen. Ein entsprechendes Argument kann unter anderem folgendermaßen aussehen:

- (1) Computationale Systeme sind Systeme, die Symbole allein syntaktisch verarbeiten.

- (2) Geistige Zustände haben einen semantischen Gehalt.
- (3) Semantischer Gehalt kann nicht allein unter Rekurs auf syntaktische Verarbeitung erklärt werden.
- (4) Also: Geistige Zustände lassen sich nicht ausschließlich unter Rekurs auf computationale Systeme verständlich machen.

Searles Gedankenexperiment kann insbesondere so verstanden werden, dass es die Prämisse (3) verständlich macht. Unter der Voraussetzung von (1) und (2) folgt dann Searles Zielthese. Sofern man die Prämisse (3) auch anderweitig plausibilisieren kann, braucht man das Gedankenexperiment also gar nicht.

Searles Gedankenexperiment geht aber nicht bloß in einer solchen Argumentation auf. Es bringt vielmehr ein kontrafaktisches Szenario hervor, das sich grundsätzlich dafür eignet, um über das Verhältnis von Syntax und Semantik nachzudenken. Damit ist das Gedankenexperiment auch für die Sprachphilosophie relevant: Lässt sich die Bedeutung von sprachlichen Zeichen unter Rekurs auf die syntaktischen Eigenschaften dieser Zeichen erklären? Oder vorsichtiger gefragt: In welchem Zusammenhang stehen die syntaktischen Eigenschaften von sprachlichen Zeichen mit ihrer Bedeutung? Das Gedankenexperiment vom chinesischen Zimmer lädt dazu ein, diesen Zusammenhang zu erkunden.

Fragen und Probleme

Das Gedankenexperiment vom chinesischen Zimmer hat besonders zwei Einwände hervorgerufen. Erstens wird kritisiert, dass das Experiment sich mit einer Position kritisch auseinandersetzt, die gänzlich unplausibel ist. Niemand geht davon aus, dass die Gehalte geistiger Zustände sich einfach dadurch erklären lassen, dass man den Geist als ein computeranaloges System begreift. Auch diejenigen, die die Analogie zwischen Geist und Computer für aufschlussreich halten, sind nicht der Mei-

nung, dass man mit dieser Analogie alles über den Geist und über den Gehalt geistiger Zustände gesagt hat. So gesehen arbeitet sich das Gedankenexperiment an einem Strohmännchen ab. Es klärt nicht viel, dies zu tun. Zweitens – und wichtiger – wird dem Gedankenexperiment vom chinesischen Zimmer entgegengehalten, dass dieses Experiment nicht das Ergebnis stützt, das es eigentlich stützen soll. Stellen wir uns vor, dass der Bewohner des chinesischen Zimmers alle Regeln auswendig lernt und dass er irgendwann dieses Zimmer verlässt. Wenn er dann auf Menschen trifft, die Chinesisch sprechen, wird er sich zweifelsohne perfekt mit ihnen unterhalten können. Er kennt ja, so die Annahme, alle Regeln, nach denen man auf chinesische Zeichen mit chinesischen Zeichen reagiert. Würden wir also sagen, dass der Bewohner des chinesischen Zimmers kein Chinesisch versteht? Zeigt das Gedankenexperiment, wenn man es fortspinnt, nicht gerade, dass sprachliche Bedeutung wesentlich durch die Zusammenhänge bestimmt ist, in denen sprachliche Zeichen mit anderen Zeichen stehen, die ihnen z. B. in einem Gespräch vorangehen oder folgen können?

Zum Weiterlesen:

Beckermann, Ansgar: »Sprachverstehende Maschinen«, in: *Erkenntnis* 28 (1988), S. 65–85.

Nimtz, Christian: »Das chinesische Zimmer«, in: Rolf W. Puster (Hrsg.): *Klassische Argumentationen der Philosophie*, Paderborn 2012.

Preston, John / Bishop, Mark (Hrsg.): *Views into the Chinese Room. New Essays on Searle and Artificial Intelligence*, Oxford 2002.

G.W.B.