

# Gibt es eine scharfe Grenze zwischen Kognitionswissenschaften und naturalistische Philosophie des Geistes?

*Marcin Milkowski, Warschau*

Es scheint, dass wenn wir nur die Naturalisten berücksichtigen, dann die Antwort auf diese Frage einfach verneinend muss: Der heutige Naturalismus bestreitet die strenge Trennung von Philosophie und Wissenschaft. Solche Überlegungen sind aber, meiner Meinung nach, etwa zu generell. Ich behaupte, dass es solche Grenze, obwohl fließende, gibt. An einem konkreten Beispiels des Problems, das von der Künstliche Intelligenz stammt, will ich die komplizierte Relationen zwischen naturalistische Philosophie des Geistes und Kognitionswissenschaften (Cognitive Science) zeigen.

## 1. Sorgen des künstlichen Intellekts

In 1969 haben John McCarthy und Patrick Hayes ein Problem gefunden. Manche behaupten, dass das Problem keine Lösung habe, andere, dass das Problem philosophisch sei. Ingenieure im Gegenteil betonen, dass es um eine technische Schwierigkeit gehe.

Das Problem ist in der logik-basierten Künstlichen Intelligenz entstanden<sup>1</sup>. Logische Zeugen sollen dazu dienen, eine adäquate Repräsentation der Welt aufzubauen, und die Satzrepräsentation in der Logik erster Stufe soll der Maschine erlauben, intelligente Entscheidungen zu treffen. Die Tatsachen in der Welt sind mithilfe der axiomatisch definierten Prädikaten ausgedrückt. Auf Grund der im Datenbank gespeicherten Axiommenge, die die Gesetze der Welt (z.B. Gesetze der Mechanik) und Handlungsarten des intelligenten Subjekts (z.B. das Verschieben von Gegenständen) beschreibt, der künstliche Intellekt sollte rationale und intelligente Entscheidungen treffen und sie gemäß handeln.

John McCarthy hat einen System der Logik erster Stufe — die sog. Situationskalkül — entwickelt, das das Wissen von Robot repräsentieren soll. Die Situationskalkül soll die Veränderungen in der Welt berücksichtigen. Die Veränderungen in der Zeit werden durch Prädikat *Besteht (Satz, Zeitliche\_Situation)* ausgedrückt. Aber hier genau stellt sich das Problem. Es wurde im originellen Fassung das Kulissenproblem (*frame problem*) genannt, weil McCarthy wollte es mithilfe des sog. Kulissenaxiomen lösen.<sup>2</sup> Die Begriffsmenge wurde mithilfe verschiedenen „Kulissen“ getrennt, indem verschiedene Kategorien entstanden wurden waren, die bestimmten, dass z.B. keine Handlung wie das Verschieben vom Gegenstände die Farbe des Gegenstands verändert.<sup>3</sup> Leider gibt es dann so viel Axiomen, wie konkreten Handlungen. Die Formalisierung ist also nicht befriedigend:

---

<sup>1</sup> Im Gegenteil zu logik-basierten Ansätzen, in der heuristischen KI hat man keine Probleme mit Wissensrepräsentation.

<sup>2</sup> P. Hayes, J. McCarthy, „Some philosophical problems from the standpoint of artificial intelligence“, in: *Machine Intelligence 4*, B. Meltzer i D. Michie (Hgs.), Edinburgh University Press, Edinburgh 1969.

<sup>3</sup> Vgl. P. Hayes, „What the frame problem is and isn't“, in: *The Robot's Dilemma. Frame Problem in Artificial Intelligence*, Z. Pylyshyn (Hg.), Ablex, Norwood 1987, S. 130.

- (1) das Vollschreiben aller Axiomen ist zeitraubend;
- (2) ein System mit vielen trivialen Axiomen ist unelegant;
- (3) eine Große Menge von Axiomen macht das System nicht leistungsfähig;
- (4) Kulissenaxiomen sind oft falsch, insbesondere in Systemen, die gleichzeitige Handlungen zulassen.<sup>4</sup>

Womit sollte man diese Kulissenaxiomen ersetzen? Die Suche nach der Antwort dieser Frage ist gerade das Kulissenproblem im weiteren Sinne.

Die Schwierigkeit ist dem gesundem Menschenverstand gar nicht bekannt. Eben kleine Kinder wissen: Wenn man die Untertasse mit einer Tasse verschiebt, dann verschiebt sich auch die Tasse. Aber in der Situationskalkül braucht man einen Axiom, oder eine Menge von Axiomen... Die Kognitionswissenschaftler sind dadurch beunruhigt: Die logische Repräsentation der Welt ist sehr schwer aufzubauen und spricht gegen die Intuitionen des Commonsense.

Diese Diskrepanz zwischen Commonsense und Situationskalkül hat, meiner Meinung nach, manche Philosophen zu den radikalen Schlussfolgerungen geführt. Die Kognitionswissenschaftler strebten nach einer technischen Lösung, während die Philosophen reduzierten das Kulissenproblem auf das Problem des induktiven Schließens.

## 2. Fliegt Tweety oder nicht?

Schnell sind verschiedene Lösungen entstanden. Hier lasse ich beiseite diese, die sehr ähnlich an den Kulissenaxiomen sind, als auch frühere Versuche, die aus der heuristischen KI entstammen (z.B. das System STRIPS von Fikes und Nilson.)<sup>5</sup> Erstens, hat McCarthy im Rahmen der Situationskalküls eine Version der nicht-monotonen Logik entwickelt, die modelliert die Schlussfolgerungen des Commonsense.<sup>6</sup> Die nicht-monotonen Logiken sind ein sehr interessantes Umfeld für sich. Ich sollte auch die Defaultlogik Raymond Reiters erwähnen, wo die Lösung besteht in der Vorraussetzungen, dass minimale Anzahl der Aspekte der Welt sich ändert. Heutzutage die interessantesten nicht-monotonen Logiken wurden von Vladimir Lifschitz<sup>7</sup> und Laura Morgenstern<sup>8</sup> entwickelt.

Zweitens, verlässt man das logik-basierte Paradigma. Z.B. Drew McDermott, der frühere Vertreter der nicht-monotonen Lösungen,<sup>9</sup> beschäftigt sich nicht mehr mit Wissensrepräsentationsforschungen, sondern prozedurale Methode entwickelt. Drittens,

<sup>4</sup> L. Morgenstern, „The Problem with Solutions to the Frame Problem“, in: *The Robot's Dilemma Revisited: The Frame Problem in AI*, K. Ford, Z. Pylyshyn (Hgs.), Ablex, Norwood, 1996.

<sup>5</sup> R. Fikes, N. Nilson, „STRIPS: A new approach to the application of theorem proving to problem solving“, *Artificial Intelligence*, 3, S. 189-208, 1971.

<sup>6</sup> J. McCarthy, „Circumscription — a form of monotonic reasoning“, *Artificial Intelligence*, 13, S. 27-39, 1980.

<sup>7</sup> V. Lifschitz, „Nested abnormality theories“, *Artificial Intelligence*, 74, S. 351-365, 1995.

<sup>8</sup> Morgenstern beschreibt das Hintergrund des Problems sehr klar; vgl. „The Problem with Solutions...“, a.a.O.

<sup>9</sup> Vgl. Drew McDermott, „We've Been Framed: Or, Why AI Is Innocent of the Frame Problem“, in: *Robot's Dilemma: The Frame Problem...*, a.a.o., S. 113, wo er behauptet, das Kulissenproblem gar kein Problem sei, also man sollte es nicht lösen. Später, in 1986, McDermott hat seinen Standpunkt um 180 Grad hemmgerissen. Vgl. dazu Steve Hanks, Drew McDermott, „Default reasoning, nonmonotonic logic, and the frame problem“, in: *Proceedings of AAAI 1986*, S. 328-333.

gibt es auch Strategien, die auf der logische Programmierung (Programmiersprache Prolog) oder Ereigniskalkül basieren.<sup>10</sup> Viertens, versucht man statistische Methoden anzusetzen.

Bevor ich kurz die nicht-monotone Logik vorstelle, möchte ich auch die mit dem Kulissenproblem verwandte Probleme erwähnen. Das Ramifikationsproblem (*ramification problem*) liegt darin, dass man implizite Darstellung der Folgen einer Handlung finden soll. Jede Handlung hat eine infinite Menge von Folgen (es ist offensichtlich bei der Satzrepräsentation des Wissens: Jeder Satz impliziert eine infinite Menge von anderen Sätzen), und keine elegante und praktische Methode kann die Folgen in Kauf nehmen. Das Qualifikationsproblem (*qualification problem*) besteht darin, dass man die notwendigen und ausreichenden Bedingungen jeder Handlung bestimmen soll. Wenn der Affe eine Banane essen soll, die Banane kann nicht durch die Scheibe getrennt werden, die Beine des Affen können nicht zum Detonator der Atombombe gebunden werden, der Affe kann sich nicht in eine Großmutter Jerry Fodors verwandeln usw. Die Lösungen des Kulissenproblems betreffen auch diese zwei verwandten Probleme. Da sind sie alle mit der Wissensrepräsentation und logischer Kalkül verbunden.

Hier bespreche ich nur nicht-monotone Logik: sie ist nicht nur interessant, aber hat vieles mit Philosophie zu tun. Die logische Repräsentation der Erkenntnis stammt aus dem Programm der logischen Philosophie: Die Kognitionswissenschaft knüpft also an die analytische Philosophie an, auch wenn es um Modelle des Schließens geht. Was ist nicht-monotone Logik? Die klassische Logik ist monoton: die Menge von Schlussfolgerungen ist monoton im Bezug auf die Menge von Prämissen. Je „mehr“ Prämissen, desto „mehr“ Schlussfolgerungen. In der nicht-monotonen Logik kann dagegen eine zusätzliche Prämisse Elemente der Menge von Schlussfolgerungen beheben. Es gibt nämlich normale, oder vorgegebene Schlussfolgerungen. Wenn Tweety ein Vogel ist, dann setzt man normalerweise, dass Tweety fliegt. Wenn wir aber hinzufügen, dass Tweety ein Pinguin ist, dann ziehen wir diese Schlussfolgerung nicht mehr.

Aus dem Standpunkt der klassischen Logik gibt es viele Enthymeme und Ellipsen in der Schlussfolgerungen des Commonsense. Sie sollen beheben werden, um der Gedankengang, das formalisiert werden kann, zu gewinnen. In der nicht-monotonen Logik einführt man die fehlenden Informationen nicht. Denn sie sind oft gar nicht zugänglich und man muss ohne sie Schlussfolgerungen ziehen. Es gibt also einige impliziten Vorraussetzungen, die so lange für wahr genommen werden, bis wir Informationen gewinnen, dass sie falsch sind.<sup>11</sup> Solche Formalisierung scheint ganz mit dem Commonsense übereinzustimmen.

Wie sieht hier also die Lösung des Kulissenproblems aus? In der nicht-monotonen Logik definiert man vorgegebene Eigenschaften der Situation, die sich unter dem Einfluss bestimmten Handlungen nicht verändern. Man muss also explizit nur die wesentlichen Veränderungen definieren. Nehmen wir z.B. Circumscription von McCarthy, also ein gut entwickeltes System von nicht-monotonen Logik.<sup>12</sup> Das Prinzip

---

<sup>10</sup> Murray Shanahan, *Solving the Frame Problem*, MIT Press, Cambridge, 1997.

<sup>11</sup> Vgl. dazu V. Lifschitz, „Logic of Common Sense“, *ACM Computing Surveys*, 27, 3 (1995), S. 343.

<sup>12</sup> Vgl. dazu Vladimir Lifschitz, „Circumscription“ in: *The Handbook of Logic in AI and Logic Programming*, D.M. Gabbay, C.J. Hogger i J.A. Robinson (Hgs.), Oxford University Press, 1994, Band 3, S. 298-352. Originelle Fassung von McCarthy: „Circumscription — a form of Non-Monotonic Reasoning“, a.a.O.; „Applications of circumscription to formalizing common sense knowledge“, *Artificial Intelligence*, 26, 3

der Circumscription ist, die Extension des Prädikats (oder einer Menge von Prädikats) auf eine besondere Weise zu beschränken. Man zeigt also Aspekte der Abnormalität des Prädikats, d.h. wann der Prädikat nicht prädiert werden kann. Z.B. die Vögel fliegen, sonst sie sind abnormal. Solche Einschränkungen, oder Circumscriptions, kann man auch in der temporalen Schlussfolgerungen mithilfe der sog. Fluents (temporal bestimmten Funktionen) benutzen.

Doch die Voraussetzungen in der Circumscription von McCarthy können falsch und antiintuitiv sein (das sog. *Yale-Shooting-Problem*: es gibt nicht übereinstimmende Modelle, die die Schlussfolgerungen repräsentieren, und wir wissen welchen wir wählen sollten.)<sup>13</sup> Das Problem entsteht deshalb, weil die nicht-monotone Logik nichtübereinstimmende Informationen als Prämisse zulässt. Wir können, dem Commonsense zufolge, manchmal auf Grund solchen Informationen Schlussfolgerungen ziehen.

Die beste Auswege aus dem Paradoxe zwei verschiedenen, aber genauso rechtfertigten Schlussfolgerungen von einer Menge von Prämissen, findet man in der Motivated-Action-Theory (MAT) von Morgenstern und insbesondere bei Lifschitz, der sog. verschachtelten Theorien von Abnormalität entwickelt hat. Diese Theorien lösen die technische Schwierigkeiten des Kulissenproblems. Lifschitz hat dann eine gute Methode (aber zu komplizierte, um sie hier darzustellen) erfunden, elegante Blocken des Kulissenaxiomen automatisch zu generieren.<sup>14</sup>

Die radikalen philosophischen Interpretationen des Kulissenproblems aber entstanden wurden, bevor Lifschitz und Morgenstern die technische Tricks erfanden.

### **3. Das kann man nicht den Hackers belassen...**

Jerry Fodor, Daniel Dennett und John Haugeland spielen die erste Geige im Streit um Wichtigkeit des Kulissenproblems für Philosophie. Natürlich, auch die Gegner von Künstlichen Intelligenz äußern sich dazu.<sup>15</sup> Hier aber spreche ich nur über den Naturalisten, da die Einwände von Dreyfus sind nicht mit diesem Problem so streng verbunden: Er spricht, wie auch immer, über die Nichtformalisierbarkeit des Know-hows, der ihm zufolge keine Satzrepräsentation haben kann.

John Haugeland kennt sich in der Problematik der Künstlichen Intelligenz sehr gut um. Er behauptet, dass KI eine neue epistemologische Schwierigkeit bereitet hat.<sup>16</sup> Obwohl er die Leistungsfähigkeit des Prinzips: „Alles bleibt unverändert, sonst etwas besonderes geschehen ist“ anerkannt,<sup>17</sup> stellt er auch fest, dass das Kulissenproblem aus der falschen Repräsentationssystem stammen kann: In der Satzrepräsentation gibt es ein Unterschied zwischen impliziten und expliziten Gehalte. In der graphischen Repräsentationen („Denkbilder“) sind aber alle Relationen direkt („komplizit“, wie Haugeland es nennt) dargestellt, also das Problem kann sich nicht melden.

Aus dem Standpunkt der logik-basierten Kognitionswissenschaft kann man erwidern, dass Haugeland etwa zu schnell die Satzrepräsentation als unpraktisch weggeworfen hat: Denn die graphischen Darstellungen des Wissens sind nicht so leicht

---

(1986), S. 89-116.

<sup>13</sup> Weitere Angaben finden sich bei Morgenstern, a.a.O.

<sup>14</sup> Technische Einzelheiten findet man in Arbeiten von Lifschitz.

<sup>15</sup> Vgl. z.B. H. L. Dreyfus, S. E. Dreyfus, „How to Stop Worrying about the Frame Problem Even though It's Computationally Insoluble,“ in: *The Robot's Dilemma: Frame Problem...*, a.a.O.

<sup>16</sup> John Haugeland, „Overview of the Frame Problem,“ in: *The Robot's Dilemma: Frame Problem...*, a.a.O.

<sup>17</sup> a.a.O., S. 84.

formalisierbar, also weißt man überhaupt nicht, ob sie keine Schwierigkeiten bereiten. Endlich hat Haugeland das Kulissenproblems auf das Holismusproblem des Schließens reduziert. Aber wo in der technischen Aufsätzen von Kognitionswissenschaftler findet man die Unterschiede zwischen implizite und explizite Informationen?<sup>18</sup> Und warum sollten wir annehmen, dass es im Schließen Holismus überhaupt gibt?

Dennett ist nicht davon gegangen.<sup>19</sup> Er erinnert uns, dass KI das menschliche Intellekt modellieren soll. Solche Formalisierungen wie Situationskalkül oder Circumscription finden allerdings keinen Platz in der Phänomenologie der Erkenntnisprozesse.<sup>20</sup> Wir haben also hier mit einem Missklang zu tun: Die komplizierten mathematischen Theorien lassen sich sehr schwer mit menschlichen Selbstverständnis vereinigen. Dennett vergleicht solche formalen ‚Erfindungen‘ in der Kognitionswissenschaften mit der Erfindung des Rads: Das Rad ist in der Natur nicht zu finden, aber ermöglicht schnellere Bewegung als Beine. Das Problem liegt aber darin, dass wenn es zu viel solche ‚kognitive Räder‘ gibt, dann die künstliche Intelligenz als Modell menschliches Geistes nicht mehr dienen kann.

Zum anderen aber, kann man mithilfe solcher Formalisierungen, Dennett zufolge, wesentliche philosophische Probleme entdecken. Ein dieser Probleme ist das Kulissenproblem, und zwar verstanden als Problem der Bestimmung relevanten Schlussfolgerungen in der Menge aller Schlussfolgerungen von vorgegebener Menge von Prämissen. Er strebt auch danach, das Kulissenproblem auf das Prädiktionsproblem, und dann auf das Induktionsproblem zu reduzieren: Es ist, seiner Meinung nach, kein besonderer Fortschritt, aber hilft, das Problem zu verstehen. Die Vertreter der logischen KI haben rasch Einwände gegen Dennett erhoben.<sup>21</sup> Das Kulissenproblem soll nicht so weit verstanden werden; es betrifft eher die Repräsentation der sich schell verändernden Welt. Dennett besteht aber nicht mehr auf seine weite Interpretation.

Das Problem des ‚kognitiven Rads‘ bleibt jedoch interessant. Die Motivation der logikbasierten Ansätzen war, das adäquate Modell des Commonsense mithilfe der Phänomenologie (im Sinne Dennetts; die kognitiven Forschungen bestanden denn oft darin, dass die Forschern die Subjekte gefragt haben, wie sie die bestimmten Problemen lösen<sup>22</sup>) zu entwickeln. Die heutigen konnektionistischen Forschungen sind auf dem neurophysiologischen Modell gestützt. Clark Glymour, der den Aufsatz Dennetts kommentiert hat,<sup>23</sup> erinnert uns, dass wir überhaupt nicht wissen, ob die menschlichen Gedankenweise die beste aller möglichen sind. Letzen Endes ist das Rad besser als Tausende von Beinen. Warum sollte dann KI sich mit dem menschlichen Geiste beschäftigen, wenn man überhaupt nicht weißt, wie er funktioniert? Klar, die Antwort

---

<sup>18</sup> Vgl. dazu Eric Lormand, „The Holorobophe’s Dilemma“ in: *Robot’s Dilemma Revisited*, K. Ford, Z. Pylyshyn (Hg.), Ablex, Norwood 1996.

<sup>19</sup> Der Aufsatz Dennetts wurde in 1984 geschrieben. Heute ist Dennett nicht völlig überzeugt, dass seine Meinungen zum diesen Thema richtig waren. Vgl. „Cognitive Wheels: The Frame Problem of AI“ in: D. Dennett, *Brainchildren*, MIT Press, 1998, S. 181-205.

<sup>20</sup> Dennetts Phänomenologie (sog. Heterophänomenologie) ist intersubjektiv und besteht darin, dass man die Berichte der Subjekten forsch, nicht die Gehalt des Bewusstseins des Forschers selber. Vgl. dazu Daniel C. Dennett, *Consciousness Explained*, Little, Brown, Boston 1991.

<sup>21</sup> Vgl. dazu Hayes, „What Frame problem is...“, a.a.O.; Lormand, a.a.O.

<sup>22</sup> Newell und Simon, a.a.O., S. 94.

<sup>23</sup> Clark Glymour, „Android Epistemology and the Frame Problem: Comments on Dennett’s «Cognitive Wheels»“, in: *The Robot’s Dilemma: Frame Problem...*, a.a.O., S. 65-75.

auf Glymours Frage lautet: Es geht eben darum, dass wir wissen wollen, wie er funktioniert. Die psychologischen Modelle ohne Formalisierung sind doch nicht so präzise wie die technische Beschreibungen im Rahmen der KI-Forschung. Man sollte also darauf rechnen, dass die KI-Wissenschaftler einen präzisen mathematischen Modell des menschlichen Geistes entwickeln. Dies aber schließt keineswegs die Möglichkeit aus, eine Intelligenz aufzubauen, die mehr leistungsfähig als die menschliche ist. Die letzte Interpretation ist heftig umstritten. Jerry Fodor ist für Ingenieure eigentlich ein *enfant terrible*. Er spricht über schlafenden Hunden, Musik der Sphären und einen Kühlschrank.<sup>24</sup> Hayes<sup>25</sup> zeigt, dass Fodor drei verschiedene Interpretationen in einem Aufsatz, der nur 10 Seiten hat, aufeinander darstellt. Erstens, versteht Fodor das Kulissenproblems als Problem Hamlets aus dem Ingenieurstandpunkt: Wann sollte man denken aufhören? Oder, mit Fodors Worten: Wie baut man eine Einrichtung auf, die rationell in dem Sinn, dass seine Mechanismen von Überzeugungsänderung in den separaten Modulen verkapselt sind?<sup>26</sup> Fodors Antwort: das weiß ich nicht, wenn ich wüsste, wäre ich reich und berühmt. Er behauptet, das Kulissenproblem betrifft das Wesen der Rationalität; wenn es in einem System gibt, dann das heißt, dass es rationell ist. Intelligente Prozesse, Fodors Meinung nach, sind nicht modular, und diese Tatsache führt zum Entstehen des Kulissenproblems.

Das Kulissenproblem ist so streng mit der Rationalität verbunden, dass es wie Musik der Sphären ist: Man kann sie nicht hören, weil sie überall ist. Er fügt auch hinzu: Die Strategie des ‚schlafenden Hundes‘ (d.h. „alles bleibt, wie es ist, sonst etwas besonderes passiert ist“) funktioniert auch gar nicht, weil bis heute die ‚intelligenten‘ Programme sind unglaublich dumm. Solche Strategie schiebt nur das Kulissenproblem aus der Theorien von Überzeugungsänderung zur Ontologie vor. Die ontologischen Problemen sind eben schwerer: man soll eine reiche kanonische Repräsentation der Tatsachen entwickeln, die meistens unverändert bleiben soll. Diese kanonische Repräsentation, oder Notation ist aber nicht leicht zu entwickeln.

Er zeigt das auf einem Beispiel. Der Philosoph schaltet den Kühlschrank ein und ab, und dies verursacht eine infinite Menge von Änderungen in seiner Meinungen: wenn er den Kühlschrank einschaltet, dann er glaubt, dass die Früchte länger frisch bleiben, sein Bier kühl wird, die Strömrechnung mehr betragen wird, der Kühlschrank laut nachts laufen wird usw. Wenn wir die relationale Eigenschaft der physischen Teilchen nach dem Einschalten des Kühlschranks Fodors als ‚Fridgeon‘ definieren, dann nach dem Einschalten wird der Zustand aller Teilchen in der Welt ändern, denn alle Teilchen diese Eigenschaft haben. Und man kann beliebig viele solche Prädikate wie ‚Fridgeon‘ definieren. Solche Prädikaten lassen sich von den ‚normalen‘ im Prinzip nicht unterscheiden. Wenn sie aber in der kanonischen Notation berücksichtigt werden, dann entsteht eine ungeheure Menge von Tatsachen, die mit jeder Handlung sich ändern. Mit anderen Worten, ohne gute Repräsentation kann man die Strategie des ‚schlafenden Hundes‘ nicht einsetzen. Als wir über solche Repräsentation noch nicht verfügen, denn die formale Ontologie des Commonsense ist noch nicht entstanden worden, das Kulissenproblem hat keine Lösung.

Fodor geht aber einen Schritt weiter. Er besteht darauf, dass das Induktionsproblem ist identisch mit dem Kulissenproblem: „Das Kulissenproblem [...] ist das Problem der nichtdeduktiven Inferenz (*nondemonstrative inference*); dann das Problem der nichtdeduktiven Inferenz [...] ist das Problem, wie funktioniert der erkennende Geist.

---

<sup>24</sup> Jerry Fodor, „Modules, Frames, Fridgeons, Sleeping Dogs, and the Music of the Spheres“, in: *The Robot's Dilemma: Frame Problem...*, a.a.O., S. 139-149.

<sup>25</sup> Vgl. Hayes, „What the Frame Problem...“, a.a.O.

<sup>26</sup> Fodor, a.a.O., S. 140.

Tut mir leid, dass McDermott fehlt die Geduld für den Philosophen, aber, offen gesagt, das Kulissenproblem ist zu wichtig, es den Hackers zu verlassen.“ Auf Einwände Hayes’ antwortet Fodor: man kann das Problem anderes nennen, aber es ist ganz egal, ob wir Krebs ‚Krebs’ nennen, oder nicht — wir werden sowieso an Krebs sterben. Die logik-basierenden KI-Forscher nur ab und zu verneinen, dass das Kulissenproblem mit der Repräsentation der Welt gebunden ist. Es ist aber fraglich, ob wir das generelle Problem der adäquaten Repräsentation überhaupt gleichzeitig lösen müssen. Dies ist vielleicht zu schwer. Die Informatiker sollen jedoch z.B. die Modelle der mechanistischen Metaphysik des 17. Jahrhunderts nicht mehr benutzen, um die Veränderungen und Dynamik besser zu repräsentieren.<sup>27</sup> Das heißt aber nicht, dass wir eine gebrauchsfertige Ontologie im Voraus bereiten sollen.

Das Induktionsproblem und das Problem, ‚wann wir denken aufhören sollen’, sind keineswegs identisch mit dem Kulissenproblem. Zum einen, wie könnte man a priori die Frage beantworten, wann wir denken aufhören sollen? Die einzige Antwort, die mir hier plausibel scheint, ist das sog. Prinzip der Gedankenökonomie. Meiner Meinung nach, es ist besser zu fragen, wann wir denken anfangen sollen. Diese Frage aber muss überhaupt nicht gestellt werden und sie hat nichts so viel mit der Modellierung der sich ändernden Welt zu tun. Das Induktionsproblem, also das Problem, wie man relevante induktive Schlussfolgerungen findet, hat ebenfalls wahrscheinlich keine apriorische Lösung. Ist das aber nicht eine übertriebene Forderung, die apriorischen Grenzen der Schlussfolgerung, die im Prinzip nicht apriorisch ist, zu ziehen? Wir finden nicht so viele deduktiven Momente in der Induktion, weil sie keine Deduktion ist.

Die philosophische Suche nach dem apriorischen Kriterium der relevanten Schlüsse, des Prüfsteins, der alle sich durch die Robohandlungen ändernde Weltzustände aufzeichnen erlaubte, könnte dann zum Scheitern verurteilt sein. Das Prinzip des ‚schlafenden Hundes’ in der nicht-monotonen Logik funktioniert bis jetzt am besten, obwohl es keine apriorische Wahrheit birgt. Diese Lösung ist provisorisch; wir haben doch nichts besseres.

Philosophen sollten vielleicht überlegen, ob etwas anderes als der Zusammenhang über die Relevanz des Schlusses entscheidet. Die Menge der in einer Situation relevanten Vorbedingungen der Handlung ist eine Menge von Merkmalen, aber die Relevanz selbst ist kein primäres Merkmal: Sie superveniert auf anderen Merkmalen. Die Relevanz ist also eine abgeleitete Eigenschaft anderer Eigenschaften in einer Situation. Die Frage, wie man relevante Vorbedingungen einer Handlung kennzeichnet, ist ähnlich an solchen Fragen: „Was ist das Kriterium des Vorhandenseins in der Raum-Zeit-Welt?“, „Was ist das Kriterium der Wahrheit?“, „Was ist das Kriterium des Gutes?“. Gut, es gibt keine solche allgemein gültigen Kriterien!

#### **4. Wie zieht man die Grenze?**

Die philosophischen Auffassungen des Kulissenproblems bei Haugeland, Dennet und Fodor scheinen nicht besonders fruchtbar. Sie reduzieren es auf bekannten Schwierigkeiten (z.B. Induktionsproblem) oder schlagen eine Revolution vor: ein anderes Darstellungssystem oder eine neue kanonische Repräsentation.

Es scheint mir, dass die Philosophen, die sich mit der formalen Philosophie beschäftigen, den KI-Forschern eigentlich helfen können. Die Kognitionswissenschaftler wissen nicht, wie man Metaphysik, Commonsense, Erkenntnistheorie, Semantik usw. formalisiert. Fodor hatte recht, es gibt Fragen, die Hackers nicht beantworten können. Solche Formalisierungen kann man mit

---

<sup>27</sup> Vgl. dazu Lars-Erik Janlert, „Modeling Change — The Frame Problem“, in: *The Robot’s Dilemma: Frame Problem...*, a.a.O., S. 32.

Phänomenologie im Sinne Dennetts verbinden: Die Selbstverständnis sollte mit den formalen KI-Systemen übereinstimmen.

Vielleicht man kann im Prinzip keine strenge Grenze zwischen Kognitionswissenschaften und Philosophie ziehen. Das könnte auch so schwer sein, wie das Erfinden der Kriterien des relevanten induktiven Schlussfolgerungen. Die Anhänger der naturalisierten Epistemologie Quines würden hier sagen, dass diese Unmöglichkeit daraus stammt, dass wir die theoretischen und empirischen Begriffen im Prinzip nicht trennen können.

Quines These ist doch in diesem Zusammenhang nicht hilfreich. Zum einen, gehören die Forschungen über die künstliche Intelligenz nicht nur zu theoretischen, aber auch zu empirischen und Ingenieurwissenschaften. Zum anderen, haben die Resultate der Kognitionswissenschaften eine ungeheure philosophische Bedeutung. Die Modelle der alltäglichen Schlussfolgerung, der künstliche Intellekt, notwendige und ausreichende Bedingungen der Erkenntnis sind typische philosophische Themen. Mehr noch: Wenn die KI-Informatiker einen Erfolg erzielen, dann werden sie auch damit beweisen, dass einige metaphysische Thesen wahr sind.

Die Relation zwischen den Naturalismus in der Philosophie und Kognitionswissenschaften ist ziemlich kompliziert. Die Grenze ist also fließend, aber sie lässt sich ziehen. Obwohl Naturalisten streben danach, ihre Philosophie und Naturwissenschaften in Einklang zu bringen, es ist doch klar, dass Philosophie keineswegs bloße Verallgemeinerung wissenschaftlichen Resultate ist. Kognitionswissenschaftler bauen die Modelle des erkennenden Geistes auf und entwickeln empirische Theorien oder funktionelle Prototype. Philosophen sind auf der Suche nach diesen wenigen Momenten, die für klar und invariant in solchen Forschungen genommen werden, und analysieren die Voraussetzungen der Forschung, z.B. das Prinzip, dass der Geist durch einen Rechner implementiert werden kann. Als es keine die weite theoretische Übereinstimmung in der Wissenschaften gibt, sie müssen auch eigene Hypothese herausarbeiten. Sie können in dieselben Forschungen teil nehmen und neue Felde der Forschung suggerieren. Die wichtigste Fragen entstehen dann, wenn sie vergleichen die alltägliche Selbstverständnis (in der intersubjektiven Phänomenologie gegeben) mit der logischen Analyse der Commonsenswissen. Bis heute haben die Naturalisten zu schnell neue Probleme auf etwas bekanntes reduziert. Sie können aber etwas mehr fruchtbares machen.