

ETHIK UND SOZIALWISSENSCHAFTEN

Streitforum für Erörterungskultur

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

EuS 1 (1990) Heft 1

Replik

Gödelsche Selbstreferentialität und maschinelle Aspekte

Dieter Wandschneider

((1)) Auf die Texte werde ich mit folgenden Abkürzungen be-
zugnehmen:

Al - Alich
Be - Beckermann
Bl - Blau
El - Eley
Fu - Fuhrmann
Kr - Krämer
Rh - Rheinwald

Ro - Roth
 Ste - Stekeler-Weithofer
 Sto - Stolz
 Stu - Stuhlmann-Laeisz
 Te - Tetens
 Va - Varga v. Kibéd
 Wa - Wandschneider

((2)) Entsprechend der Thematik meines Artikels habe ich zwei Teile vorgesehen. Zunächst geht es um das Gödelproblem selbst (wobei ich mich auf die Diskussion der Unbeweisbarkeit des Gödelschen Ausdrucks G beschränken werde, die ja auch für Gödels 2. Theorem fundamental ist) (I) und sodann um die Frage möglicher Konsequenzen für 'Denkmaschinen' (II). - Vorab möchte ich den Kritikern für ihre Mühe danken und zugleich um Entschuldigung und Verständnis für die sehr punktuelle Bezugnahme auf die einzelnen Beiträge bitten, denen man so natürlich nicht entfernt gerecht werden kann. Ich mußte mich hier auf die zentral diskutierten und kritisierten Fragen konzentrieren. Im übrigen betrachte ich diese Rückkopplung, verbunden mit der Möglichkeit, nachträglich noch Klarstellungen und Korrekturen vornehmen zu können, als eine sehr begrüßenswerte Innovation in der Form wissenschaftlicher Auseinandersetzung.

I

((3)) Zunächst also zum Gödelproblem. Sehr kontrovers ist die von mir besonders herausgestellte Frage der Selbstreferentialität von G diskutiert worden. Ich beziehe mich hier einfachheitshalber auf Fuhrmanns Darstellung ((Fu 8)) des Diagonalisierungslemmas für G (wobei E hier der Existenzquantor, '-' die Negation, g die Gödelzahl von G , ' g ' das zugeordnete Zahlzeichen und G jetzt keinName, sondern eine Formel ist). Im System S ist also

DLG $G \leftrightarrow \neg(\exists y)\text{bew}('g', y)$

beweisbar.¹ Die Selbstreferenz besteht hier ersichtlich darin, daß mit ' g ' in der rechts stehenden Formel auf den Satz G bezuggenommen ist, der dieser Formel selbst äquivalent ist. Zugleich ist ein möglicher Grund für die wiederholte Abwertung oder Leugnung der Selbstreferentialität von Gödels Ausdruck (z.B. ((Be 2)), ((Rh Anm. 6 (dagegen Rh 7, 12))), ((Ste 10 f)), ((Va 10 f))) deutlich: Wer nur die rechte Formel betrachtet, die bereits alles Wesentliche zu enthalten scheint, sieht nicht mehr und nicht weniger als einen formalen Ausdruck, aber keinerlei 'Referenz'.

((4)) Tetens macht einen Vorschlag zur Güte: Der menschliche Logiker, der die arithmetische Kodierung der Metatheorie versteht und mitvollzieht, erkennt die inhaltliche Selbstreferenz der Gödelformel. "Aber das System S 'vollzieht' die Gödelisierung und Arithmetisierung der Metatheorie nicht mit" ((Te 6)). Gleichwohl: Gödels Beweis ist es, der die durch Arithmetisierung ermöglichte Selbstreferenz nachvollziehen muß. Fuhrmanns Darstellung macht dies unmittelbar

deutlich: Aus der Annahme der Beweisbarkeit von G (immer in S) würde die Beweisbarkeit von $(\exists y)\text{bew}('g', y)$ folgen. Gemäß (DLG), und das heißt eben aufgrund der Selbstreferentialität von G , ist dies aber äquivalent $\neg G^2$ und damit die Reductio ad absurdum der Annahme. Die Selbstreferentialität in dem angegebenen Sinne ist für die Argumentation also essentiell; andernfalls wäre diese nicht möglich. Hätte (DLG) z.B. die Form $H \leftrightarrow \neg(\exists y)\text{bew}('g', y)$ (mit einem G nicht äquivalenten Satz H), wäre hier also keine 'Diagonalbeziehung' und damit Selbstreferenz gegeben, so wäre die vorherige Folgerung unmöglich: Die Beweisbarkeitsannahme für G mit der Konsequenz, daß $(\exists y)\text{bew}('g', y)$ beweisbar ist, würde nicht mehr zur Beweisbarkeit von $\neg G$, sondern zu der von $\neg H$ führen, und das wäre eben keine Reductio ad absurdum der Beweisbarkeitsannahme von G , d.h. G könnte in diesem Fall sehr wohl beweisbar sein. Rheinwalds etwas gequältes Zugeständnis, daß "irgendeine Art von - vielleicht verschlungenem - Selbstbezug" nur für die "Unvollständigkeits-Beweise" eine Rolle spielen könnte ((Rh 7)), sucht es so hinzustellen, als habe der Beweis nichts mit dem bewiesenen Sachverhalt zu tun. Doch wie könnte man beides trennen - und hier 'lebt' der Beweis, wie dargelegt, tatsächlich von der in G enthaltenen Selbstreferenz.³ Diese wiederum beruht auf der technischen Operation der Diagonalisierung. Es war übrigens nicht mein Ehrgeiz, wie Fuhrmann annimmt, eine darüber "in irgendeiner Weise hinausgehende Selbstreferentialität" nachzuweisen ((Fu 13)). Mein Hinweis ((Wa 19)) sollte lediglich besagen: Wenn in der 'Diagonalisierung' nur eine "formal-technische Operation" gesehen wird, so ist der Blick dafür verstellt, daß es gerade die so realisierte Selbstreferentialität von G ist, auf dem der Beweis der Unbeweisbarkeit von G entscheidend beruht.

((5)) Aus diesem Grund beruht er aber auch nicht spezifisch auf den Axiomen und damit auf dem besonderen Beweisbegriff von S . Man könnte geltend machen, daß der Beweisbegriff von S via Gödelisierung natürlich die Gestalt des Ausdrucks G mitbestimmt. Dessen Unbeweisbarkeit ist insofern allerdings S -spezifisch. Das von mir Gemeinte sollte ich daher präziser so ausdrücken, daß, wie immer der Beweisbegriff in einem System gewählt werden mag, die in ihm konstruierbare 'Gödelformel' (deren Konstruierbarkeit und Konsistenz des Systems vorausgesetzt) prinzipiell unbeweisbar ist. Ich denke, daß Rheinwald⁴ und Varga (vgl. ((Rh 8)), ((Va 26))) mit einer solchen Formulierung einverstanden sein könnten. Tatsächlich hängt dieser Umstand "gar nicht in erster Linie ab von dem Beweisbegriff ..., sondern von dem Ausdrucksreichtum" des Systems. Er ist "ganz unabhängig von einer näheren Präzisierung des Beweisbegriffs"; so F. v. Kutschera (1964, 86f), der anschließend gegen die Gödelsche Beweisidee überhaupt Vorbehalte geltend macht (nicht nur, wie Varga es hinstellt, gegen deren semantische Auffassung ((Va 5)), die v. Kutschera, im Gegenteil, selbst vertritt (1964, 7, 75 ff, 90 f u.a.)). Diese weitgehende Unabhängigkeit des Gödelschen Resultats von den deduktiven Mitteln des Systems, die zwar vorausgesetzt sind, aber noch nicht das Prinzipielle der Unbeweisbarkeit von G erklären, beruht, wie dargelegt, allein auf der Selbstreferenz von G .

((6)) 'Referenz' ist nun aber ein semantischer Begriff, und in diesem Sinne spreche ich von 'semantischen Strukturen'. Das ist verschiedentlich auf Widerspruch gestoßen. Es wird darauf hingewiesen (z.B. ((Bl 5)), ((Sto 4))), daß die Selbstreferenz nicht semantischer, sondern formal-syntaktischer Natur sei. Hier liegt, denke ich, ein Mißverständnis vor. Daß G bei inhaltlicher Interpretation eine syntaktische Aussage macht, ist richtig, aber der Umstand, daß eine Interpretation vorliegt, ist gleichwohl ein semantischer Sachverhalt. Oder anders gesagt: Der Inhalt der Aussage ist in diesem Fall sicher kein semantischer Gegenstand (wie etwa 'Referenz' oder 'Wahrheit'),⁵ aber (als Bedeutungsgehalt) dennoch Gegenstand der Semantik. (Ähnlich: 'Tisch' bedeutet nichts Semantisches, aber als Bedeutung ist 'Tisch' Gegenstand der Semantik.) Wenn ich im vorliegenden Zusammenhang von 'semantischen Strukturen' spreche, soll damit also nicht die von G gemachte Aussage selbst charakterisiert sein, sondern: daß mit G überhaupt eine Aussage verbunden ist.

((7)) Den Grund für das Auftreten solcher semantischer Strukturen sehe ich in der Gödelisierung. Dem wird verschiedentlich entgegengehalten, daß damit lediglich eine Zuordnung im Sinne einer bloßen Parallelisierung syntaktischer und zahlentheoretischer Sachverhalte vollzogen ist (z.B. ((Stu 10)), ((Te 7))). Das ist sicher nicht falsch, aber es ist zu wenig. Wenn man hier quasi zwei voneinander unabhängige Schichten sieht, so ist damit verkannt, daß Gödels Beweis gerade die Interdependenz beider benötigt: Er muß auf die Referenzbeziehung von 'g' und G - gewissermaßen eine Querverbindung zwischen den 'Schichten' - rekurrieren. Aus dem Grunde, denke ich, darf diese "sekundäre" semantische Beziehung (neben der "primären" zwischen einer Gödelzahl und deren Zahlzeichen, vgl. ((Va 4 und Anm. 1))) nicht vernachlässigt werden. Es geht, mit anderen Worten, gar nicht darum, die inhaltliche Deutung der Gödelformel im Detail nachzuvollziehen ((Te 6)); für Gödels Beweis hätte das in der Tat nur "heuristischen" Wert ((Va 6)), vgl. auch ((Be 2))). Aber die Zuordnung von G zu g, und über die 'primäre' Semantik auch zu 'g', ist eben nicht irrelevant. Sie scheint mir aus dem angegebenen Grund vielmehr zentral für den Beweis zu sein.

((8)) Angemerkt sei noch, daß Tetens' Darstellung von der Rolle, die ich der Selbstreferenz von G für Gödels Beweis zubillige ((Te 4)), so nicht zutreffend ist. "These (h)" braucht nicht als 'These' formuliert und "begründet" zu werden, sondern versteht sich von selbst: Indem G seine eigene Unbeweisbarkeit behauptet, ist G genau dann wahr, wenn G unbeweisbar ist (immer in S). Aber daraus läßt sich nicht schon folgern (und das tue ich auch nicht), daß die Unbeweisbarkeit von G in S "allein auf dem selbstreferentiellen semantischen Inhalt der Formel G" beruht ((Te 4)) - es sei denn, man macht die zusätzliche Voraussetzung, daß S korrekt ist, d.h. nur Sätze zu beweisen gestattet, die bei Interpretation wahr sind (vgl. ((Wa 11 f)); in ((Stu 14)) fehlt offenbar ein 'nicht'). Diese semantische Argumentation für die aus der Selbstreferentialität von G - unter der Vorausset-

zung der Korrektheit des Systems - zu erschließende Unbeweisbarkeit des Satzes bildet in meinem Text im übrigen nur einen Vorspann zur Analyse des eigentlichen Gödelschen Beweises, der statt der Korrektheit die schwächere Bedingung formaler Widerspruchsfreiheit (und der Omega-Widerspruchsfreiheit) benötigt und nicht semantisch, sondern syntaktisch ist. Aber auch und gerade für diesen Beweis gilt ja, wie dargelegt, daß die Selbstreferentialität von G entscheidend eingeht. Nicht die explizite Bedeutung des Satzes ist für den Beweis wesentlich, sondern daß in G die Gödelzahl für G selbst bzw. das zugeordnete Zahlzeichen auftritt.

((9)) Es ist in diesem Zusammenhang instruktiv, einmal zu überlegen, was geschieht, wenn G selbst zum Axiom erklärt würde. G ist jetzt also beweisbar. Wird das System S dadurch aber nicht inkonsistent? Keineswegs, es hat sich nur verändert; es ist so nicht mehr das System S, sondern ein neues System S'. G ist insofern nach wie vor unbeweisbar in S. Gehen wir nun etwas mehr ins Detail. Durch die Hinzufügung von G zu den übrigen Axiomen verändert sich der Beweisbegriff, und damit verschiebt sich auch die Zuordnung der Gödelzahlen zu den Ausdrücken von S', zumindest solcher Ausdrücke, die den Beweisbegriff enthalten. Infolgedessen gilt, daß g nun nicht mehr die Gödelzahl von G, sondern irgendeines anderen Ausdrucks oder möglicherweise gar keine Gödelzahl mehr ist. G enthält aber nach wie vor das Zahlzeichen 'g' für g, das jetzt freilich nicht mehr auf G selbst referiert, mit anderen Worten: G ist in S' nicht mehr selbstreferentiell und kann eben darum, so möchte ich im Sinne der früheren Überlegungen fortfahren, in S' beweisbar sein. Zugleich läßt sich nun ein anderer Ausdruck G' mit der Gödelzahl g' konstruieren, der in S' selbstreferentiell und damit 'aus Gödelschen Gründen' unbeweisbar ist. Auch im Vorgang der Systemerweiterung durch G zeigt sich so (indirekt) der Tatbestand der 'Unbeweisbarkeit aufgrund von Selbstreferentialität'.

((10)) Von daher wird auch meine Unterscheidung zwischen dem Satz G und dem durch G ausgedrückten Sachverhalt deutlicher ((Wa 24 ff, 32 f)). Diese, das habe ich ja ausgeführt, differenziert zwischen der S-Formulierung des Sachverhalts (der S-Unbeweisbarkeit von G) durch den Satz G selbst und einer in eine andere Sprache übersetzten Formulierung Ü desselben Sachverhalts. Die S-Formulierung G ist selbstreferentiell und darum (s.o.) unbeweisbar; die übersetzte Formulierung Ü hingegen macht zwar dieselbe Aussage, aber - als Übersetzung - über einen von Ü selbst verschiedenen Satz, ist also nicht mehr selbstreferentiell und kann daher, im Unterschied zu G, beweisbar sein. Das, scheint mir, ist das ganze Geheimnis des Übergangs in ein anderes System, von dem her der Sachverhalt der Unbeweisbarkeit von G bewiesen werden kann. So führt auch Gödel seinen Beweis, nämlich - zu einer paradoxen Formulierung zusammengezogen - den Beweis für G, wonach G unbeweisbar ist, aber er formuliert dies in einer von G verschiedenen sprachlichen Form, so daß daraus nicht mehr auf die gleichzeitige Beweisbarkeit von -G geschlossen werden kann.⁶ Das scheint mir der rationale Kern der von Stuhlmann-Laeis pointierten

Auffassung zu sein, derzufolge Gödels Beweis auf einer Objektivierungsleistung beruht ((Stu 16 f)), vgl. auch ((Kr 6)). Wenn Gödel also etwas über S Hinausgehendes benötigt, dann jedenfalls eine weitere sprachliche Ebene. Meine Aussage, daß sich der Logiker bei seinem Beweis ganz in S bewege ((Wa 23, 27)), ist insofern verfehlt (vgl. hierzu auch Blaus Kritik ((Bl 1 ff))) und mit meiner Formulierung, daß er den von ihm bewiesenen Sachverhalt in einer Metasprache mitteilen müsse, inkompatibel.⁷

((11)) Wenn bei diesen Überlegungen auf den bei inhaltlicher Deutung durch G ausgedrückten Sachverhalt bezuggenommen wurde, dann also nur deshalb, um darauf hinzuweisen, daß G (bei inhaltlicher Deutung) und Ü lediglich verschiedene sprachliche Formulierungen desselben Sachverhalts sind. Wird das übersehen, so können daraus, wie ich aus eigener Erfahrung weiß, immense Unklarheiten entstehen (wobei ich mir vorstellen kann, daß solche begleitenden Erwägungen auf den Logiker, den vorrangig die formale Triftigkeit des Beweisgangs selbst interessiert, eher befremdlich wirken mögen).

((12)) Ist nun die Selbstreferentialität von G beweisrelevant und 'Referenz' ein semantischer Begriff, dann muß in der Tat eine semantische Struktur für das Gödelsche Resultat verantwortlich gemacht werden. Zugleich ist deutlich, daß diese Struktur erst durch die Gödelisierung ermöglicht ist. Auch das klingt befremdlich, aber so stellt es sich mir dar, und ich stelle dies hiermit nochmals zur Diskussion.

((13)) Es scheint mir in diesem Zusammenhang klärend zu sein, der Frage weiter nachzugehen, warum G nicht axiomatisch beweisbar sein kann. Unmöglich ist dies, so hatte sich schon gezeigt, weil gemäß (DLG) aus der Beweisbarkeit von G auf die Beweisbarkeit von $\neg G$ geschlossen werden kann. Daß dies auf der Selbstreferentialität von G beruht, habe ich dargelegt. Die zugrundeliegende Struktur habe ich im Text ((Wa 10) näher verdeutlicht: Für einen Satz, der seine eigene Unbeweisbarkeit aussagt, kann es nur zwei Möglichkeiten geben: Er kann nur entweder unbeweisbar und wahr oder beweisbar und falsch sein; tertium non datur aufgrund der Selbstreferentialität. Betrachtet man, um das Wahrheitsprädikat hier aus dem Spiel zu lassen, statt der Wahrheit von G deren objektsprachliches Korrelat, d.h. G selbst, und statt der Falschheit von G entsprechend $\neg G$, so ist - informell gesprochen - die Beweisbarkeit von G mit $\neg G$ 'gekoppelt' (was im Gödelschen Beweis ja entscheidend ist), oder anders gesagt: Aufgrund seiner spezifischen, selbstreferentiellen Konstruktion schließt der Satz G seine Beweisbarkeit durch sich selbst aus; er steht seiner Beweisbarkeit gewissermaßen selbst im Wege. G ist damit per definitionem außerhalb der 'Reichweite' der Axiome von S. Sind Axiome die Grundlage formaler Beweise, so könnte man den Satz G geradezu ein Anti-Axiom nennen: Er selbst garantiert seine eigene Unbeweisbarkeit in S.

((14)) Nun drückt G in zahlentheoretischer Perspektive ja eine arithmetische Wahrheit aus, nämlich die, daß die Gödel-

zahl g von G nicht zur Klasse T der Gödelzahlen der Theoreme (von S) gehört. Es gibt also in der Tat arithmetische Wahrheiten, die nicht axiomatisch beweisbar sind, was immer wieder irritierend wirkt. Aber man muß sich klar machen, daß hier sozusagen jener 'anti-axiomatische' Charakter von G zum Tragen kommt:

((15)) Die Klasse T kann über die Gödelisierung arithmetisch (nach Maßgabe des Begriffs der 'Beweisbarkeit') definiert werden. Nachdem diese Zahlklasse aufgrund der Axiome und Deduktionsregeln von S festliegt, wird von Gödel ein Ausdruck konstruiert, dessen Gödelzahl garantiert nicht zu T gehört. Garantierbar ist das aber nur in der Weise, daß dieser Ausdruck einerseits auf T rekurriert - hier durch den Beweisbarkeitsbegriff -, um sich, bildlich gesprochen, 'davon abzu stoßen', d.h. die Gödelzahl g des Ausdrucks wird als ein Nichtelement von T konstruiert. Und andererseits muß der Ausdruck auf sich selbst bezugnehmen, denn er soll ja durch sich selbst die Nichtzugehörigkeit von g zu T garantieren. Die allgemeine Konstruktionsvorschrift für einen solchen Ausdruck muß also tatsächlich eine Selbstreferenz enthalten, durch die er sich als denjenigen Ausdruck bestimmt, dessen Gödelzahl gerade nicht zu T gehört. Beides, der Rekurs auf T (durch 'bew') und der Selbstbezug (G/'g'), ist in $G \leftrightarrow \neg(\exists y)\text{bew}('g', y)$ deutlich erkennbar. Die Entsprechung mit dem Cantorsche Diagonalverfahren ist ebenfalls offenkundig.⁸

((16)) Der Mangel des axiomatischen Beweisverfahrens - wenn man diesen Topos einmal aufnimmt - besteht im Lichte dieser Überlegungen also darin, daß Gödels Ausdruck durch seine spezifische, selbstreferentielle Konstruktion gewissermaßen 'außer Reichweite' der S-Axiome ist. Was hier über die Möglichkeiten des formal-axiomatischen Verfahrens hinausgeht, beruht solchermassen auf 'Referenz' und damit auf einer semantischen Struktur. So gesehen wird fast eine Tautologie daraus: Das über die formale Ebene der Axiome Hinausgehende ist genau das Semantische an Gödels Formel. Das Bemerkenswerte daran ist dies, daß die Arithmetik derartige Möglichkeiten einschließt; und zwar - damit komme ich zum Ausgangspunkt dieser Überlegungen zurück - offenbar auf der Grundlage der Gödelisierung der Ausdrücke. Das Zahlzeichen einer Gödelzahl ist demgemäß nicht einfach eine Zeichengestalt, sondern darüberhinaus Repräsentant, 'Referent' einer anderen (ihr durch Gödelisierung zugeordneten) Zeichengestalt. Es weist über sein bloßes Formsein hinaus auf etwas von ihm selbst Verschiedenes. Die rein formale Ebene ist damit verlassen.

((17)) Vargas Hinweis auf die von R. M. Smullyan entwickelten Sprachen, in denen Selbstreferenz möglich ist ((Va 12)), verstehe ich in diesem Zusammenhang eher als eine Bestätigung. Zentral für Smullyans Konstruktion ist nämlich, Quine folgend, das Verfahren der Bezeichnung von Ausdrücken durch deren 'Anführung'.⁹ Auch hierbei handelt es sich also um Referenzstrukturen. Daß es neben der Gödelisierung alternative Verfahren zur Kodifizierung von Ausdrücken gibt (vgl. z.B. ((Al 5))), heißt also nicht, daß in diesen Fällen

keine Referenzbeziehungen im Spiele wären.

II

((18)) Ich wende mich nun meinem zweiten Thema zu: der Frage, ob die durch Gödels Theoreme formulierten Einschränkungen auch die Möglichkeiten Künstlicher Intelligenz limitieren. Aus Gödels Resultaten, das ist die hier vertretene Auffassung, ergeben sich dafür keine Argumente.

((19)) Grundsätzlich gesehen hat man die folgende Situation: Ist der Gödelsche Ausdruck G in S unbeweisbar, so kann der Logiker dies in einem Metasystem beweisen. Charakteristisch für den Logiker ist also die Möglichkeit, zu einem Metasystem überzugehen, um von dort aus die Beschränkungen des Objektsystems sichtbar zu machen. Für den Mensch-Maschine-Vergleich mit Bezug auf das Gödelsche Problem wäre somit vor allem zu klären, ob die Möglichkeit eines solchen Übergangs nur für den Menschen oder grundsätzlich auch für die Maschine besteht. Das ist zweifellos ein weites Feld, dessen Bearbeitung in einem umfassenderen Rahmen vorzunehmen wäre. Zwei Punkte meines Textes möchte ich aber doch noch einmal deutlich machen und in ihrer Aussage bekräftigen:

((20)) (1) Aus Gödels Theoremen folgt nichts, was für oder gegen die Möglichkeit einer Maschine spräche, einen Systemübergang im angegebenen Sinne zu vollziehen. Trotzdem ist damit eine immer wieder zu beobachtende Unklarheit dieser Diskussion bezeichnet. Man begegnet immer wieder Argumentationen der folgenden Art: Maschinen sind technische Realisierungen formaler Systeme; gemäß Gödels Theoremen gibt es grundsätzliche Grenzen formaler Systeme. Also gibt es grundsätzliche Grenzen maschineller Systeme. Das scheint so unmittelbar aus Gödels Resultat zu folgen. Das Unzutreffende dieser Argumentation wird aber sofort deutlich, wenn man sich vergegenwärtigt, daß die erste Prämisse (Maschinen sind technische Realisierungen formaler Systeme) jedenfalls nicht aus Gödels Theoremen folgt. Über den Charakter von Maschinen sagen diese nichts aus, so daß von daher auch nichts über prinzipielle Grenzen von Maschinen zu erschließen ist. Ein von mir offenbar provoziertes Mißverständnis besteht in der Annahme einiger Kritiker (z.B. ((Be 1)), ((Te 4))), als solle das eben Gesagte aus dem semantischen bzw. selbstreferentiellen Charakter von G erschlossen werden, was natürlich keinen Sinn ergibt.

((21)) Und weiter: Es kann meiner Ansicht nach nicht gefolgert werden, daß jenes Metasystem, von dem aus Gödels Beweis geführt wird, über eine grundsätzlich andere Logik verfügt. Im Sinne der entwickelten Auffassung scheint mir nur gefordert zu sein, daß es von S unabhängige sprachliche Mittel bereitstellt.

((22)) (2) Der zweite Punkt betrifft die Frage (jetzt ohne den verfehlten Rekurs auf Gödels Theoreme): Ist möglicherweise auch eine Maschine befähigt, den Übergang zu einem

sprachlich erweiterten System zu vollziehen, so daß ihr die Durchführung des Gödelschen Beweises mit maschineneigenen Mitteln möglich würde? - eine Frage, die für die Beurteilung des Verhältnisses von Mensch und Maschine sicher zentral ist (vgl. ((Va 16))), auch wenn sie, wie Roth betont ((Ro 5)), gewiß nicht die konkrete Entwicklungsarbeit der KI betrifft. Auch diesbezüglich möchte ich meine grundsätzliche Auffassung ((Wa 29)) bekräftigen, daß die Festlegung der Maschine auf irgendwie ingeschränkte Operationsmöglichkeiten schon eine Vorentscheidung des Mensch-Maschine-Problems und damit eine Petitio principii wäre. Ich drücke mich hier etwas vorsichtiger aus als im Haupttext, denn meine Formulierung, daß die Maschine nicht von vornherein als ein "fixiertes" Gebilde mit "rigiden" Systemstrukturen zu betrachten sei, ist in diesem Zusammenhang mißverständlich: Was sollte die Maschine sein, wenn sie nicht in bestimmter Weise strukturiert wäre? Diese Unklarheit ist zu Recht kritisiert worden (vgl. ((Al 17)), ((Be 6)), ((Ste 1 f)), ((Sto 5)), ((Va 18))).

((23)) Auf der anderen Seite scheint es mir aus heutiger Sicht unmöglich zu sein zu sagen, was eine Maschine je können wird und was nicht. Das Modell der Turingmaschine ist nicht etwa schon die Antwort auf diese Frage, weil es uns im Grunde nur sagt, daß die Maschine ein operationsfähiges, algorithmisch beschreibbares System ist. Man kann aber zum Beispiel - zumindest heute - nicht behaupten, das menschliche Gehirn sei etwas ganz anderes. Insofern macht der allgemeine Hinweis auf die Simulierbarkeit beliebiger Maschinen durch Turingmaschinen noch nicht die spezifische Differenz von Gehirn und Maschine sichtbar. Die bloße Versicherung, Gehirnprozesse seien, eben als Gehirnprozesse, von prinzipiell anderer Art als Maschinenoperationen, wofür ich - unbeschadet der Eigenständigkeit logischer Geltungszusammenhänge - nicht die geringste Legitimation sehe, wäre nur wieder die genannte Petitio. Sollte also auch das Gehirn turingmodellierbar sein, so würde das bedeuten, daß die Möglichkeiten von Turingmaschinen noch völlig unabsehbar sind. In der Tat zeigt die noch in Anfängen steckende Theorie der neuronalen Netze, mit welchen prinzipiellen Verifikationsproblemen (auf die Stekeler-Weithofer zu Recht hinweist ((Ste 13))) und damit auch Unvorhersagbarkeiten man in diesem Feld zu rechnen hat. Ähnliche Konsequenzen legen sich aufgrund der Ergebnisse der Komplexitätstheorie nahe (z.B. Rice's Theorem bezüglich des Input-Output-Verhaltens von Systemen). Ich will damit nicht etwa sagen, daß Unbestimmtheit hier das letzte Wort wäre, sondern nur, daß man sich Forscher Unmöglichkeitsaussagen in Sachen KI enthalten sollte (vgl. z.B. Krämers Titelwahl "Das Scheitern der universalen Denkmaschine", ferner ((Ste 13)), vorsichtiger ((Va 33))).

((24)) Ich selbst möchte - im Sinne eines Gedankenexperiments - einmal eine anderslautende Konjektur wagen: Eine 'Denkmaschine', die in der Lage sein soll, Gödels Beweis zu führen, müßte, scheint mir, über die generelle Kompetenz verfügen, eine Metasprache zu ihrem formalen System zu konstruieren. Das ist sicher nur eine notwendige, keine hinreichende Bedingung. Varga hat recht, wenn er auf die grund-

sätzliche "Nichteliminierbarkeit des impliziten Kontexts" hinweist ((Va 21)), durch den ein Formalsystem überhaupt erst eingeführt werden kann. Ich abstrahiere jetzt davon, indem ich all dieses mit an den Programmierer delegiere, der den Zeichenvorrat, die Formationsregeln, Axiome, Deduktionsregeln usw. des Systems spezifiziert. Die fragliche Maschine müßte nun befähigt sein, metasprachliche Mittel zu konstruieren, um über ihre Strukturen und Operationen Auskunft geben zu können. (Das setzt natürlich auch schon eine entsprechende 'Intention' voraus, die ich hier - um auf diese Weise eine Minimalbedingung herauszupräparieren - ebenfalls an den Programmierer (z. B. die Evolution) delegiere.) Auf die Frage 'Ist G beweisbar?' könnte eine solche Maschine auf der Metaebene zunächst die Annahme, G sei beweisbar, formulieren und sodann die bekannten Beweisschritte vollziehen: Aufgrund der Annahme existiert die Gödelzahl eines Beweises für G. Dies kann über die Gödelisierung zahlentheoretisch formuliert und formal in S repräsentiert werden. An dieser Stelle wird also der Übergang von G zu g und weiter zu 'g' vollzogen und damit die Referenz zwischen G und 'g' ins Spiel gebracht und für den Beweis eingesetzt: Die Annahme der Beweisbarkeit von G wird jetzt mit Hilfe von 'g' formuliert und führt so zu einem Ausdruck, der in S formal beweisbar ist und die inhaltliche Aussage 'G ist beweisbar' repräsentiert. Wird dieser nun mit G verglichen,¹⁰ so ergibt sich, daß er gerade äquivalent -G ist. Dies erweist sich als Widerspruch zur Annahme, die damit widerlegt ist. - Etwas von der Art stelle ich mir als metasprachlichen Bericht der Maschine vor und hoffe nur, die Grenze zur Science-fiction hin damit noch nicht überschritten zu haben. Möglicherweise kann die von Blau entwickelte 'Reflexionslogik' (vgl. ((Bl 5))) hier zu Präzisierungen führen.

((25)) Man beachte: Wesentlich ist nach dem Gesagten, daß es wirklich zwei Sprachebenen gibt (so zu Recht ((Kr 6))): Die auf der Metaebene formulierte Mitteilung der S-Unbeweisbarkeit von G ist dann kein selbstreferentieller Satz mehr wie im Fall des objektsprachlichen Satzes G (vgl. ((10))) - andernfalls könnte sie (Konsistenz vorausgesetzt) nicht beweisbar sein. Die Beziehung zwischen den beiden Sprachebenen ist aber semantischer Natur (was nicht mit der hier als semantisch charakterisierten Beziehung zwischen 'g' und G zu verwechseln ist). Wenn man bedenkt, daß eine Maschine zur Etablierung einer neuen Sprachebene zudem in der Lage sein müßte, neue Zeichen zu generieren, so würde das bedeuten, daß sie auch über eine gewisse pragmatische Kompetenz verfügen müßte, kurz: Derartige Leistungen gehen eindeutig über den Bereich des rein Formalen hinaus. Aber gehen sie damit auch über das hinaus, was einer Maschine grundsätzlich möglich ist? Der Begriff des Algorithmus muß, scheint mir, auch auf die Möglichkeit semantischer und pragmatischer Maschinenstrukturen hin durchdacht werden.

((26)) Das damit sichtbar werdende viel weiter reichende Projekt der KI war hier nicht Gegenstand der Untersuchung. Die Frage etwa, ob das Beweisen von Theoremen der entscheidende Punkt menschlicher Intelligenz sei ((Al 11)), muß hier nicht gestellt und beantwortet werden. Mit der The-

matisierung eines bestimmten Problems ist ja nicht geulnet, daß es auch noch andere, nicht weniger wichtige Fragen auf diesem Gebiet gibt. Und weiter: Für die Klärung der Struktur einer wirklichen 'Intelligenz' geben die hier vorgebrachten Überlegungen keine näheren Anhaltspunkte:¹¹ Was ist z. B. Empfindung,¹² Intuition, Verstehen, Intentionalität, Reflexion, Urteilskraft, Bewußtsein, Selbstbewußtsein, Ichidentität, Weltverständnis, Intersubjektivität usw.? Bevor sich die KI solcher Fragen annimmt, wäre deren philosophische Klärung vonnöten, die ihrerseits, scheint mir, bisher kaum ansatzweise geleistet ist, vgl. ((El 8 f)). Möglicherweise sind diesbezüglich auch die von Fuhrmann erwähnten 'parakonsistenten' Logiken von Bedeutung ((Fu 17 f)), die zumindest für die Logik dialektisch-semantischer Inferenz eine zentrale Rolle spielen.¹³

((27)) Ich möchte abschließend noch einmal unterstreichen, daß das hier erprobte, völlig neue Publikationsverfahren, so aufwendig und anstrengend es für die Beteiligten ist, doch eine neue Qualität wissenschaftlicher Kommunikation darstellt. Die gewiß streitbaren Beiträge meiner Kritiker haben Akzente gesetzt, die zu einem neuen Durchdenken der Sache und zu weiterer Verdeutlichung nötigten. Ich wünschte mir sehr, daß diese Praxis Schule macht.

Anmerkungen

1 Vgl. auch D. van Dalen, Algorithms and Decision Problems: A Crash Course in Recursion Theory, in: Handbook of Philosophical Logic, Vol. I, ed. D. Gabbay/F. Guenther, Dordrecht/Boston/Lancaster 1983, S. 461.

2 Zur Verdeutlichung des Grundsätzlichen dasselbe in salopper Schreibweise: Nur im Fall des selbstreferentiellen Satzes G, mit $G \leftrightarrow \neg(G \text{ ist beweisbar})$, gilt, daß 'G ist beweisbar' äquivalent -G ist. - Ich "identifiziere" -G also nicht mit G, sondern mit 'G ist beweisbar' - offenbar ein Versehen in ((Fu 12)). - Zur Beachtung: Statt 'G ist beweisbar' müßte ich genaugenommen "'G' ist beweisbar" schreiben. Ich verzichte in diesem Text auf die Anführung von G, da die Gefahr eines Mißverständnisses nicht besteht.

3 Natürlich behaupte ich nicht, der Grund für die Unbeweisbarkeit einer wahren Formel in irgendeinem System sei generell deren Selbstbezüglichkeit, vgl. ((Rh 7)).

4 Deren Eindruck, ich würde "nur die Ableitungsregeln zu den Beweismitteln zählen" ((Rh 10 u. Anm. 10)) im übrigen unzutreffend ist, vgl. z. B. ((Wa 34, die letzten vier Sätze)).

5 Aus diesem Grund sollte ich hier nicht von 'semantischer Selbstreferenz' sprechen; denn das legt das Mißverständnis nahe, es handle sich hierbei um eine Aussage mit semantischem Inhalt (vgl. ((Bl 5))).

6 Ob in S selbst ein von G verschiedener Satz G' möglich wäre, der bei inhaltlicher Deutung dasselbe über G aussagt wie G selbst, als solcher aber nicht mehr selbstreferentiell ist und insofern grundsätzlich beweisbar sein könnte (vgl. ((Wa Anm. 48)), ((Be 3 ff))), bleibt eine interessante und weiterhin offene Frage. (Gemäß (DLG) dürfte G' auch nicht äquivalent G sein.)

7 Die mit der Wahrheit von G verbundenen sprachlichen Aspekte lasse ich hier ganz außer Betracht.

8 Vgl. Fuhrmanns Hinweis auf die "Überabzählbarkeit" solcher Wahrheiten ((Fu 14)).

9 Vgl. R. M. Smullyan, Languages in which Self Reference is Possible, in: J. Hintikka (ed.), The Philosophy of Mathematics, Oxford University Press 1969, S. 66, R1, R2.

10 Darauf muß man freilich erst einmal kommen, und dabei spielt natürlich die Intuition eine entscheidende Rolle. Thema ist hier aber nicht die Beweisheuristik, sondern nur die prinzipielle Möglichkeit eines metasprachlichen Beweisgangs für die Maschine.

11 Roths gewiß als originell zu bezeichnende Aussage, daß ein Modell des Geistes "eine ausgesprochen schwache Leistung" der KI-Forschung wäre ((Ro 5)) und "maschinelles Denken dem menschlichen a priori überlegen" sei ((Ro 7)), will ich daher auch auf sich beruhen lassen.

12 Hierzu D. Wandschneider, Anfänge des Seelischen in der Natur in der Deutung der hegelschen Naturphilosophie und in systemtheoretischer Rekonstruktion, in: M. J. Petry (ed.), Hegel und die Naturwissenschaften, Stuttgart 1987.

13 Vgl. die in ((Wa Anm. 33)) genannte Arbeit (ersch. vorauss. 1990), 8. Abschn.

Adresse

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

H. Tetens : Te

M. Varga v. Kibéd : Va

Replik von D. Wandschneider : War

((2)) Im folgenden möchte ich zunächst in I das Anliegen von D. Wandschneider referieren, dann in II die Hauptpunkte der Kritiken skizzieren, in III die Replik von D. Wandschneider kritisch beleuchten und schließlich in IV eine Zusammenfassung der mir wesentlich scheinenden Ergebnisse der schriftlichen Diskussion angeben.

Vorab möchte ich jedoch schon sagen, daß ich alle Kritiken mit großem Gewinn für mich gelesen habe. Einige von ihnen sind knapper und berühren nicht alle Punkte von Wa, andere sind sehr umfangreich. Manche zeigen ein hervorragendes didaktisches Geschick, die Gödel-Theoreme korrekt darzustellen. Schließlich gehen die Erläuterungen etwa von Ste weit über den gesteckten Rahmen hinaus. Es ist mir unmöglich, alle Kritiken en detail zu würdigen. Ich werde mich auf die Hauptpunkte beschränken und - die Autoren mögen mir dies verzeihen - die hierzu vorliegenden Stellungnahmen nur sehr summarisch referieren.

Es ist mir ein Bedürfnis, den Kritikern zu danken. Durch ihre Beiträge ist das von Wandschneider angesprochene Thema so substantiell behandelt, daß der Leser wirklich einen Gewinn davon hat.