Von Geistern und Maschinen. Ein Dialog über J.R.Lucas, "Minds, Machines and Gödel"

Dr. Christian Mann

11.07.2001

Die Szene

Wir befinden uns in einer Bar, irgendwann in ferner Zukunft, den Ort lassen wir offen. Die Bar ist eindrucksvoll bestückt: Neben diversen mehr oder minder alkoholischen Getränken findet man auch eine beeindruckendes Sortiment von Schmieröl. Ansonsten entspricht die Einrichtung eher einem alten englischen Pub, d.h. es sind auch verschiedentlich Regale zu finden, auf denen allerlei alte Folianten, wie man sie dieser Tage auf den Flohmärkten zu Hauf' findet, ihr "Gnadenbrot" fristen.

Unter anderem liegen ältere Ausgaben von Hofstadter, Putnam und Searle herum, sowie vereinzelte Jahrgänge von Mind *und* Philosophy*.*

An den Tischen hat sich ein rechtes Panoptikum verschiedenartiger Gestalten versammelt. An der Bar sitzen zwei offenbar miteinander wohl vertraute Gäste in philosophischer Unterhaltung versunken...

1. Akt – die These

C3PO

...Ich hab' neulich mal wieder in den alten Büchern aus dem 20. Jahrhundert gestöbert und bin dort auf einen ganz interessanten Text von einem Herrn *Lucas* gestoßen; beeindruckender Titel: "Minds, Machines and Gödel", ziemlich alt übrigens, von 1961.

R2D2

Du meine Güte!
Damals konnten ja die Computer kaum Schach spielen... (Und wenn sie es versucht haben, dann konnte man mit der Abwärme fast eine kleine Stadt heizen!)

C3PO

Nichtsdestotrotz haben wir dort eine recht interessante These: Lucas behauptet, daß es *grundsätzlich* keine Maschine mit einem menschlichen Geist aufnehmen könne!

R2D2

Moment: was soll das heissen?

C3PO

Na ja, daß eben der menschliche Geist grundsätzlich nicht durch ein "mechanistisches" Modell abbildbar wäre, und zwar – nun kommt der Clue des Ganzen – wegen des berühmt-berüchtigten Gödelschen Theorem, welches bekanntlich besagt, daß es in jedem konsistenten System welches "stark genug ist, um einfache Arithmetik zu produzieren"[a] auch Sätze gibt, die zwar offensichtlich wahr, aber nicht innerhalb des Systems beweisbar sind!

[a] J.R.Lucas, "Minds, Machines and Gödel", S.112, in: "Philosophy" 36, S.112 ff; die Übersetzung dieses wie auch aller folgenden Zitate stammt vom "Chronisten"...

R2D2

Und was hat das mit der Frage eines mechanistischen Modells des Geistes zu tun?

C3PO

Eins nach dem Anderen! Fangen wir mit dem Argument an...
Ist der Satz "Dieser Satz ist unbeweisbar-im-System" wahr oder falsch?

R2D2

Na, ja...

C3PO

...Angenommen, er wäre beweisbar-im-System:
Dann müßte er – da beweisbar – eben wahr sein, d.h. aber er wäre unbeweisbar-im-System, d.h. er wäre eben *nicht* wahr! Das wäre also ein Widerspruch!

R2D2

Na dann muß er ja eigentlich unbeweisbar-im-System sein...

C3PO

...also (einsehbar) wahr, aber ohne Beweis!
Anders betrachtet: Wäre der Satz einfach *falsch*, dann kommen wir zwangsläufig zu einem Widerspruch, nämlich dem eben genannten, der sich aus der dann notwendigen Beweisbarkeit-im-System ergibt... Die Unbeweisbarkeit-im-System (mithin die Wahrheit) des Satzes ist also unmittelbar einsichtig, nur nicht mit den Mitteln des Systems als wahr erkennbar.

R2D2

So weit, so gut.
Aber können wir nun auf die Frage der Vergleichbarkeit von Geist und Maschine zurückkommen?

C3PO

Nun, wir können eine Maschine im klassisch-kybernetischen Sinne als "Instanziierung" eines formalen Systems ansehen. *Wenn* eine solche Maschine nun (a) konsistent und (b) einfacher Arithmetik fähig ist, *dann* gibt es mindestens einen – gödelschen – Satz, den diese Maschine **nicht *als wahr* aufstellen kann**, obwohl der menschliche Geist ihn ohne weiteres – wie gesehen – als wahr einsehen kann.
Ergo kann eine solche Maschine kein vollständiges und adäquates Modell des menschlichen Geistes darstellen!

R2D2

Also, so einfach ist das doch nun auch wieder nicht:
Im Prinzip erkennt doch auch der menschliche Geist diesen Satz nur als "wahr, aber unbeweisbar-im-System", weil er sich in einen erweiterten Kontext begibt, d.h. in ein Metasystem zu dem vorher betrachteten System, innerhalb dessen er den "Gödelsatz" eben doch als wahr – für das untergeordnete System – beweisen kann!
Meine erste Reaktion wäre, auch die Maschine mit einem entsprechenden Mechanismus auszustatten, der das Erkennen – und als wahr deklarieren – derartiger Gödelsätze erlaubt...

C3PO

Das letztere nützt dir allerdings wenig, da es sich nun um eine neue, andere Maschine als die zuvor betrachtete handelt, die also – so sie konsistent ist, und nur konsistente Maschinen lohnt es sich überhaupt in unserem Kontext zu betrachten – ein *anderes* formales System implementiert.
Der Geist geht in die nächsten Runde mit der Konstruktion eines neuen Gödelsatzes für das neue formale System (der 2. Maschine), und wir sind so schlau als wie zuvor!
"Dank Gödels Theorem hat der Geist immer das letzte Wort."[a]

[a] J.R.Lucas, "Minds, Machines and Gödel", S.116

R2D2

Aber wenn es sich um ein standardisiertes "Kochrezept" handelt, nach dem die Gödelsätze produziert werden – und nur so können wir sicher sein, daß sie sich für *jedes* hinreichend mächtige formale System konstruieren lassen –, was hindert mich denn dann daran, eben diese Prozedur in meiner Maschine zu implementieren, und jedes Resultat der Anwendung dieser Prozedur als "wahr, aber im System unbeweisbar" zu deklarieren?

C3PO

Wie bereits gesagt, wir betrachten dann ein anderes System, als das vorherige...
Wir können die nach deinem Schema produzierten Gödelsätze als (abzählbar unendliche) Menge von Axiomen dem neuen System hinzufügen, oder auch die Produktionsregel für diese Sätze – ein menschlicher Geist kann nichtsdestoweniger dieses neue System als formales System betrachten und wiederum einen zugehörigen Gödelsatz hinzufügen!
Das mechanistische Modell muß stets *finit und definit* sein, um implementierbar zu bleiben. Es instanziiert damit immer nur *ein*, und zwar ein *bestimmtes* System, welches stets dem Gödelschen Theorem unterworfen bleibt...

R2D2

Nur als Marginalie bemerkt:
"Der Menschliche Geist", den du hier so vehement lobpreist, dürfte in der hier zur Diskussion stehenden Qualität bestenfalls in 2 Promille der Individuen der Species *homo sapiens* anzutreffen sein – und auch das wohl nur im westlichen Kulturkreis (anderen Kulturen dürfte es häufig dem Sinn für den "sittlichen Nährwert" des Gödelschen Theorems ermangeln!)
Die Frage der Superiorität des Geistes über die Maschine ist mit deinen (bzw. den Lucasschen) Erörterungen noch keineswegs geklärt!

C3PO

Zugegeben, aber das war ja auch gar nicht das Ziel des Ganzen Manövers.
Es handelt sich um eine Art *Spiel*, welches zeigen soll, daß der mechanistische Ansatz, den menschlichen Geist über ein formales (in der Regel Computer-)Modell zu verstehen, im Grundsatz unhaltbar ist.
Es geht nicht darum zu beweisen, daß der Geist "besser" ist als die Maschine, sondern nur darum, daß er *anders* ist.
"Sicherlich, die Maschine kann vieles tun, was ein menschlicher Geist nicht kann: Aber wenn es etwas gibt, das die Maschine nicht tun kann, aber der Geist kann es, dann – wie immer trivial die Angelegenheit auch sein mag – können wir beides nicht gleichsetzen, und können auch niemals hoffen, ein mechanisches Modell zu besitzen, welches den Geist adäquat repräsentiert." [a]

[a] J.R.Lucas, "Minds, Machines and Gödel", S.118

2. Akt – Noch ein Theorem...

C3PO

Es gibt da aber noch ein kleines Problem:
"Gödels Theorem gilt nur für konsistente Systeme. Alles was wir *formal* beweisen können, ist daß *wenn* ein System vollständig ist, dann ist der [darin enthaltene] Gödelsatz unbeweisbar-im-System. Um kategorisch sagen zu können, daß der Gödelsatz unbeweisbar-im-System und daher wahr ist, müssen wir nicht nur mit einem konsistenten System arbeiten, sondern wir müssen auch in der Lage sein zu sagen, daß es konsistent ist. Und, wie Gödel in seinem zweiten Theorem – einem Korollar zu seinem ersten – zeigt, es ist unmöglich in einem konsistenten System zu beweisen, daß ebendieses System konsistent ist." [a]
Der menschliche Geist muß also, um Gödels (erstes) Theorem "gegen" die Maschine anwenden zu können, feststellen können, daß diese Maschine (bzw. das ihr zugrunde liegende formale System) konsistent ist. Dies ist aber nicht absolut beweisbar, sondern nur im Rahmen des Systems des menschlichen Geistes.
Der menschliche Geist kann also nur feststellen, daß die Maschine konsistent ist – vorausgesetzt, der Geist ist dies auch!

[a] J.R.Lucas, "Minds, Machines and Gödel", S.120

R2D2

Man könnte also – dem Vorschlag Putnams folgend – den Geist als Maschine, allerdings als *inkonsistente* ansehen, und nun mit ihm gleichzuziehen versuchen, indem man eine ebenso inkonsistente Maschine als Modell entwickelt...
...was aber erheblich an Charme vermissen ließe, da ein inkonsistentes qua "kreatives" System eben so ziemlich aller Vorteile, die man sich von einem mechanischen Modell des Geistes verspricht, verlustig geht!
Zudem erscheint der menschliche Geist zugegebenermaßen als zwar fehlerträchtiges, aber nichtsdestotrotz konsistentes System.

C3PO

Ganz so schlimm ist die Lage auch noch nicht, daß wir den menschlichen Geist als Grundsätzlich inkonsistent ansehen müßten...
Gödels Theorem besagt erst einmal, daß innerhalb eines konsistenten Systems ein Satz, der die Konsistenz eben dieses Systems behauptet, nicht – formal – bewiesen werden kann. Dies gilt natürlich auch für den menschlichen Geist – wenn er denn eine Maschine wäre. "Für einen Geist, der keine Maschine ist, folgt eine derartige Konklusion keineswegs" [a]...

[a] J.R.Lucas, "Minds, Machines and Gödel", S.124

R2D2

...Was nun eine *petitio principii* reinsten Wassers ist...

C3PO

...Was wiederum nichts daran ändert, daß Gödels Theorem nur besagt, ein Geist könne keinen *formalen* Konsistenzbeweis für ein System innerhalb des Systems führen.
Nichts spricht dagegen, aus dem System herauszutreten und es "von außen" zu betrachten; und nichts spricht dagegen, mehr oder minder informelle Argumente für die Konsistenz auch "innerhalb" des Systems zu suchen (und vor allem auch: zu finden).
Der menschliche Geist nun ist *sich seiner selbst bewußt*, weshalb er sich aus sich selbst heraus in einem Reflexionsakt betrachten und sein eigenes Handlen analysieren kann; er ist somit in der Lage – zwar nicht formal zu beweisen, aber doch – zu *entscheiden*, daß er konsistent ist! Gödels Theorem unterscheidet insofern auch *selbstbewußtes Wesen* von *inanimierten Objekten*[a].
Der Kern des Gödelschen Theorems liegt in dessen Selbstbezüglichkeit, und eben diese ist auch ein wesentliches Merkmal des Phänomens des Selbstbewußtseins. "Bei den ersten und einfachsten Versuchen zu philosophieren, wird man von Fragen ergriffen, ob, wenn man etwas weiß, man auch weiß, daß man es weiß, und worüber man nachdenkt, wenn man über sich selbst nachdenkt, und was denn da denkt."[b] Nach einer Weile des "Herumirrens" lernt der Geist dann, den rekursiven Frageprozeß zu "beherrschen", ohne in einen malignen infiniten Regress zu verfallen...

[a] [*nota bene:* "inanimate" = "geistlos"]
[b] J.R.Lucas, "Minds, Machines and Gödel", S.124

R2D2

Aber was unterscheidet dann "formal" den Geist von der oben avisierten Maschine, die in der Lage ist, systematisch Gödelsätze zu produzieren bzw. zu erkennen, und dann als wahr zu deklarieren?

C3PO

"[...] Ein bewußtes Wesen kann mit derartigen Gödelschen Fragen in einer Weise umgehen, die eine Maschine nicht beherrscht, da ein bewußtes Wesen sowohl sich selbst als auch sein Handeln in Betracht ziehen kann, und doch [dadurch] kein Anderes wird als das, was das Handeln vollzieht. Eine Maschine kann zwar gewissermaßen ihr eigenes handeln "berücksichtigen", aber sie kann dies nicht "miteinbeziehen" ohne eine andere Maschine zu werden, nämlich die alte Maschine mit einem "neuen Teil". Aber es ist inhärenter Bestandteil der Idee eines bewußten Geistes, daß er über sich selbst reflektieren und sein eigens Handeln kritisieren kann, und es ist kein zusätzlicher Bestandteil hierfür notwendig: Er ist stets schon vollständig, und hat keine Achilles-Ferse."[a]

[a] J.R.Lucas, "Minds, Machines and Gödel", S.125

R2D2

Turing[a] hat gegen derartige Thesen entgegengehalten, daß es so etwas wie eine "kritische Masse" auch für Intelligenz bzw. Bewußtsein gibt, d.h. daß erst ab einer bestimmten Komplexität eines Systems mit – dann als emergent zu verstehenden – Phänomenen wie Bewußtsein etc. zu rechnen ist.
Die meisten Gehirne, wie auch (zu Turings und Lucas' Zeiten) alle Maschinen liegen dabei im "sub-kritischen" Bereich. Es ist doch durchaus vorstellbar, daß ab einer gewissen Komplexität ein "Qualitätssprung" einsetzt, der auch einer Maschine zu weniger berechenbarem, gleichwohl aber konsistentem Verhalten verhilft!?

[a] A.M.Turing, "Computing Machinery and Intelligence", in: "Mind", 1950, S.433 ff.

C3PO

Moment:
Der Witz an dem mechanistischen Unterfangen, den Geist über ein mechanisches, formales Modell zu verstehen, ist doch eben die Determiniertheit des mechanischen Modells! "Wenn der Mechanist eine Maschine produziert, die [nicht mehr aus ihren initialen Zuständen, sowie ihrer Konstruktion heraus verstehbar ist], dann ist das nicht länger eine Maschine im Sinne unserer Diskussion, gleichgültig, wie sie auch immer konstruiert worden ist. Wir sollten dann eher sagen, daß er einen Geist geschaffen hat[...]"[a]
"Kurz gesagt, faktisch ist jedes System, daß nicht von Gödels Fragestellung betroffen ist, *eo ipso* keine Turing-Maschine, i.e. keine Maschine im hier zugrunde liegenden Sinne."[b]

[a] J.R.Lucas, "Minds, Machines and Gödel", S.126
[b] J.R.Lucas, ebd.

R2D2

Das letztere heißt natürlich, daß wir die ganze bisherige Zeit über ein "Scheingefecht" geführt haben:
Du erklärst ja hiermit *per definitionem* jedes System, welches die von dir als Voraussetzung für die Kommensurabilität mit dem menschlichen Geist in Anschlag gebrachten Leistungen erbringen kann, als *keine Maschine im Sinne des mechanistischen Ansatzes*.
Der Mechanist hat also in deinem "Spiel" *a priori* keine Chance!
Was du allerdings voraussetzt, ist, daß das mechanistische Programm mit einer Maschine, die – ähnlich dem menschlichen Geist – nicht-deterministisch handelt, tatsächlich *ad absurdum* geführt wäre. Es wird nicht berücksichtigt, daß doch eventuell auch diese Maschine (und damit auch der ihr kommensurable, menschliche Geist) deterministisch sein könnte – nur eben auf einer Komplexitätsstufe, die für einen "unbewaffneten" (i.e. nicht von Maschinen unterstützten) menschlichen Geist nicht mehr adäquat erfaßbar ist!?

3. Akt – Einige Einwände...

C3PO

Um es nochmal klarzustellen:
Der Sinn des ganzen Papiers von Lucas ist der Nachweis, daß der menschliche Geist *keine* wie auch immer komplexe *Turing*-Maschine ist!
Daß hierfür der erheblich komplexere Weg über Gödels Theoreme beschritten wurde (anstatt Turings eigenen, recht einfachen Beweis der Beschränktheit von Turing-Maschinen heranzuziehen), liegt in dem Aspekt der *Wahrheit* eines Satzes begründet, der in Gödels Theoremen wie auch generell im menschlichen Geist eine wesentliche Rolle spielt.
Turings Satz ist ein rein syntaktisch erklärbares Phänomen, Gödels Theoreme hingegen "leben" von der Semantik!
Turing betont denselben – negativen – Aspekt wie Gödel, wenn es darum geht nachzuweisen, was eine Maschine *nicht* kann. Gödel gibt uns darüber hinaus einen zweiten, positiven Aspekt, indem er erlaubt darzulegen (letztlich anhand des dem menschlichen Geist inhärenten Begriffs der *Wahrheit*), daß der menschliche Geist etwas *kann*, was die Maschine eben nicht kann...

Ich kann mir allerdings immer noch nicht vorstellen, daß das alles unwidersprochen hingenommen wurde – auch nicht im 3. Viertel des 20. Jahrhunderts!
Es gibt doch eine ganze Reihe *prima facie* Einwände, die vorgebracht werden könnten, als da seien

- die Idealisierung, die in Lucas' Verwendung von Gödels Theoremen implizit vorgenommen wird:

  Weder der menschliche Geist, noch die konkurrierende Maschine sind in der angenommenen Weise unlimitiert; der Geist steckt in einem sterblichen Körper, irrt sich ganz gerne einmal, und ist – wie bereits bemerkt – im seltensten Fall in der Lage, die Gödelschen Theoreme zu verstehen geschweige denn derartige Sätze gezielt hervorzubringen; die Maschinen sind – auch heute noch – keineswegs mit unendlichem Arbeitsspeicher gesegnet (i.e. sie sind keine idealen Turing-Maschinen), und haben auch nicht "das ewige Leben"!

- die Frage, ob dem Prozeß des iterierten "ausgödelisierens" der Maschine durch den Geist nicht doch Grenzen gesetzt sind:

  Ich meine mich zu erinnern, bei Hofstadter in "Gödel, Escher, Bach" etwas über ein ganz interessantes Theorem von Church und Kleene gelesen zu haben, welches besagt, daß sich nicht uneingeschränkt Namen für transfinite Ordinalzahlen finden lassen; vorgeblich läßt sich ja nicht nur ein "einfaches" System ausgödelisieren, sondern ebenso ein System, welches die besagte Funktion zur Produktion/Identifikation von Gödelsätzen enthält!? "Im Kern ist dies der Schritt von ω, der unendlichen Folge von durch den Gödelisierungsoperator produzierten Gödelsätzen, zu $\omega + 1$, der nächsten transfiniten Ordinalzahl"[a]; der Geist hat also möglicherweise Schwierigkeiten, *ad infinitum* weiterzumachen...

- Zuguterletzt wäre da noch die Frage, woher du denn *weißt*, daß der Geist *tatsächlich* immer neue Gödelsätze für jedes erweiterte System finden kann? – Schließlich bist du bisher einen *Beweis* für diese Annahme schuldig geblieben (und wenn du ihn denn erbringen könntest, kann ich ihn auch in der Maschine implementieren, mithin eine Maschine bauen, die ihrerseits die von dir exclusiv für den menschlichen Geist reklamierten Leistungen – nämlich des "Ausgödelisierens" *jeder* konkreten Maschine – erbringen kann...)

[a] J.R.Lucas, "Minds, Machines and Gödel: A Retrospect", S.110, in: Millican/Clark (hrsg.), "Machines and Thought, Vol.1, Oxford 1996, S.103 ff.

C3PO

Die meisten dieser gängigen *prima-facie*-Einwände hat Lucas selbst behandelt (ich habe allerdings nur ein doch erheblich späteres Dokument vom Ende des 20. Jh. gelesen[a].)
Was nun den ersten Einwand angeht, so ist wohl kaum anzunehmen, daß der Maschine in unserem Spiel deutliche Vorteile daraus erwachsen, daß sie – wie auch der menschliche Geist – *de facto* nur auf begrenzte Fähigkeiten zurückgreifen kann?

[a] vgl. J.R.Lucas, "Minds, Machines and Gödel: A Retrospect"

R2D2

Wie man's nimmt:
Wenn der Geist nur eine endliche Zeit lebt, so kann ich doch immer eine ebenso endliche Maschine bauen, die den gesammten Output eines jeden gegebenen Geistes vollständig (zugegeben: re-)produzieren kann, mithin auch deinen Dialog mit einer anderen Maschine!

C3PO

Dies sei dir gerne konzediert, allerdings – wie du bereits bemerkt hast – geht dies nur *ex post facto*! Das Ziel des Mechanisten ist ja, eine Maschine zu bauen, die nicht nur den Output produziert, den ein Geist zuvor produziert hat, sondern auch den Output, den ein Geist erst noch produzieren wird...

R2D2

Und was ist mit Hofstadters transfiniten Ordinalzahlen?

C3PO

Auch das wird dem Mechanisten nicht sonderlich weiterhelfen:
So, wie Lucas das "Spiel", den Wettbewerb konzipiert hat, ist nicht der Geist daran, stets die nächsthöhere Stufe zu erklimmen, sondern es muß im Gegenteil die Maschine in "Vorleistung" treten, indem sie eben als Antwort auf den Gödelsatz auf einer bestimmten Stufe die besagte Funktion implementiert, die gleich alle Gödelsätze dieser Stufe produziert & für wahr erklärt.
Das ist aber genau der besagte Schritt von ω zu $\omega + 1$!
Alles, was der Geist dann tun muß, ist, einen weiteren Gödelsatz auf der soeben erklommenen Stufe zu generieren...

R2D2

So weit, so gut (der schlecht, wenn man die mechanistische Perspektive einnehmen möchte...) Kommen wir zum dritten Punkt:
Dein (bzw. des Herrn Lucas) Ziel ist es ja, *nachzuweisen*, daß der menschliche Geist etwas kann, was eine Maschine *per se* nicht kann. Dafür mußt du aber auch *nachweisen*, daß der Geist die besagten Fähigkeiten des stets erneuten Ausgödelisierens der Maschine besitzt!
Bis dato erscheint mir dein diesbezüglicher Optimismus vor allem darauf zu beruhen, daß Gödel dir ein einfaches "Kochrezept" zur Produktion von Gödelsätzen (eben dieses Rezept, welches der Mechanist in seiner Maschine zu implementieren gedenkt) an die Hand gegeben hat. Wenn du aber nun auf der Basis diese Kochrezeptes einen (formalen) Beweis für deine Annahme erbringst, so steht der Implementierung dieses Verfahrens in einer Maschine nichts mehr im Wege; die dann entstandene Maschine könnte aber eben das, was du – wie gesagt – exclusiv für den Geist reklamierst: jede beliebige Maschine ausgödelisieren!

C3PO

Klar, "Catch 22"...
Der Kern deines Argumentes ist doch die Annahme, daß jedes informelle Argument entweder formalisierbar sein muß, oder aber ungültig ist (das ist jetzt fast ein Zitat von Lucas)! Ich vertrete dagegen einen gewissen *Stil* der Argumentation, der auf Einsehbarkeit, nicht auf formalen Beweisen basiert...
"Zugegeben, wir können einem engstirnigen Mechanisten nicht *beweisen* daß wir immer weiter schreiten können. Aber wir können zu einer wohlbegründeten Zuversicht diesbezüglich gelangen, die uns – und dem ehemaligen Mechanisten, wenn er denn vernünftig und nicht engstirnig ist – gute Gründe für die Zurückweisung des Mechanismus liefert."[a]

[a] J.R.Lucas, "Minds, Machines and Gödel: A Retrospect", S. 113

R2D2

Hoppla!
Der Kern *deines* letzten Argumentes ist ja wohl die Annahme, daß *meine* Annahme nicht stimmt, und daß jeder, der *deine* Annahme nicht teilt, *per se* mal "engstirnig" sein muß!?

C3PO

Pardon, aber die Engstirnigkeit zeigt sich in der Weigerung (oder auch Unfähigkeit), die Perspektive zu wechseln!
Nimm ein Beispiel:
"In der Argumentation gegen einen Finitisten, der das mathematische Prinzip der vollständigen Induktion nicht akzeptiert, kann ich auf einer Meta-Ebene erkennen, daß ich, wenn er $F(0)$ zugibt und es gilt $(\forall x)(F(x) \rightarrow F(x+1))$, ohne Furcht vor Widersprüchen $(\forall x)F(x)$ behaupten kann. Ich kann darauf vertrauen, auch wenn ich keinen finitistischen Beweis dafür habe. Ich kann nur – *vis-a-vis* dem Finitisten – darauf hinweisen, daß *wenn* er meine Behauptung in irgend einem spezifischen Fall bestreitet, ich ihn zurückweisen kann." [a]
Ich verwende Gödel in ähnlicher Weise, um den Mechanisten zurückzuweisen; das Argument lebt aber davon, daß diese Zurückweisung eine *taktische* und keine *strategische* ist (ob das Argument als strategisches ähnlich effektiv wäre, darob bin ich mir nicht so sicher...)

[a] J.R.Lucas, "Minds, Machines and Gödel: A Retrospect", S. 114

R2D2

"Die Idee eines völlig intuitiven, unformalisierbaren Argumentes erweckt Verdacht: Wenn es überzeugen kann, ist es vermittelbar, und wenn es vermittelbar ist, dann kann es formuliert und in formalen Begriffen ausgedrückt werden."" [a]

[a] J.R.Lucas, "Minds, Machines and Gödel: A Retrospect", S. 114

C3PO

Natürlich, jedes Argument kann formalisiert werden (ob es dadurch aber überzeugender wird, lassen wir mal dahingestellt...)
Auch Gödels Argument für den Beweis, daß der besagte Gödelsatz unbeweisbar-im-System ist, ist formalisierbar! Das Ergebnis ist allerdings lediglich, daß *wenn* das System konsistent ist, *dann* ist der Gödelsatz ("Dieser Satz ist unbeweisbar-im-System") wahr:

$$\vdash Cons(ENT) \rightarrow G^{a}$$

Da nun G nicht beweisbar ist, muß dies umgekehrt auch für *Cons(ENT)* gelten – und ebendies ist Gödels zweites Theorem.

[a] Wobei *Cons(ENT)* ein Satz, der die Konsistenz der *Elementary Number Theory (= ENT)* behauptet, und G der besagte Gödelsatz ist.

R2D2

Gut, aber was ist mit Putnams Einwand, daß der Schluß von

Ich kann erkennen, daß $(Cons(ENT) \rightarrow G)$

auf

$Cons(ENT) \rightarrow$ Ich kann erkennen, daß (G)

unzuläßig ist? – Schließlich benötigst du den letzteren Satz, um dann mittels Gödel zu dem Satz

$Cons(ENT) \rightarrow$ Eine ENT-Maschine kann nicht erkennen, daß (G)

zu gelangen.

C3PO

Auch Putnam scheitert an der dialektischen Natur von Gödels Argument:
Erst einmal ist es bereits der *Mechanist*, der die Konsistenz seiner Maschine behaupten muß, um überhaupt mit seinem Anspruch eines adäquaten Modells des Geistes ernst genommen zu werden. "Das Gedankenexperiment, einmal begonnen, muß durchgedacht werden. Und wenn es durchgedacht ist, dann ist es auf den Hörnern eines Dilemmas aufgespießt. Entweder, die Maschine beweist in ihrem System den Gödelsatz, oder nicht: wenn sie es kann, ist sie inkonsistent, und daher nicht äquivalent mit dem Geist; wenn sie es nicht kann, ist sie konsistent, und der Geist kann daher den Gödelsatz als wahr behaupten. In beiden Fällen ist die Maschine nicht äquivalent mit dem Geist, und die mechanistische These scheitert."[a]

[a] J.R.Lucas, "Minds, Machines and Gödel: A Retrospect", S. 119

R2D2

Letztlich bleibt also alles beim alten:

- Die Maschine hat bei dem Spiel keine Chance, da sie – zumindest nach deiner Meinung – ein bestimmtes finites, formales System implementieren muß, und daher zwangsläufig die Wahrheit "ihres" Gödelsatzes nicht erkennen kann.
- Jegliche Forderung an den Geist, seinen Optimismus hinsichtlich der Möglichkeit des iterierten Ausgödelisierens der Maschine zu beweisen, wird als "Immunisierungsstrategie" der mechanistischen These zurückgewiesen.
- Statt dessen wird mit der "Einsehbarkeit" des Argumentes plädiert, die vor allem darauf basiert, daß das Argument als *per se* informell dargestellt wird...

Ich würde das durchaus als Immunisierungsstrategie seitens der mentalistischen These ansehen!

C3PO

"Gödels Theorem ist eine elaborierte Form des berühmten Kreter-Paradoxes von Epimenides."[a] Es aufzulösen bedarf es der Fähigkeit, aus dem System herauszutreten, und es quasi "von außen" zu betrachten.
Eben dies ist die wesentliche Fähigkeit des Bewußtseins, welches den menschlichen Geist auszeichnet; und eben diese Fähigkeit spreche ich tatsächlich der Implementierung eines – deterministischen – formalen Systems ab!

[a] J.R.Lucas, "Minds, Machines and Gödel: A Retrospect", S. 104

Letzter Akt – Und wie "gödelfest" ist der Geist?

R2D2

Na gut, aber wechseln wir doch einmal die Perspektive...
Wenn wir uns einmal *in* ein formales System (gleichgültig, welcher Komplexität) hineinversetzen – dein Argument ist doch, daß *dieses* System *seinen "eigenen"* Gödelsatz nicht als wahr erkennen kann, weil dafür das System "aus sich heraustreten" können müßte, um sich reflexiv selbst zu betrachten!?

C3PO

Exakt!

R2D2

Du behauptest weiter, daß der menschliche Geist im Gegensatz dazu *sehr wohl* in der Lage ist, sich selbst quasi von außen zu betrachten – Kraft der im inhärenten Eigenschaft, ein selbstbewußtes Wesen zu sein!?

C3PO

Exakt!

R2D2

Du argumentierst weiter, daß das formale System – erweitert um eine Funktion zur Produktion/Identifikation von Gödelsätzen – zwar in der Lage sei, die Gödelsätze für das *ursprüngliche* System korrekt zu behandeln, nicht aber die Gödelsätze für das *erweiterte* System!?

C3PO

Exakt!

R2D2

Wenn ein formales System aber seine "eigenen" Gödelsätze prinzipiell nicht korrekt behandeln kann, kann es dann überhaupt feststellen, daß es gerade "ausgödelisiert" wird? Kann es einen Gödelsatz von einem herkömmlichen Widerspruch unterscheiden?

C3PO

Eben nicht!
Das war ja gerade der Clou des ganzen Arguments, daß eben *dies* der Maschine nicht möglich ist, wohl aber dem menschlichen Geist... ich sehe, du beginnst meine Position allmählich zu begreifen!

R2D2

Und damit wären wir auch schon beim "neuralgischen Punkt":
Wenn der Geist *doch* ein – zwar überaus komplexes, aber nichtsdestoweniger – formales System wäre, *dann* könnte er doch *ebenfalls* den auf ihn zugeschnittenen Gödelsatz nicht angemessen behandeln! Er würde ihn ebenso wie jedes andere formale System zwar ggfls. nach formalen Kriterien erkennen können, aber eben nicht *als wahr* – mithin könnte er ihn nicht als genuinen Gödelsatz in deinem Sinne einstufen!?

C3PO

Wenn er denn ein formales System (i.e. eine Art Maschine) wäre, *dann* hättest du sicherlich recht!

R2D2

Nach deiner Prämisse ist dies natürlich nicht der Fall...
Aber wie du ja bereits zugegeben hast, du kannst nicht *beweisen*, daß der Geist jedes beliebige gegebene System "ausgödelisieren" kann, aber du bist – aus für jeden (nicht engstirnigen) Menschen prinzipiell gut einsehbaren Gründen – zuversichtlich, daß er jedes gegebene System ausgödelisieren und den dabei produzierten Gödelsatz auch tatsächlich *als wahr* erkennen kann.
Ich bin auch durchaus bereit, dir bis hier zu folgen, mit der Einschränkung allerdings, daß der Geist das gegebene System stets *von außen*, i.e. von einer Meta-Ebene aus betrachten können muß, um es ausgödelisieren zu können!

C3PO

Ich sehe schon, worauf du hinaus willst:
Deine Einschränkung scheint den Geist *selbst* aus den betrachtbaren Systemen auszuschließen, und damit ist er nicht – prinzipiell – besser d'ran, als die von ihm betrachteten Maschinen...
Du wirst dich vielleicht wundern, aber ich akzeptiere deine Modifikation voll und ganz – nur möchte ich darauf hinweisen, daß sich der Geist (eben in seiner Eigenschaft als selbstbewußtes Wesen) *sehr wohl* selbst von außen betrachten kann (mithin auch nicht in das Dilemma der Universellen Turing-Maschine rutscht, die nicht entscheiden kann, ob sie selbst terminiert, oder nicht!)

R2D2

Ich wundere mich keineswegs darüber!
Die Frage ist nur, *wie "gödelfest" ist der Geist selbst?* – Dein Optimismus hinsichtlich der Leistungsfähigkeit des Geistes rührt schließlich nicht unwesentlich von der Annahme her, daß der Geist seinen "eigenen" Gödelsatz wie den jedes anderen Systems zu behandeln in der Lage ist...
Was aber, wenn diese Annahme nicht stimmt? Was aber, wenn der Geist *doch nicht* in der Lage ist, sich selbst (notabene: konsistent!) "von außen" zu betrachten?
Kann er diesen Mangel überhaupt feststellen?

C3PO

Er wäre in diesem Falle ja eben *doch* im weitesten Sinne ein formales System, und daß er es dann nicht könnte, habe ich dir ja bereits konzediert...

R2D2

...Und nun bitte ich dich, auch fair zu sein:
Erinnere dich, daß das "Spiel", von dem hier so viel die Rede war, keinesfalls zwischen dem Geist und der Maschine (*pars pro toto*) gespielt wurde, sondern zwischen dem Geist und dem *Schöpfer* der Maschine – dem "Mechanisten". *Letzterer* versuchte dann mutmaßlich, die Maschine durch Erweiterungen stets so zu verbessern, daß sie den "Attacken" des Geistes standzuhalten in der Lage wäre...
In gleicher Weise müßte der Ansprechpartner für (z.B.) eine Maschine, die den Geist ausgödelisiert, nicht der Geist sein, sondern sein "Schöpfer"!
Der Geist kann *hinsichtlich seiner selbst* kaum mit Sicherheit entscheiden, ob er die von dir im zugesprochenen Fähigkeiten besitzt, oder nicht. Dies bleibt einer wie auch immer gearteten außenstehenden Instanz vorbehalten!

C3PO

Moment!
Ich hatte von Anfang an gesagt, daß es für meine Annahme *keinen formalen Beweis* gibt, insofern gebe ich dir dies auch gerne zu.
Nichtsdestoweniger habe ich einige gute Gründe angeführt, die mich – wie gesagt – zu der *Zuversicht* führen, des Spiel gegen die Maschine stets gewinnen zu können...

R2D2

Du hast es selbst gesagt: "Das Gedankenexperiment, einmal begonnen, muß durchgedacht werden."[a]
Meine Argumente sehen folgendermaßen aus:

- Die Maschine – oder wer auch immer erfolgreich ausgödelisiert wird – *merkt nicht*, daß sie das Spiel verliert!
- Die Zuversicht, das Spiel zu gewinnen, beruht allein auf der Fähigkeit, seinen Gegner von außen, als abgeschlossenes System, betrachten zu können.
- Um einen qualitativen Vorsprung vor *jeder* Maschine behaupten zu können, muß der Geist *sicher* sein können, daß er auch sich selbst "von außen" betrachten kann – und eben dies kann er nicht!

[a] J.R.Lucas, "Minds, Machines and Gödel: A Retrospect", S. 119

C3PO

Er kann es nicht *formal beweisen*!
Aber er kann es aus – notwendig! – informellen Gründen sehr wohl selbst wissen!

R2D2

Ja, ja:
"Wahr ist, was ich klar und deutlich einsehe", hat schon Descartes in seinen Meditationen festgestellt (aber hier ist der Schritt zur Intentionalismus-Debatte doch fatal kurz, und die wollen wir uns wohl doch für später aufheben!?)
Eben dies kann ich jedoch auch für eine Maschine in Anspruch nehmen, die den Geist ausgödelisiert – und der Geist kann *eben aufgrund der Gödelschen Theoreme* niemals *wissen*, ob ich recht habe, oder du!
Ich bin mir mit dir an sich ganz einig:
Ich sage nicht, daß es so ist – nur, daß es so sein könnte, und wir dies grundsätzlich nicht entscheiden können.

Wenn ich etwas aus unserem Gespräch gelernt habe, dann *nicht*, daß Geist und Maschine grundsätzlich verschieden sind, sondern *nur*, daß der Geist aufgrund des Gödelschen Theorems niemals entscheiden kann, ob ihm eine gegebene Maschine äquivalent ist oder nicht (wie er auch niemals entscheiden kann, ob er sich selbst quasi "von außen" betrachtet, oder nur ein unzureichendes "selbstgestricktes" Modell seiner selbst...)
Das Gleiche gilt natürlich *vice versa* für die Maschine...

C3PO

...womit wir also glücklich bei einem wohlfundierten Agnostizismus angelangt wären:

"Und wir seh'n betroffen
den Vorhang zu, und alle Fragen offen"

Eigentlich sollte es ja ein "Thesenpapier" werden...

... was herausgekommen ist, ist aber ein kleiner Dialog, der in einer vagen Zukunft handelt, und über dessen Protagonisten ich mir selbst nicht recht im Klaren bin (sind sie denn nun menschlicher oder vielleicht doch maschineller Natur?)

Der vorliegende Text ist am Rande einiger geruhsamer Urlaubstage am Wörthersee entstanden, die ich auch aber nicht nur zur Vorbereitung des Blockseminares "Maschinenintelligenz und Maschinenbewußtsein I"[1] genutzt habe. Da ich dem universitären Leben bereits seit einiger Zeit (zugunsten des schnöden Broterwerbs) entsagen muß, blieb mir auch keine andere Wahl. Infolgedessen konnte ich auch nicht auf eine größere Ansammlung von Literatur zurückgreifen, sondern mußte zur Erstellung des vorliegenden Papieres mit den beiden vorzustellenden Aufsätzen vorlieb nehmen (was sich vor allem in einem eklatanten Mangel an Primärzitaten weiterer Literatur ausdrückt – der geneigte Leser möge es mir nachsehen!) Die Form des Dialoges habe ich aus einer Reihe verschiedener Gründe gewählt:

- Der Dialog als literarische Form vertritt nur bedingt einen wissenschaftlichen Anspruch, paßt also insofern recht gut zu meinem mangelnden Zugriff auf weitere Sekundärliteratur...
- Lucas modelliert in seinen Aufsätzen ein *Spiel* von stark dialogischem Charakter:

  Ein **Mentalist** versucht durch beständig iteriertes "Ausgödelisieren" der Maschinen eines **Mechanisten** den letzteren von der Sinnlosigkeit seines Unterfangens (der Erstellung eines mechanischen Modells des Geistes zum Verständnis desselben) zu überzeugen...
- Zudem kann der Dialog als genuin *philosophische* Ausdrucksform auf eine nicht unbeachtliche Tradition zurückblicken!
- Und *last but not least*:

  Einen Dialog zu schreiben macht einfach mehr Spaß, und ich bin schließlich zu meinem Vergnügen hier am Wörthersee!

Pörtschach am Wörthersee, den 11.07.2001
Dr. Christian Mann

[1] Prof.Dr. E.Brendel / Prof.Dr. Th.Metzinger, Universität Mainz, SS 2001