# ETHIK
# UND
# SOZIALWISSENSCHAFTEN

Streitforum für Erwägungskultur

Herausgegeben von Frank Benseler, Bettina Blanck, Rainer Greshoff, Werner Loh

EuS 1 (1990) Heft 1

# Die Gödeltheoreme und das Problem Künstlicher Intelligenz

Dieter Wandschneider

Zusammenfassung: Gödels Unvollständigkeitstheoreme sind immer wieder als Beweis für die Existenz prinzipieller Leistungsgrenzen formaler und damit grundsätzlich auch technischer Systeme gedeutet worden. Sie scheinen, so gesehen, auch zu implizieren, daß die dem Computer zugrundeliegende Logik von grundsätzlich anderer Art ist als die des Denkens, das solche Grenzen zu entdecken und zu beweisen vermag. Im folgenden wird gezeigt, daß diese Auffassung, die für das Problem Künstlicher Intelligenz gravierende Konsequenzen hätte, unhaltbar ist. Wesentlich ist, daß durch Gödels Verfahren semantische Strukturen im System der Arithmetik etabliert werden, und das Unvollständigkeitsproblem ergibt sich danach allein aus der so ermöglichten Selbstreferenz des von Gödel konstruierten Ausdrucks, nicht aus einem grundsätzlichen Unterschied der logischen Möglichkeiten von Denken und Computer.

Summary: Again and again Gödel's incompleteness theorems have been interpreted as a proof that there exist limits for formal systems on grounds of principle and that thereby such limits also exist in principle for technical systems. From this point of view they also seem to imply that the logic underlying the computer is fundamentally different from that of the thinking, which is able to discover and prove such limits. In what follows we will demonstrate that the thesis, which would have serious consequences for the problem of artifical intelligence, is untenable. The vital point can here be seen in the introduction of semantic structures into the system of arithmetic in consequence of Gödel's method. Then the problem of incompleteness solely results from the self-reference of Gödel's formula which is made possible in this way and not from a difference in principle between the logical possibilities of the thinking on the one hand and those of the computer on the other hand.

((1)) Nicht nur unser Leben ist heute von Grund auf durch die Technik geprägt, sondern auch unser Weltbild und Wirklichkeitsbegriff. Unser Realitätsverständnis ist weithin über technische Modelle vermittelt. Nicht nur, wie Heidegger vordem beklagt hatte, daß 'der Rheinstrom' in dieser Perspektive auf seinen rein energetischen Aspekt verkürzt erscheint[1] - inzwischen denken wir uns selbst durch und durch technisch: den menschlichen Leib, unbeschadet 'psychosomatischer' Vorbehalte, als biochemische Maschine und das Gehirn als Computer, dem damit umgekehrt denkanaloge Leistungen zugetraut werden. Vor allem diese letztere Idee einer 'Denkmaschine' ist sicher keine moderne Erfindung, sondern hat die neuzeitliche Phantasie immer wieder beschäftigt - erinnert sei an Pascal und Leibniz -, doch heute zeichnen sich früher nicht zu erahnende Realisierungsmöglichkeiten ab. Diese Situation stellt für die Philosophie zweifellos eine Herausforderung dar. Das Ausmaß des technisch Möglichen trifft uns in gewissem Sinne unvorbereitet - man denke nur an die Gentechnik - und verlangt daher um so dringlicher auch philosophische Aufarbeitung, Klärung und Orientierung.

((2)) Das Interesse an Denkmaschinen erscheint in philosophischer Perspektive als durchaus ambivalent: In ihm kommt keineswegs nur das Motiv technischer Effizienzsteigerung zum Ausdruck, sondern - J.v.Neumann und A.M.Turing sind diesbezüglich Kronzeugen - ebenso der Wunsch, die Denkleistung selbst noch im technischen Modell zu objektivieren. Es ist die philosophische Idee der Selbsteinholung des Denkens in der Weise, daß der bewußtlosen Materie gewissermaßen das Denken beigebracht wird - eine erst heute realistisch erscheinende Möglichkeit und sicher die durchschlagendste Form 'technikphilosophischer' Demonstration. Die moderne Technik, das ist bei aller berechtigten Kritik an ihr auch in Rechnung zu stellen, erschließt so zugleich qualitativ neue Möglichkeiten menschlichen Selbstverständnisses.

((3)) Im folgenden soll das Problem technischer Selbstobjektivierbarkeit des Denkens nur in sehr grundsätzlicher Hinsicht erörtert werden: Es soll der Frage nachgegangen werden, ob für Computer prinzipielle Leistungsgrenzen bezüglich ihrer logischen Möglichkeiten anzunehmen sind oder umgekehrt: ob die dem Denken zugrundeliegende Logik möglicherweise nicht objektivierbar, d.h. nicht in effektive Computeroperationen übersetzbar ist. Diese Problemstellung verweist auf eine neuere mathematische Forschungsrichtung, die Theorie der sogenannten rekursiven, d.h. grundsätzlich effektiv berechenbaren Funktionen, die für die Frage formaler Beweisbarkeit und damit auch Objektivierbarkeit logischer Beziehungen zentral ist. Von besonderem Interesse sind in diesem Zusammenhang vor allem

zwei von K.Gödel schon 1931 bewiesene Theoreme[2], die - damals völlig überraschend - Grenzen formaler Beweisbarkeit markieren, durch welche die Möglichkeiten formaler und damit grundsätzlich auch maschineller Systeme wesentlich eingeschränkt erscheinen, nicht hingegen das Denken des Logikers, das ja einen Beweis für die Existenz solcher Beweisbarkeitsgrenzen liefert. Gödels Theoreme scheinen so zu implizieren, daß die dem Computer zugrundeliegende Logik von prinzipiell anderer Art ist als die des Denkens oder schärfer: daß jede Form artifizieller Intelligenz menschlicher Intelligenz notwendig unterlegen ist. So jedenfalls ist Gödels Resultat immer wieder gedeutet worden; es wird hier darum gehen, in diesem Punkt zu einer Klärung zu kommen. Das Ergebnis wird negativ sein, d.h. es wird gezeigt werden, daß durch Gödels Argumentation keine Grenzen artifizieller Intelligenz markiert sind und insbesondere eine Zweiweltentheorie der Logik von daher entschieden nicht zu begründen ist. Im folgenden soll der durch Gödels Theoreme charakterisierte Tatbestand zunächst näher erläutert, zweitens auf die ihm zugrundeliegende Argumentationsstruktur hin analysiert und drittens hinsichtlich seiner Bedeutung für das Problem Künstlicher Intelligenz beurteilt werden.

## I

((4)) Zunächst zum Technischen der Gödelschen Konstruktion.[3] Vorausgesetzt ist ein logisches System von der Art, daß in ihm die elementare Arithmetik entwickelt werden kann. Davon zu unterscheiden ist zunächst ein diesem System zugeordnetes Metasystem, in dem metatheoretische Aussagen über das System formulierbar sind, z.B. daß ein Ausdruck 'wohlgeformt', 'beweisbar' usw. ist. Vermittels eines Kunstgriffs gelang es Gödel nun, einen Teil der Metatheorie des Systems im System selbst zu formalisieren. Dieses so erweiterte Logiksystem soll im folgenden als das System S bezeichnet werden. Gödels Verfahren, heute als Arithmetisierung oder auch als Gödelisierung bezeichnet, besteht bekanntlich darin, daß den Grundzeichen des Systems, den aus diesen gebildeten Formeln sowie den Folgen von Formeln eineindeutig natürliche Zahlen - 'Gödelzahlen' genannt - zugeordnet werden. Wie dies geschieht, ist hier nicht wesentlich. Daß die Zahlen, die den Zeichen, Formeln und Formelfolgen des Systems zugeordnet sind, ihrerseits erst mit Hilfe dieser im System verfügbaren Ausdrücken definiert worden sind, dürfte in diesem Zusammenhang unerheblich sein; es handelt sich dabei um eine 'Abbildung' des schon etablierten Systems in sich. Entscheidend für das Folgende ist, daß es auf diese Weise möglich wird, einen Teil der Metatheorie des Systems in dieses selbst zu integrieren, konkret: Durch die Arithmetisierung der verschiedenen Ausdrücke können bestimmte Klassen derselben, z.B. die Klasse der beweisbaren Formeln, durch rein arithmetische Beziehungen charakterisiert werden. Die Gödelzahlen der im System beweisbaren Theoreme gehören dementsprechend einer wohlbestimmten Zahlklasse T an. Der arithmetischen Aussage, daß eine Zahl z zur Zahlklasse T gehört, korrespondiert also die metatheoretische Aussage, daß die z zugeordnete Formel ein Theorem ist. - Übrigens beruht diese Möglichkeit, einen Teil der Metatheorie in das System der Arithmetik zu integrieren, offenbar darauf, daß die Ausdrücke eines formalen Systems Eigenschaften besitzen, die selbst zahlenmäßig faßbar sind, z.B. ihre endliche Abzählbarkeit, wohlbestimmte Reihenfolge usw., Eigenschaften also, die für formale Systeme überhaupt charakteristisch zu sein scheinen.

((5)) Von besonderer Bedeutung ist nun der Umstand, daß mit der Arithmetisierung der Ausdrücke eine semantische Ebene im System S etabliert ist. Denn jede Gödelzahl hat ja aufgrund dieser Zuordnung eine Interpretation, d.h. sie referiert auf den ihr zugeordneten Ausdruck: ein Grundzeichen, eine Formel, insbesondere etwa auch eine beweisbare Formel usw.[4] Soweit ich sehe, bleibt dieser strukturell nicht folgenlose Begleitumstand der Arithmetisierung bei Gödel selbst und in der Literatur merkwürdig unterbelichtet. Die Darstellungen vermitteln häufig den Eindruck, als sei der formale Charakter der Arithmetik das Entscheidende und die Arithmetisierung lediglich eine Übersetzung metasprachlicher Ausdrücke in Zahlenrelationen.[5] Doch dabei gerät aus dem Blick, daß umgekehrt diejenigen Zahlen, die Gödelzahlen sind, nun nicht mehr nur Zahlen sind, sondern außerdem eine Interpretation besitzen und damit nicht mehr nur in formalen, sondern auch in semantischen Relationen stehen. Man sagt auch, daß sie durch ihre formalen Eigenschaften zugleich semantische Beziehungen 'widerspiegeln'. Die ganze Konstruktion ist dadurch schon im Ansatz semantisch orientiert. Solange dies nicht berücksichtigt wird, ist ein Verständnis des den Gödeltheoremen zugrundeliegenden Prinzips m.E. nicht möglich.

((6)) Was besagen diese nun konkret? Mit Hilfe des Arithmetisierungsverfahrens ist es Gödel gelungen, einen Ausdruck - nennen wir ihn G[6] - zu konstruieren, der in arithmetischer Kodierung seine eigene Unbeweisbarkeit aussagt. Gödels Ausdruck G ist genauer eine arithmetische Relation zwischen Gödelzahlen, und entsprechend der für G (per Arithmetisierung) festgelegten Interpretation repräsentiert G die Aussage, daß keine Zahl die Gödelzahl eines Beweises für G selbst ist. Das von Gödel in diesem Zusammenhang bewiesene sogenannte Unvollständigkeitstheorem besagt nun, daß dieser Ausdruck G, sofern das zugrundegelegte formale System widerspruchsfrei ist, prinzipiell nicht beweisbar ist und eben deswegen (denn genau dies drückt er ja aus) eine wahre Aussage darstellt: In diesem Sinne also ist das System unvollständig, d.h. es enthält einen Satz, der zwar wahr und dennoch grundsätzlich unbeweisbar ist.[7] Das hat weiter zur Folge, so Gödels 2.Theorem, daß die in S formalisierte Aussage, das System S sei widerspruchsfrei, prinzipiell nicht innerhalb des Systems S selbst bewiesen werden kann. - Es versteht sich in diesem Zusammenhang von selbst, daß der Beweisbegriff immer relativ auf ein bestimmtes System und die in diesem installierte Arithmetisierung zu verstehen ist, so daß die Prädikate 'beweisbar' und 'unbeweisbar' im folgenden stets durch den Zusatz 'im Rahmen des Systems S' ergänzt zu denken sind.

((7)) Nun wäre zu erwarten, daß die Vollständigkeit oder Unvollständigkeit eines Systems eine Frage der im System verfügbaren deduktiven Möglichkeiten ist, d.h. stärkere formale Mittel sollten auch stärkere Resultate ermöglichen. Es mußte daher im höchsten Maße überraschen, daß das Unvollständigkeitstheorem demgegenüber eine äußerste Grenze formaler Beweisbarkeit markiert: Der von Gödel konstruierte Ausdruck G ist prinzipiell, d.h. unabhängig von allen Beweismodalitäten, unbeweisbar und gleichwohl wahr[8] - ein höchst merkwürdiges Resultat, das zudem das für die moderne Logik so charakteristische Ideal zu entwerten drohte, das im Hilbertschen Formalismus prägnanten Ausdruck gefunden hatte. Hilbert soll denn auch - einem von H.Meschkowski wiedergegebenen Zitat zufolge[9] - "etwas ärgerlich" gewesen sein, als er durch Paul Bernays von Gödels Resultaten erfuhr. Quine spricht in diesem Zusammenhang gar von der "zweiten großen modernen Grundlagenkrise der Mathematik".[10] In der Tat sah sich die mathematische Logik damit vor die Frage gestellt, ob der Formalisierbarkeit und damit Objektivierbarkeit logischer Verhältnisse nicht womöglich prinzipielle, unüberschreitbare Grenzen gezogen sind.[11]

((8)) Nun kann, wie etwa durch die Untersuchungen von A.Church und A.M.Turing deutlich geworden ist, eine Maschine in gewissem Sinne als eine Realisierung formaler Systeme verstanden werden.[12] Lassen sich aus Gödels Unvollständigkeitstheorem, so fragt sich, somit auch Rückschlüsse auf Möglichkeiten und Grenzen logischer Strukturen, wie sie in Maschinen realisierbar sind, ziehen? Wird es nie eine Maschine geben können, die das Denken des Logikers nachzubilden vermag? Das wäre ein Tatbestand, der auch technikphilosophisch und insbesondere für das ganze Gebiet der Künstlichen Intelligenz gravierende Konsequenzen hätte. Wird hier gewissermaßen "die Achillesferse der kybernetischen Maschine" sichtbar, wie J.R.Lucas meint[13], in dem Sinne nämlich, daß eine Maschine, die das System der elementaren Arithmetik enthält, stets einen Ausdruck bilden kann, der für sie selbst unbeweisbar, für uns hingegen als wahr einsehbar ist? Lucas hält es von daher für ausgemacht, daß "keine Maschine ein vollständiges oder adäquates Modell des Geistes sein kann" (44), d.h. "wir können keine Maschine bauen, die geistartiges Verhalten in jeder Hinsicht zu simulieren vermag". "Wir können niemals, nicht einmal im Prinzip, ein technisches Modell des Geistes besitzen" (47). Ähnliche Äußerungen finden sich etwa bei E.Nagel und J.R.Newman[14], G.Frey[15], H.Meschkowski[16] und vielen anderen[17], bis hin zu Popularisierungen im Stile D.R.Hofstadters[18] oder gar 'postmoderner' Inanspruchnahme bei J.-F.Lyotard[19] - die Gödeltheoreme sind inzwischen auch so etwas wie ein Mythos.

## II

((9)) Um bezüglich der gestellten Frage zu einer Klärung zu kommen, soll im folgenden versucht werden, den eigentlichen Grund für den von Gödel entdeckten Sachverhalt auszumachen. Entscheidende Bedeutung, das ist hier die These, kommt dabei dem semantischen Aspekt des Gödelschen Ausdrucks G zu und insbesondere seinem selbstreferentiellen Charakter.

((10)) Eine erste Hinsicht, in der sich die Selbstreferentialität von G geltend macht, ist die für selbstbezügliche Aussagen charakteristische Kopplung von Satzaussage und Satzeigenschaft: Gödels Satz G entspricht ja der Aussage 'Dieser (hier angeschriebene) Satz ist unbeweisbar'. Aufgrund dieser Selbstreferentialität kann es nur zwei Möglichkeiten geben: G kann entweder die Eigenschaft der Unbeweisbarkeit haben und damit wahr sein, oder G ist beweisbar und dann, G's Bedeutung entsprechend, falsch. Diese Kopplung von Wahrheit und Unbeweisbarkeit einerseits und Falschheit und Beweisbarkeit andererseits beruht allein auf der semantischen Selbstreferentialität von G.[20]

((11)) Ist das betrachtete System nun insbesondere ein semantisch korrektes System, d.h. ein solches, in dem alle beweisbaren Sätze stets wahre Sätze sind, so scheidet die letztgenannte der beiden Alternativen aus; denn G kann dann nicht beweisbar und zugleich falsch sein, und es bleibt somit nur die andere Möglichkeit, daß G unbeweisbar und wahr ist.

((12)) Diese einfache Überlegung zeigt, daß ein Ausdruck wie G, der seine eigene Unbeweisbarkeit aussagt, allein aufgrund seiner eigentümlichen Referenzbeziehung unbeweisbar und zugleich wahr sein muß, sofern das System korrekt ist, d.h. keine falschen Sätze zu beweisen gestattet.[21] Mit dieser Einsicht wird gleichsam ein erster Blick hinter die Kulissen des Unvollständigkeitstheorems möglich: Die sich hier ergebende Unbeweisbarkeit und damit Wahrheit des Satzes G ist keineswegs durch eine Schwäche der im System verfügbaren deduktiven Möglichkeiten bedingt, sondern - bei vorausgesetzter Korrektheit des Systems - allein durch die selbstreferentielle Struktur von G selbst. Die Selbstreferentialität von G ist somit durchaus kein nebensächlicher Umstand, wie Äußerungen von Gödel selber oder auch Stegmüller vermuten lassen könnten.[22]

((13)) Nun macht die hier durchgeführte Argumentation mit der Annahme semantisch korrekter Systeme allerdings eine stärkere Voraussetzung als Gödels eigener Beweis (bzw. der entsprechende Beweis von B.Rosser[23] - zwischen beiden Versionen soll hier nicht unterschieden werden); dieser benötigt nur die schwächere Bedingung formaler Widerspruchsfreiheit.[24] Daß dieser Beweisgang tatsächlich noch mehr auf der selbstreferentiellen Struktur des Satzes G beruht als die zuvor entwickelte Argumentation, soll im folgenden kurz skizziert werden: Der Satz G, der seine eigene Unbeweisbarkeit formuliert, so hatte sich zunächst gezeigt, ist aufgrund seiner Selbstreferentialität entweder unbeweisbar und wahr oder beweisbar und falsch. Die letztgenannte Möglichkeit scheidet, wie dargelegt, aus, wenn das System semantisch korrekt ist. Sie scheidet aber nach der Argumentation Gödels auch aus, wenn die schwächere Bedingung formaler Widerspruchsfreiheit erfüllt ist, wobei, wie sich zeigen wird, die Selbstreferentialität von G auch in diesen

Nachweis entscheidend mit eingeht. Gödel schließt wie folgt:

((14)) 1.Schritt: Wäre G beweisbar, so wäre die Metaaussage 'G ist beweisbar' gültig und, da (wie Gödel zeigt[25]) rekursiv durch G fundiert und im System S 'formal ausdrückbar'[26], ihrerseits beweisbar. Mit der Beweisbarkeit von G wäre somit auch die Metaaussage 'G ist beweisbar' beweisbar, die ihrerseits aber - und das ist hier die Pointe - gerade die Negation von G ist, mit anderen Worten: Wäre G beweisbar, so wäre auch Non-G beweisbar, ein Widerspruch also. Die Annahme der Beweisbarkeit von G ist damit qua reductio ad absurdum widerlegt, sofern das System widerspruchsfrei ist. In diesem Fall bleibt also auch hier nur die Möglichkeit, daß der Satz G unbeweisbar und damit, im Hinblick auf seine Aussage, freilich auch wahr ist.

((15)) 2.Schritt: Gödel hat weiter gezeigt, daß G darüberhinaus ein im System S unentscheidbarer Satz ist in dem Sinne, daß mit der eben bewiesenen Unbeweisbarkeit von G auch die Unbeweisbarkeit von Non-G impliziert ist. Dadurch ist übrigens ein Konzept formaler (syntaktischer) Unvollständigkeit[27] charakterisiert, das, im Unterschied zu dem früher angegebenen semantischen Unvollständigkeitsbegriff ('wahr, aber nicht beweisbar'), ohne das (in S selbst nicht definierte) Prädikat 'wahr' auskommt. - Der Beweisgang kann auch hier nur grob angedeutet werden: Das vorherige Resultat, daß G nicht beweisbar ist, impliziert für jede einzelne Zahl die formal beweisbare Aussage, daß keine derselben die Gödelzahl eines Beweises für G ist, mithin unbeschränkt viele formal beweisbare Einzelaussagen dieses Inhalts. Daraus folgt freilich noch nicht die formale Beweisbarkeit der generellen Aussage 'G ist überhaupt unbeweisbar' - und das wäre die Aussage G selbst -, d.h. es folgt daraus nicht (was desaströs wäre[28]) die Beweisbarkeit von G. Obwohl also G selbst nicht formal beweisbar ist, kann andererseits doch bereits aufgrund der beweisbaren Einzelaussagen ausgeschlossen werden, daß die Negation von G beweisbar ist: Ein solcher Schluß wird möglich, falls für das System die Bedingung der sogenannten Omega-Widerspruchsfreiheit erfüllt ist (die also stärker als die einfacher Widerspruchsfreiheit ist).[29]

((16)) Gödels 2.Theorem schließlich hängt auch wesentlich mit der Unbeweisbarkeit von G zusammen: Das im 1. Schritt erzielte Resultat bedeutet nach Gödel weiter die formale Beweisbarkeit einer Implikation der Form: 'Die Widerspruchsfreiheit des Systems impliziert G'.[30] Wäre folglich die Aussage der Widerspruchsfreiheit selbst, d.h. das Vorderglied der Implikation, innerhalb des Systems beweisbar, so müßte auch das Hinterglied, also G, innerhalb des Systems beweisbar sein, im Widerspruch zum Ergebnis des Unvollständigkeitstheorems, daß G innerhalb des Systems (sofern dieses widerspruchsfrei ist) nicht beweisbar sein kann, mit anderen Worten: Aufgrund der Unbeweisbarkeit von G ist es auch unmöglich, daß der Satz von S, der die Widerspruchsfreiheit von S formuliert, in diesem System selbst beweisbar ist.

((17)) Angesichts dieser außerordentlich irritierenden Resultate Gödels soll im folgenden versucht werden, so etwas wie eine Erklärung dafür zu finden: Wesentlich ist offenbar, daß (im 1. Schritt) mit der Annahme der Beweisbarkeit von G auch die Beweisbarkeit der Metaaussage 'G ist beweisbar' impliziert ist, die ihrerseits aber gerade die Negation des Satzes G ist und dadurch zum Widerspruch führt. Voraussetzung dieser reductio ad absurdum ist aber, daß sowohl die Metaaussage 'G ist beweisbar' als auch G selbst eine Beweisbarkeitsaussage über G machen, mit anderen Worten: Wesentlich ist die Selbstreferentialität von G. Nur so können beide Sätze - G und 'G ist beweisbar' - über denselben Satz G aussagen, in diesem Fall sogar Entgegengesetztes.[31] Aber auch der 2.Beweisschritt wäre ohne die Selbstreferentialität von G nicht möglich: Der zuvor bewiesene Sachverhalt, daß G unbeweisbar ist, führt hier über die für jede einzelne Zahl beweisbare Einzelaussage, daß diese Zahl nicht Gödelzahl eines Beweises für G ist, 'fast' zum Beweis der generellen Aussage 'G ist überhaupt unbeweisbar', und das heißt eben G, woraus dann über die Omega-Widerspruchsfreiheit auf die Nichtbeweisbarkeit der Negation von G geschlossen werden kann. Auch hier ist also wesentlich, daß die Aussage 'G ist unbeweisbar' mit G selbst identisch, G also selbstreferentiell ist. Und schließlich hängt auch das für die Metamathematik besonders irritierende 2.Theorem Gödels, wonach die S-Aussage der Widerspruchsfreiheit des Systems nicht in diesem selbst bewiesen werden kann, wie dargelegt, an der innersystemischen Unbeweisbarkeit von G, die ihrerseits auf der Selbstreferentialität von G beruht.

((18)) Mit der Selbstreferentialität von G ist daher m.E. die entscheidende Bedingung der Gödelschen Argumentation erfaßt. Gödels Problem betrifft solchermaßen, landläufigen Vorstellungen zum Trotz, keinen wesentlich beweistheoretischen, sondern im Grunde einen semantischen Tatbestand. Dies würde im übrigen die - schon konstatierte - weitgehende Unabhängigkeit der Gödelschen Resultate von der speziellen Wahl der Axiome und des Beweisbegriffs in S erklären - ein rein beweistheoretisch sonst unverständlicher Sachverhalt. Gödels metamathematische Verdienste sollen dabei nicht verkleinert werden. Vor allem sein ingeniöses Verfahren zur arithmetischen Repräsentation bestimmter metatheoretischer Prädikate und der Nachweis ihres primitiv-rekursiven Charakters stellt zweifellos eine beweistheoretische Glanzleistung dar.[32] Aber der damit auch verbundene exorbitante technische Aufwand dürfte eher zur Verschleierung als zur Klärung des zugrundeliegenden semantischen Sachverhalts beigetragen haben.

((19)) Dem entspricht, daß insbesondere die mit der Selbstreferentialität von G involvierten Eigentümlichkeiten des Satzes verdeckt blieben. Der hier einschlägige mathematische Begriff der 'Diagonalisierung' benennt nur die formaltechnische Operation, vermittels derer bei einem arithmetischen Funktionsausdruck ein Selbstbezug herstellbar wird, der in semantischer Perspektive freilich folgenreiche Konsequenzen hat: g sei etwa die Gödelzahl der Formel G, die ihre eigene Unbeweisbarkeit aussagt, in arithmetisierter Form also die Aussage 'g ist kein Element von T' ist (wobei T

wieder die Klasse der Gödelzahlen jener Formeln bezeichnet, die Theoreme sind). g bedeutet, mit anderen Worten, im Grunde die Aussage 'g ist kein Element von T', deren Sinn seinerseits wiederum von g abhängig ist. Hier liegt also, worauf v.Kutschera (1964, 88 ff) aufmerksam gemacht hat, eine semantische Zirkularität vor, d.h. der Satz G ist, indem er eine Aussage über sich selbst macht, semantisch unfundiert.[33] v.Kutschera nennt ihn darum "bedeutungslos" (im Sinne von 'gehaltleer') und schließt daraus, daß Gödels Beweis unter semantischem Aspekt eine Petitio principii enthält (92).

((20)) Das ändert indes nichts daran, daß der Satz G formal möglich ist und, gerade aufgrund seiner Unfundiertheit, eine semantisch interessante Struktur darstellt - einen Satz nämlich, der (aufgrund der durch G.Gentzen gesicherten Widerspruchsfreiheit der Arithmetik) wahr, aber, da unfundiert, von unbestimmtem Gehalt ist. Doch auch unter allgemein sprachtheoretischem Gesichtspunkt ist bemerkenswert, daß Objekt- und Metasprache im Sinne des Gödelschen Verfahrens vereinigt werden können und auf diese Weise sprachliche Selbstbezüge konstruierbar sind, ohne daß damit Antinomien in das System eingeschleust werden (obwohl Gödels Konstruktion, worauf dieser selbst schon hinweist, in Analogie zur Richardschen Antinomie oder auch zur Lügnerantinomie konzipiert ist[34]). Dies zeigt, daß Selbstreferentialität sowie die Verschmelzung von Objekt-und Metasprache, sprachliche Möglichkeiten also, wie sie für natürliche Sprachen charakteristisch sind, in künstlichen Sprachsystemen sehr wohl antinomienfrei nachgebildet werden können. Das ist zweifellos eine sprachtheoretisch eminent wichtige Einsicht, die wir ebenfalls der Gödelschen Konstruktion verdanken. - Soviel zur Analyse der Gödelschen Resultate, wobei gewisse Vereinfachungen in dem hier vorgegebenen Rahmen unvermeidlich waren.

## III

((21)) Was folgt nun aus diesen Überlegungen für die eingangs gestellte Frage bezüglich des Verhältnisses von Denken und Computer? Es geht hier wohlgemerkt weder um spezifisch automatentheoretische Probleme noch um mathematische Probleme effektiver Berechenbarkeit, wie sie etwa im Zusammenhang mit der von A.Church formulierten Hypothese ('Eine zahlentheoretische Funktion ist genau dann berechenbar, wenn sie rekursiv ist') diskutiert werden[35] noch gar um die avancierte Problemstellung: 'Sind bewußtseinsanaloge Maschinen möglich?'.[36] Das ist hier nicht zu untersuchen. Es ist vielmehr um eine Klärung der Frage zu tun, ob die Gödeltheoreme zu der Konsequenz nötigen, daß die Logik des Denkens und die der Maschinen von prinzipiell verschiedener Art ist.

((22)) Nun, an Gödels Resultat, das steht fest, ist nicht zu deuteln. Eine Maschine, deren System hinreichend ausdrucksreich ist, kann danach einen Ausdruck von der Art des Gödelschen Satzes G bilden, der für sie strikt unbeweisbar ist, während der Logiker ihn als wahr erweisen kann, sofern das System widerspruchsfrei ist. Er kann darüberhinaus - im Sinne des 2.Gödeltheorems - beweisen, daß die S-Aussage der Widerspruchsfreiheit von S nicht innerhalb des Systems selbst bewiesen werden kann. Noch einmal also die Frage: Besitzt der Logiker eine der Maschine überlegene Logik? Gibt es zwei grundsätzlich differente logische Welten - die der Maschine und die des Denkens?

((23)) Hierzu wäre zunächst und vor allem zu klären, in welchem logischen System sich eigentlich der Logiker selbst bei der Herleitung der Gödeltheoreme bewegt. Offenbar doch in dem von ihm untersuchten System S, mit dessen Ausdrücken und Schlußregeln er ja operiert. Alle für die Beweise benötigten deduktiven Mittel sind keine anderen als die im System S selbst verfügbaren Mittel, und das heißt: Das Denken des Logikers ist nicht erst durch Teilhabe an einer anderen, höheren als der in S formalisierten Logik zu solchen Beweisen befähigt, sondern bleibt dabei ganz innerhalb des durch S selbst vorgegebenen Rahmens. Eine mit dieser Logik ausgestattete Maschine hätte, mit anderen Worten, exakt die gleiche Beweispotenz wie der Logiker.

((24)) Für ein angemessenes Verständnis dieser Zusammenhänge ist es freilich wichtig, bei Beweisen im System S klar zu unterscheiden zwischen dem Beweis eines arithmetischen Sachverhalts und dem Beweis des diesen Sachverhalt formulierenden Ausdrucks. Eine solche Differenzierung mag prima vista befremden, denn: Kann ein Sachverhalt anders als in sprachlicher Formulierung diskutierbar sein?[37] Natürlich nicht. Überhaupt muß alles, was zu einem Beweis gehört - nicht nur sein Resultat, sondern auch die Beweisstrategie, die einzelnen Beweisschritte usw. - , irgendwie sprachlich mitgeteilt werden. Im Normalfall geschieht das aber in einer unthematischen Metasprache, die dem thematisierten formalen System nicht angehört. Diese hat reine Mitteilungsfunktion, ist also selber nicht Gegenstand der beweistheoretischen Untersuchung.

((25)) Im Fall der Gödelschen Konstruktion liegen die Verhältnisse dagegen insofern anders, als die Metatheorie des untersuchten Systems S hier zum Teil ebenfalls in S enthalten ist. Bestimmte metatheoretische Aussagen über formale Strukturen von S treten somit selber als formale Strukturen von S, deren Beweisbarkeit zu klären ist, in Erscheinung, konkret: Die beweistheoretische Metaaussage, die den Sachverhalt der formalen Unbeweisbarkeit von G im System S formuliert, ist selber ein formaler Ausdruck von S, nämlich G selbst. Beides - Ausdruck und Sachverhalt - kann wiederum in einer unthematischen, nicht S selbst angehörenden Metasprache unterschieden werden, wie es ja auch hier geschieht. In diesem Sinne ist festzustellen, daß Gödel den Sachverhalt der Unbeweisbarkeit von G mit Hilfe der in S gegebenen deduktiven Möglichkeiten beweist und in einer neutralen Metasprache mitteilt, während der in S formalisierte metatheoretische Satz, daß G unbeweisbar ist - und das ist ja der Ausdruck G selbst -, im System S (sofern es widerspruchsfrei ist) nicht beweisbar sein kann; denn der Beweis von G wäre gerade das

Gegenteil des tatsächlich bewiesenen Sachverhalts der Unbeweisbarkeit von G. Das System wäre in diesem Fall nicht mehr widerspruchsfrei.

((26)) Wenn somit der Sachverhalt der Unbeweisbarkeit von G verbalisiert wird (und nur so kann er diskutierbar sein), so wird, inhaltlich gesehen, die gleiche Aussage formuliert, die auch G macht, aber: Dies ist G in einer übersetzten Form; übersetzt in eine, wie schon gesagt, von S verschiedene Metasprache, während G dem System S selbst angehört, d.h. selber ein formaler Ausdruck von S ist.

((27)) Mit diesen Klärungen kann die zuvor formulierte Auffassung, wonach sich der Logiker beim Beweis der Gödeltheoreme ganz im Rahmen des Systems S bewegt, weiter präzisiert werden: Er beweist in diesem Rahmen, so ist jetzt deutlich, den Sachverhalt der Unbeweisbarkeit von G und teilt dies in einer neutralen, nicht S angehörenden Metasprache mit, die als solche aber natürlich nicht seine deduktiven Möglichkeiten in S verändert, d.h. der Rahmen der in S gegebenen Beweismöglichkeiten wird dabei nicht überschritten, und eine entsprechend ausgestattete Maschine (System S - einschließlich der Arithmetisierung metatheoretischer Ausdrücke - plus S-unabhängige Mitteilungsfunktion) würde exakt dasselbe leisten. Neben der beweistheoretisch neutralen Form, die Unbeweisbarkeit von G in einer von S verschiedenen Metasprache mitzuteilen, gibt es aber noch die in S formalisierte metasprachliche Aussage dieses Sachverhalts, nämlich G selber. Da es sich hierbei, wie gesagt, um eine Struktur von S selbst handelt, stellt sich für diese auch die Frage ihrer Beweisbarkeit in S, die, wie wir gesehen haben, eindeutig negativ zu beantworten ist, d.h. der Ausdruck G ist im System S auf keinen Fall beweisbar: weder für die Maschine noch für den Logiker. Auch in diesem Punkt ergibt sich also keine Differenz bezüglich Denken und Maschine. "Das Gödeltheorem", so bemerkt zu Recht M.Arbib, "stellt für uns selbst nicht weniger eine Einschränkung dar als für den Computer".[38] In ähnlichem Sinne äußert sich auch H.Putnam.[39] - Die Implikationsaussage 'Die Widerspruchsfreiheit des Systems impliziert G' schließlich ist, was für das 2.Gödeltheorem ja entscheidend ist, im System S selbst und damit auch für die Maschine formal beweisbar. In allen diesen Punkten ist also in der Tat gleiche Beweispotenz für Denken und Computer zu konstatieren.

((28)) Es mag vielleicht eingewendet werden, wesentlich sei nicht, welche Beweismöglichkeiten innerhalb des Systems S existieren oder nicht existieren, sondern vielmehr der Umstand, daß der (in einem widerspruchsfreien System) als unbeweisbar erweisbare Satz G vom Logiker dennoch als wahr erkennbar ist. Sind die Möglichkeiten der Maschine damit nicht überboten? Hierzu ist zunächst zu sagen, daß der Wahrheitserweis eines Satzes keineswegs überragende logische Fähigkeiten, sondern nur die Möglichkeit voraussetzt, den in ihm ausgedrückten Sachverhalt bestimmen und dessen Bestehen feststellen zu können. Ohne Frage sind dazu grundsätzlich (und schon längst) auch Maschinen in der Lage. Trotzdem ergeben sich im Zusammenhang mit dem Wahrheitsbegriff gewisse Komplikationen. Einem von Tarski bewiesenen Theorem zufolge ist dieser nämlich in S selbst nicht verfügbar; andernfalls wären, wie Tarski zeigt, zusammen mit der Möglichkeit, über Substitutionsoperationen selbstreferentielle Sätze zu bilden, die Bedingungen für das Auftreten der Wahrheitsantinomie ('Dieser Satz ist falsch') gegeben.[40] Der Grund dafür ist, allgemein gesagt, der, daß Objekt- und Metasprache im System S eben nicht mehr strikt getrennt sind, sondern, wie eingangs dargelegt, ein Teil der Metasprache via Arithmetisierung in die Objektsprache integriert ist. Mit dem Wahrheitsprädikat wird dann - jedenfalls im Normalfall[41] - die Wahrheitsantinomie möglich. Daß G wahr ist, kann infolgedessen nicht in S selbst, sondern nur in einer von S verschiedenen Metasprache formuliert werden, wie sie (s.o.) für die Darstellung der Beweisoperationen ohnehin benötigt wird.

((29)) Etwas anders verhält es sich im Fall des Widerspruchsfreiheitsbeweises für S. Denn der Satz 'S ist widerspruchsfrei' ist ja, wie schon im Zusammenhang mit dem 2.Gödeltheorem bemerkt, im System S sehr wohl formalisierbar, in dieser Form aber nicht beweisbar, weil sonst auch G beweisbar wäre - was nach dem 1.Gödeltheorem ausgeschlossen ist. Aber dafür sind, wie sich gezeigt hat, nicht beweistechnische, sondern rein semantische Gründe verantwortlich: Der Ausdruck G kann aufgrund seiner im System S realisierten Selbstreferentialität prinzipiell nicht beweisbar sein. Hiermit ist in der Tat eine Einschränkung bezeichnet, über die sich der Logiker, wie dargelegt, allerdings durch Übergang zu einer von S verschiedenen Metasprache hinwegsetzen kann. Auf der anderen Seite spricht nichts dagegen, daß derartiges auch einer geeigneten Maschine möglich ist. Wir sind zwar geneigt, die Maschine als ein fixiertes Gebilde zu betrachten, wie es dem herkömmlichen - im Grunde mechanistischen - Begriff des maschinellen oder auch formalen Systems entspricht. Durch eine solche begriffliche Festlegung wäre das Mensch-Maschine-Problem freilich schon vorentschieden. Die darin enthaltene Petitio ist nur so vermeidbar, daß der Begriff formaler bzw. maschineller Systeme nicht von vornherein auf rigide Systemstrukturen restringiert wird. Die Möglichkeit der Einbeziehung zusätzlicher Ausdrucksmittel ist durch Gödels Resultate jedenfalls in keiner Weise ausgeschlossen, sondern im Gegenteil vielmehr nahegelegt.

((30)) Man könnte daran denken, auch die neuen Ausdrucksmittel zu formalisieren, sie außerdem in die Arithmetisierung einzubeziehen und auf diese Weise vom System S zu einem erweiterten System S' überzugehen. Dadurch verschiebt sich aber das gesamte System der Gödelzahlen.[42] Doch in dem erweiterten System S' kann auch wieder ein selbstreferentieller 'Gödelscher' Ausdruck - dem jetzt die arithmetische Relation G' entsprechen möge - gebildet werden, d.h. durch Systemerweiterung ist dem Gödelschen Problem jedenfalls nicht zu entgehen. Das aber gilt für das Denken ebenso wie für die Maschine; beide finden immer wieder das Gödelsche Problem vor.

((31)) Diese Überlegungen zeigen also, daß - auch wenn es zunächst so scheinen könnte - in keinem Fall ein grundsätzlicher Unterschied bezüglich der logischen Möglichkeiten des Denkens und der Maschine auszumachen ist. Aus Gödels Theoremen ist dafür jedenfalls kein Argument zu gewinnen, und nur dies stand hier zur Klärung an.

((32)) Daß Gödels Resultate in dieser Hinsicht oft falsch eingeschätzt worden sind, ist sicher auch auf eine Reihe irreführender Formulierungen in der Gödelliteratur zurückzuführen. So zum Beispiel, wenn gesagt wird, daß G nicht mit den Mitteln des Systems selber beweisbar sei.[43] Nach den früheren Darlegungen kann es nur heißen, daß der Satz G aufgrund seiner selbstreferentiellen Struktur nicht beweisbar sein kann; denn von den beweistechnischen Mitteln des Systems hängt die Unbeweisbarkeit des Satzes G ja tatsächlich nicht ab.[44] Irreführend ist auch die Aussage bei Nagel/Newman, derzufolge Gödels Theoreme "eine grundlegende Begrenzung für die Reichweite der axiomatischen Methode" bedeuten sollen (1964, 93)[45], denn, so die Folgerung, die von den Autoren gezogen wird: Die Existenz eines Satzes, der, wie G, wahr und gleichwohl unbeweisbar ist, zeige, daß es "arithmetische Wahrheiten gibt, die nicht formal beweisbar sind" (1964, 99, ähnlich 85, 96).[46] Auch hier kommt es darauf an, zwischen dem Beweis des 'Gödelschen' Sachverhalts und des ihm in S korrespondierenden formalen Ausdrucks klar zu unterscheiden; denn der Sachverhalt der Unbeweisbarkeit von G und damit die ihm entsprechende 'arithmetische Wahrheit' wird ja tatsächlich streng bewiesen, nur eben nicht der Ausdruck G, der eben diesen Sachverhalt mit den formalen Mitteln von S selbst formuliert. Daß ein derartiger Ausdruck nicht beweisbar ist, beruht nach dem Vorigen andererseits nicht auf einem Mangel des in S installierten formalen Beweisverfahrens, sondern hat, wie gesagt, rein semantische Gründe. Oder anders gewendet: Daß kein Computer je imstande sein wird, G zu beweisen, beruht keinesfalls darauf, daß es für eine Maschine etwa unmöglich wäre, bestimmte Operationen auszuführen, wie sie dem Denken möglich sind, sondern ist semantisch bedingt und stellt damit für das Denken ebenso eine Unmöglichkeit dar wie für die Maschine.

((33)) Dennoch, Gödels Resultat bleibt irritierend: Wie kommt es, daß ein Ausdruck vom Typus des Satzes G derart unüberwindliche Schwierigkeiten involviert? Der Grund dafür, so hat sich gezeigt, ist das Auftreten semantischer Beziehungen, wie sie durch die Arithmetisierung metatheoretischer Ausdrücke generiert werden, und insbesondere der Umstand, daß G ein selbstreferentieller Ausdruck ist, der seine eigene Unbeweisbarkeit aussagt. Nur aufgrund dieser Selbstreferentialität kann G nicht beweisbar sein[47], da andernfalls auch Non-G beweisbar wäre (s. Teil II). Von daher versteht man sofort, daß der durch G ausgedrückte Sachverhalt (die Unbeweisbarkeit von G) in einer von G verschiedenen sprachlichen Form sehr wohl beweisbar sein kann; denn eine solche Aussage ist ja nicht mehr selbstreferentiell: Indem sie über G spricht, spricht sie nicht über sich selbst. Dies ist, was Gödel tut, indem er den Beweis der Unbeweisbarkeit von G in einer von S verschiedenen Metasprache formuliert. Hier wird also keineswegs eine Grenze des formalen Beweisverfahrens sichtbar, wie immer wieder behauptet, sondern eine Unmöglichkeit, die sich allein aus der Selbstreferentialität von G ergibt. Und genau aus diesem Grunde - nicht wegen einer prinzipiellen Beschränktheit formaler Systeme - kann es eine beweisbare Aussage der Unbeweisbarkeit von G nur außerhalb des Systems S geben, denn auf diese Weise ist die Formulierung des 'Gödelschen Sachverhalts' eben nicht mehr selbstreferentiell.[48]

((34)) Auf der anderen Seite könnte selbst der Ausdruck G, als rein formales Gebilde genommen, sehr wohl beweisbar sein, wenn er nicht außerdem semantisch selbstreferentiell wäre. Selbstreferentiell ist G aber nur aufgrund der in S arrangierten Arithmetisierung eines Teils der Metatheorie von S. Ohne Arithmetisierung könnte G also durchaus beweisbar sein, was am einfachsten wieder dadurch erreichbar wäre, daß G als zusätzliches Axiom zu den übrigen hinzugefügt würde. Der Tatbestand der prinzipiellen Unbeweisbarkeit von G ergibt sich vielmehr erst aufgrund der Arithmetisierung: Durch diese werden im Rahmen der Arithmetik zusätzliche, semantische Strukturen etabliert, die einerseits die Konstruktion eines selbstreferentiellen Ausdrucks wie G ermöglichen und andererseits auch neue 'arithmetische Wahrheiten' generieren, z.B. die, daß eine bestimmte Zahl z nicht zur Klasse T der Gödelzahlen der Theoreme gehört. Natürlich 'gibt' es die so charakterisierten Zahlen auch unabhängig von den per Arithmetisierung getroffenen Zuordnungen, aber eben nicht als die Klasse T der entsprechenden Gödelzahlen: Diese ist vielmehr erst durch die solchen Zahlen gemeinsame Bedeutung (hier: die Beweisbarkeit einer Formel) festgelegt, die ihnen durch die Arithmetisierung zugeordnet wurde; d.h. T ist in der Tat semantisch definiert, und auch die mit der Existenz von T zusammenhängenden arithmetischen Wahrheiten beruhen auf diesen semantischen Strukturen, die erst durch die Arithmetisierung in das System hineingekommen sind. Die Wahrheit derartiger arithmetischer Beziehungen hat ihren Grund somit nicht mehr nur in den arithmetischen Axiomen selbst, sondern in zusätzlichen, aus dem Arithmetisierungsverfahren stammenden semantischen Bedingungen. Die Axiome selbst verbürgen diese Wahrheiten also tatsächlich nicht, aber das hat ersichtlich nichts mit Gödels Resultat zu tun (sondern mit jenen semantischen Strukturen) und bedeutet daher auch nicht, daß es grundsätzlich keine Axiomatisierung solcher Sachverhalte geben könne. Durch Gödel ausgeschlossen, noch einmal gesagt, ist nur, daß der selbstreferentielle Satz G Axiom oder aus Axiomen herleitbar sein kann, dies aber nicht wegen einer prinzipiellen Defizienz der axiomatischen Methode, sondern allein wegen seiner Selbstreferenz. In nicht-selbstreferentieller Form hingegen wäre die durch G formulierte Wahrheit sehr wohl axiomatisierbar.

((35)) Zurück zur Frage der Differenz von Denken und Maschine: Das Denken, so ist aufgrund der entwickelten Überlegungen festzustellen, hat es mit der Konstruktion eines selbstreferentiellen Ausdrucks G gewissermaßen selbst so

eingerichtet, daß G notwendig unbeweisbar ist. Durch die Arithmetisierung von S und die dadurch erst ermöglichte Formulierung von G hat es selbst in das System eine Unmöglichkeit hineingelegt, die als unüberwindliche objektive Schranke zu deuten ein Mißverständnis wäre. Gödels Theoreme, soviel ist damit klar, können weder als Beleg für eine grundsätzliche Differenz der Logik des Denkens und der Maschine noch auch im Sinne einer prinzipiellen Überlegenheit des Denkens über die Maschine gedeutet werden. Oder umgekehrt: Gödels Resultat bedeutet nicht, daß die Maschine auf eine Primitivlogik festgelegt wäre, die an die Logik des Denkens niemals heranreichte. Die Auffassung, daß Mensch und Maschine derselben Logik unterstehen, ist durch Gödels Konstruktion jedenfalls in keiner Weise widerlegt oder auch nur tangiert.

((36)) Gewiß - um möglichen Mißverständnissen vorzubeugen: An der faktischen Superiorität des Denkens im Vergleich mit faktischen Maschinen ist nicht im mindesten zu zweifeln. Aber dieses Faktum begründet keine prinzipielle Differenz. Man könnte einwenden, daß das Denken seinen technischen Produkten, einfach durch seine Urheberschaft, überlegen sei. Ein Blick auf moderne Computer lehrt, daß dies in quantitativer Hinsicht heute schon nicht mehr generell zutreffend ist - man denke nur an die extremen Rechengeschwindigkeiten und Speicherkapazitäten solcher Maschinen. Doch es wäre ignorant, die gewaltige qualitative Differenz von Denken und Computer für die gegenwärtige Situation zu leugnen. Noch sind wir es, nicht die Maschinen, die denken und so unter anderem auch Logiksysteme und Maschinen erfinden; und solche Gebilde sind, worauf wiederum Lucas (1964, 48 ff) eindringlich hingewiesen hat, beschränkte, fixierte Gestalten, denen wir, als deren Bildner, stets viele Schritte voraus sind. Wir haben die Fähigkeit selbstüberholender Reflexion, die uns instandsetzt, unseren jeweiligen Zustand immer noch zu überbieten. Wir sind uns gewissermaßen selbst voraus und dabei - das ist wesentlich - dennoch dasselbe identische Subjekt. Dieses schon aus der philosophischen Tradition geläufige Argument - zu erinnern wäre an Kants 'transzendentale Apperzeption', Fichtes Reflexionsbegriff, Hegels Nachweis der Begriffsstruktur von Subjektivität oder auch an anthropologische Konzepte bei Scheler, Plessner, Gehlen und nicht zuletzt auch bei Heidegger - dieses bekannte Argument, das auch von Lucas ins Spiel gebracht wird, ist schwerlich bestreitbar. Aber Lucas irrt, wenn er meint, daß sich die technische Rekonstruierbarkeit eines solchen Subjekts aufgrund des Gödelschen Unvollständigkeitstheorems grundsätzlich verbiete. Zumindest der Übergang zu neuen, erweiterten Systemen wird durch Gödels Resultat nicht im geringsten ausgeschlossen, in gewissem Sinne sogar erzwungen.[49]

((37)) Eine - im Gegensatz zu Lucas' tendenziell antimechanistischer Gödeldeutung - dezidiert mechanistische Position wird in dem anregenden Buch von J.C.Webb (1980) vertreten. Hatte Lucas Gödels Resultate als die "Achillesferse" der Maschine bezeichnet (1964, 47), so nennt Webb sie demgegenüber "guardian angels" (1980, 202, 208), Schutzengel des Mechanismus in dem Sinne, daß die Unvollständigkeit formaler Systeme im Blick auf das Verhalten von Maschinen zu nicht berechenbaren, unvorhersagbaren Prozessen führe (200, 209, 245 f) und selbst die Möglichkeit von "self-reflection" involviere (246, 235), wovon, so Webb, "frühere Mechanisten nur hätten träumen können" (235). Gödels Theoreme seien, recht verstanden, "genau das, was der Doktor dem Mechanismus verordnet" habe (2oo), so daß "Mechanisten ihren Glücksstemen für seine Beweise danken" könnten (245). Webb illustriert seine Auffassung am Beispiel einer "Gödelmaschine", wie er sie nennt, die mit anderen Maschinen kommuniziert "wie eine Person, die genau zu denen hallo sagt, die zu ihr hallo sagen"(234), ein Vorgang, der durch Hineinverlegung in die Gödelmaschine auch "more 'introspective'" gestaltet werden könne (235).

((38)) Nun, das ist wohl zu einfach: Die Möglichkeit unbestimmten, nicht berechenbaren Maschinenverhaltens aufgrund Gödelscher Unvollständigkeit[50] soll eine Affinität zu menschlichem, sich selbst als frei verstehenden Handeln suggerieren (245 f) - eine sicher nicht weniger dubiose mechanistische Vereinnahmung Gödels als die antimechanistische, denn: Ist für das Handeln wirklich Unberechenbarkeit spezifisch, um als menschliches gelten zu können, und ist für das Problem des Selbstbewußtseins irgendetwas gewonnen durch den Nachweis, daß eine solche 'Gödelmaschine' zu sich 'hallo' sagen kann? Derartiges wäre zudem mit simpleren, 'nicht-Gödelschen' Mitteln erzielbar. Hier werden improvisierte, simplifizierende Geistmodelle in Anschlag gebracht, und J.R.Searles Kritik dieser Art von 'Kognitionswissenschaft' ist nur zu berechtigt.[51] Der Gedanke andererseits, daß Gödels Theoreme möglicherweise auch etwas zur Klärung der Probleme des Selbst und der Willensfreiheit beitragen können, ist indessen nicht von der Hand zu weisen und mag in verschiedener Hinsicht sogar manches für sich haben. Aber beiläufige Versicherungen oder dunkle Andeutungen sind diesbezüglich, auch wenn sie, wie bei Hofstadter etwa, gehäuft auftreten[52], wenig hilfreich.

((39)) Abschließend noch eine Bemerkung zum Grundsätzlichen: Die entwickelten Überlegungen lassen sich im weiteren Sinne als einen Beitrag zum Problem Künstlicher Intelligenz verstehen und so gewissermaßen als ein Stück technikphilosophischer Reflexion in anthropologischer Absicht. Technische Rationalität, so war eingangs gesagt worden, ist auch und gerade als Objektivation menschlicher Möglichkeiten, als deren Wiederholung und sogar Überbietung in objektiven Systemen zu begreifen. Insofern ist es, bei aller Gefahr der Verselbständigung und Entgleisung der Technik, sicher verfehlt, in dieser nur eine negative Macht zu sehen. Ihr Telos ist keineswegs nur Beherrschung, Ausbeutung oder gar Perversion der Natur, sondern nicht zuletzt auch deren intellektuelle Durchdringung und in eins damit Selbstobjektivation des intelligenten Wesens. Namentlich auch die 'Denkmaschine' ist als ein Versuch technischer Selbsteinholung des Denkens zu verstehen. Schon von daher ist es unwahrscheinlich, daß die Logik der Maschinen von der des Denkens prinzipiell verschieden ist. Mehr noch: Der einfa-

che Gedanke, daß auch das Gehirn ein durch und durch physisch bestimmtes und insofern grundsätzlich technisch rekonstruierbares System sein muß, scheint mir unabweisbar zu sein, und es ist wichtig zu realisieren - darum war es hier allein zu tun -, daß Gödels Resultate einer solchen Annahme entschieden nicht entgegenstehen. Gewiß, das ist nur ein negatives Resultat, aber dennoch, wie ich hoffe, immerhin ein Beitrag zur Klärung einer notorischen Unklarheit des Mensch-Maschine-Problems und damit auch des Projekts Künstlicher Intelligenz.

## Anmerkungen

1) M.Heidegger 1962, Die Technik und die Kehre, Pfullingen 1962, 15.

2) K.Gödel 1931, Über formal unentscheibare Sätze der Principia Mathematica und verwandter Systeme I, in: Monatshefte für Mathematik und Physik, Bd. XXXVIII, dort Satz VI und Satz XI.

3) Vgl. hierzu die instruktiven Darstellungen von W.K.Essler 1964, Aufzählbarkeit und Cantorsches Diagonalverfahren, Diss. München 1964, § 8; E.Nagel/J.R.Newman 1964, Der Gödelsche Beweis, Wien/München 1964, Kap. VIII; W.Stegmüller 1973, Unvollständigkeit und Unentscheidbarkeit, Wien/New York $^{3}$1973.

4) Vgl. Essler 1964, 44. Auf die Bedeutung semantischer Beziehungen in metamathematischen Kontexten hat F.v.Kutschera 1964, Die Antinomien der Logik, Freiburg/München 1964, 75 ff, hingewiesen.

5) Gödel 1931, 176; vgl. auch Stegmüller 1973, 7 f.

6) G ist hier also der Name dieses Ausdrucks.

7) So die semantische Fassung des Unvollständigkeitsbegriffs, vgl. W.K.Essler 1969, Einführung in die Logik, Stuttgart $^{2}$1969, 246.

8) Vgl. z.B. Essler 1964, 48.

9) H.Meschkowski 1978a, Richtigkeit und Wahrheit in der Mathematik, Mannheim/Wien/Zürich $^{2}$1978, 76.

10) W.V.O.Quine 1978, Was es gibt, in: W.Stegmüller (Hg.) 1978, Das Universalienproblem, Darmstadt 1978, 121.

11) Vgl. auch C. Thiel 1972, Grundlagenkrise und Grundlagenstreit, Meisenheim am Glan 1972, 123 ff; P. Weibel/ E. Köhler 1986, Gödels Unentscheidbarkeitsbeweis. Ideengeschichtliche Konturen eines berühmten mathematischen Satzes, in: Gödelsatz, Möbiusschleife, Computer-Ich. Wien 1986.

12) Hierzu z.B. J.C.Webb 1980, Mechanism, Mentalism, and Metamathematics, Dordrecht (Holland) 1980, Ch.IV; ferner S.Körner 1968, Philosophie der Mathematik, München 1968, 115.

13) J.R.Lucas 1964, Minds, Machines, and Gödel, in: A.R.Anderson (Hg.) 1964, Minds and Machines, Englewood Cliffs, New Jersey 1964, 47. - Vgl. hierzu auch G.Franck 1987, Menschlicher Geist und künstliche Intelligenz, in: MERKUR, Bd.41 (1987) 950 ff; W.Wieser 1987, Spiegelungen und Spiegelfechtereien, ebd. 943 f.

14) Nagel/Newman 1964, Kap. VIII.

15) G.Frey 1965, Sprache - Ausdruck des Bewußtseins, Stuttgart 1965, 37 ff, 41 ff; ähnliche, sich wiederholende Formulierungen in anderen Arbeiten des Autors: Frey 1966, Sind bewußtseinsanaloge Maschinen möglich?, in: STUDIUM GENERALE, 19.Jahrgang, 1966; Frey 1967, Die Mathematisierung unserer Welt, Stuttgart 1967, 130 ff; Frey 1968, Können Maschinen Bewußtsein haben?, in: n+m. Naturwissenschaft und Medizin, Nr. 24, 1968.

16) H.Meschkowski 1978b, Problemgeschichte der neueren Mathematik (1800-1950), Mannheim/Wien/Zürich 1978, 289.

17) Hierzu auch Webb 1980, Ch. IV.

18) D.R.Hofstadter 1985, Gödel, Escher, Bach, Stuttgart 1985.

19) J.-F.Lyotard 1979, La Condition Postmoderne, Paris 1979, 70.

20) Ein nicht-selbstreferentieller Satz könnte einem anderen Satz sehr wohl die Eigenschaft der Unbeweisbarkeit zusprechen, ohne dadurch selbst als unbeweisbar und wahr oder als beweisbar und falsch qualifiziert zu sein.

21) Hierzu auch D.Wandschneider 1979, Formalität und Wahrheit, in: G.Simon/E.Straßner (Hg.), Sprechen - Denken - Praxis, Weinheim/Basel 1979.

22) Gödel 1931, Anm. 15; Stegmüller 1973, 10. Auf die Wichtigkeit der selbstreferentiellen Struktur von G hat dagegen v.Kutschera 1964, Kap. II.3, nachdrücklich aufmerksam gemacht.

23) Vgl. B.Rosser 1939, An Informal Exposition of Proofs of Gödel's Theorems and Church's Theorem, in: Journal of Symbolic Logic, Vol. 4, 1939.

24) Gödel 1931, 176. Schwächer deshalb, weil Korrektheit formale Widerspruchsfreiheit impliziert: Ist ein System nämlich nicht widerspruchsfrei, so ist mit einem Satz A zugleich Nicht-A beweisbar. Einer von beiden ist aber falsch, so daß ein falscher Satz beweisbar und das System damit nicht mehr korrekt ist. Doch das Umgekehrte gilt nicht: Ist ein System nicht korrekt, so ist ein falscher Satz F beweisbar. Nicht-F ist also ein wahrer Satz. Wäre das System vollständig (alle wahren Sätze sind auch beweisbar), so wäre (außer F) folglich auch Nicht-F beweisbar, das System mithin nicht widerspruchsfrei. Ist das System hingegen unvollständig, so ist dieser Schluß nicht möglich, d.h. ein nicht korrektes System kann, sofern es unvollständig ist, durchaus widerspruchsfrei sein.

25) Dem liegt Satz V in Gödel 1931 zugrunde; vgl. dort auch Anm. 39.

26) Zum Begriff der formalen Ausdrückbarkeit z.B. Stegmüller 1973, 21; vgl. auch Essler 1964, 44.

27) Vgl. Essler 1969, 245.

28) Folgte die Beweisbarkeit von G, so wäre das ein Widerspruch zu dem vorherigen Resultat der Unbeweisbarkeit von G, d.h. das System wäre nicht widerspruchsfrei.

29) Mit der Bedingung einfacher Widerspruchsfreiheit könnte der Übergang zur Unbeweisbarkeit von Nicht-G nur vollzogen werden, wenn die generelle Beweisbarkeit von G selbst gesichert wäre (was hier nicht gegeben ist). Mit der stärkeren Bedingung der Omega-Widerspruchsfreiheit ist hierfür schon die Beweisbarkeit der 'Einzelaussagen' zureichend. Rosser (1939) hat aber gezeigt, daß bei einem geeignet modifizierten Ausdruck G' schon die einfache Widerspruchsfreiheit ausreicht, um die Unbeweisbarkeit von Nicht-G' zu beweisen.

30) Gödel 1931, 4.Kap.

31) Vgl. Anm. 47.

32) Vgl. Weibel/Köhler 1986, 89.

33) Zur Unfundiertheit selbstreferentieller Ausdrücke in antinomischen Strukturen s. D.Wandschneider 1986, Das Antinomienproblem und seine pragmatische Dimension, in: H.Stachowiak 1986, PRAGMATIK, hg.v. H.Stachowiak, Hamburg 1986 ff, Bd. IV.

34) Gödel 1931, 175; vgl. auch Essler 1964, 46 f; Stegmüller 1973, 3 ff.

35) Vgl. z.B. Webb 1980.

36) So der Titel von Frey 1966, eine Arbeit, die, wie auch ihre Folgever-

sionen (s. Anm. 15), in einigen Punkten, so scheint mir, eher zur Verwirrung der Diskussion über dieses Thema beigetragen hat; hierzu Wandschneider 1979, 262.

37) Ich nehme hiermit Bezug auf eine Frage von C.Stetter (Aachen), dem ich im übrigen für anregende, klärende Gespräche über diese Thematik danken möchte.

38) M.Arbib 1964, Brains, Machines, and Mathematics, New York/San Francisco/Toronto/London 1964, 140.

39) H.Putnam 1964, Minds and Machines, in: Anderson 1964, 77.

40) Vgl. R.C.Lyndon 1966, Notes on Logic, New York/Toronto/ London/ Melbourne 1966, 84 f; vgl. auch Essler 1964, 46 f, sowie W.K.Essler 1972, Analytische Philosophie I, Stuttgart 1972, 147 ff.

41) Wie sich das Auftreten von Antinomien ohne Trennung von Sprachstufen vermeiden läßt, habe ich in D.Wandschneider 1974, Zum Antinomienproblem der Logik, in: RATIO, Bd.16, 1974, und in Wandschneider 1986 entwickelt.

42) Für klärende Gespräche über diesen Punkt möchte ich W.K. Essler (Frankfurt/M.) danken.

43) Z.B. Frey 1966, 200.

44) Eine Formulierung wie die bei Essler 1972: Gödel habe "gezeigt, daß man logische und mathematische Systeme nur mit solchen Mitteln als widerspruchsfrei erweisen kann, die in der betreffenden Sprache nicht vollständig formulierbar sind"(149 f, Hvh.D.W.), rekurriert demgegenüber auf den Umstand, daß die im System S verfügbaren sprachlichen Ausdrucksmittel für die Formulierung eines Widerspruchsfreiheitsbeweises untauglich sind, insofern sie ein in S grundsätzlich unbeweisbares Element - eben G - enthalten.

45) Auch G.Vollmer 1986, Was können wir wissen? Bd.2: Die Erkenntnis der Natur, Stuttgart 1986, betrachtet dies zumindest als eine mögliche Deutung (284). Die beiden anderen von Vollmer genannten Möglichkeiten - Widersprüchlichkeit bzw. stochastische Funktion des Gehirns (283 f) - betreffen lediglich die Frage der faktischen Gehirnfunktionen und nicht mehr das Problem, wie die Gödeltheoreme in der Computerperspektive zu deuten sind.

46) Ähnlich Gödel selbst in einem unveröffentlichten Vortrag, zitiert bei Weibel/Köhler 1986, 98.

47) Inwiefern die Negation ('Unbeweisbarkeit') dafür wesentlich ist, soll hier nicht mehr untersucht werden. Aber es leuchtet unmittelbar ein, daß der positive Satz 'Dieser Satz ist beweisbar' keine Schwierigkeiten dieser Art involvieren würde; vgl. auch Wandschneider 1986.

48) In diesem Zusammenhang legt sich ein weiterer Gedanke nahe: Angenommen, man könnte die Unbeweisbarkeit von G (die ja von G selbst formuliert wird) auch in einer von G abweichenden sprachlichen Form H ausdrücken, die aber ebenfalls nur die in S verfügbaren formalen Mittel benötigte. H machte inhaltlich also die gleiche Aussage wie G, aber - das ist hier wesentlich - in nicht-selbstreferentieller Form, denn H spricht über G, nicht über sich selbst. Gödels Argumentation, die ja nur aufgrund der Selbstreferentialität von G möglich war (s. o.), würde somit für den Ausdruck H nicht mehr zutreffen und dessen Beweisbarkeit nicht mehr prinzipiell ausschließen. Man könnte den Ausdruck H also etwa als zusätzliches Axiom wählen, ihm dadurch Beweisbarkeit sichern und hätte dergestalt sogar eine im System S formalisierte und gleichwohl beweisbare Aussage desselben Sachverhalts, der auch von G formuliert wird, in dieser Form aber aufgrund der Selbstreferentialität von G nicht beweisbar sein kann. Mit der Beweisbarkeit von H wäre so der häufig artikulierte Mangel, daß die durch G formulierte arithmetische Wahrheit im System S selbst grundsätzlich nicht formal-axiomatisch beweisbar ist, behoben. Daß G selber in S formal unbeweisbar ist, hätte dann nichts Bemerkenswertes mehr: Es gibt trivialerweise viele Ausdrücke, die in S (sofern das System widerspruchsfrei ist) nicht beweisbar sind. - Freilich ist die Konstruierbarkeit einer solchen Formel H keineswegs selbstverständlich. Am einfachsten erschiene es, in der Formulierung von H auf die Gödelzahl g von G bezugzunehmen. Dies verbietet sich aber, weil die Hinzunahme von H zu den Axiomen (über den Beweisbegriff) zu einer Verschiebung des Systems der Gödelzahlen führen müßte mit der Folge, daß g nun nicht mehr die Gödelzahl von G wäre und H daraufhin auch keine Aussage mehr über G machte. Der beabsichtigte Effekt wäre damit verfehlt. Man müßte gewissermaßen im voraus wissen, zu welcher Gödelzahl von G die Hinzunahme des Axioms H führt, und das Ergebnis dieser Operation andererseits schon vorab für die Formulierung von H kennen. Mir ist nicht klar, ob es dafür überhaupt ein Verfahren geben könnte. - Vielleicht wäre es aber auch möglich, für H eine allgemeinere, G indirekt charakterisierende Formulierung zu finden, in der nicht mehr vorab auf G bzw. g bezuggenommen werden müßte, etwa ein Satz der Art: 'Jede selbstreferentielle Unbeweisbarkeitsaussage ist unbeweisbar', der also die Aussage der Unbeweisbarkeit von G mit einschließen würde. Ließe sich ein solcher Satz ohne Bezugnahme auf G in Gödelzahlen kodieren, so könnte H ohne weiteres als zusätzliches Axiom gewählt werden. Die Gödelzahl g von G würde sich daraufhin zwar - etwa in g' - verändern, aber das würde sich nicht wiederum auf H und damit auch auf g' selbst auswirken usw., d. h. die Arithmetisierung bliebe eindeutig. Ob eine derartige Möglichkeit besteht, wäre zu prüfen. Gewisse Anhaltspunkte dafür entnehme ich A.A. Johnstone 1981, Self-Reference, the Double Life and Gödel, in: Logique et Analyse, Bd.24 (1981), 41 ff.

49) Die Frage der Selbstidentität eines solchen Systems ist freilich eines der schwierigsten und bislang ungeklärten philosophischen Probleme, das auch hier offenbleiben muß, zumal es im gegenwärtigen Zusammenhang nur um die viel allgemeinere Frage der Mensch und Maschine zugrundeliegenden Logik geht.

50) Dies ist eine Folge davon, daß der Satz G, aufgrund seiner Selbstreferentialität, eine unfundierte und damit unbestimmte Aussage darstellt; s. Abschn. II.

51) J.R.Searle 1986, Geist, Hirn und Wissenschaft, Frankfurt/M. 1986, Kap. II und bes. III.

52) Vgl. Hofstadter 1985, 741 ff, 753 ff, 76o ff.

Adresse

Prof. Dr. Dieter Wandschneider, (privat:) Theresienstraße 18, D-5100 Aachen; (dienstlich:) Philosophisches Institut der RWTH Aachen, Eilfschornsteinstr.16, D-5100 Aachen