Cognitive Wheels: The Frame Problem of AI

by D. Dennett

eingereicht von:

Leopold Meuer 0616190

Thomas Rieger 0616432

Inhaltsverzeichnis

1. „Framing“

Das aus dem Bereich der amerikanischen Cartoon-Zeichner entlehnte Wort „Framing“ meint ursprünglicherweise den „Rahmen“, in dem sich eine Cartoon oder Comic Szene abspielt. Die Handlung ist hierbei im Vordergrund eines stets gleich bleibenden Bildhintergrundes eingebettet. Sprich die Kulisse oder der Kontext bleibt gleich, nur die Handlung ändert sich. Die Handlungen sind ihrerseits beschrieben durch das was sie ändern, während alles andere, das nicht von der Handlung betroffen ist (der „Rahmen“), unverändert bleibt. (http://en.wikipedia.org/wiki/Frame_problem)

2. Das Roboter Dilemma

Man stelle sich folgendes Szenario vor:

Wissenschaftler haben einen Roboter R1 entwickelt. Er soll seine Ersatzbatterie, die sich auf einem rollbaren Wagen in einem Raum befindet, bergen. Auf dem Wagen befindet sich ebenfalls eine Zeitbombe. Was macht unser Roboter R1 nun? Er betritt den Raum, fährt den Wagen auf dem seine Ersatzbatterie liegt hinaus und kümmert sich nicht weiter um den gefährlichen Appendix seiner Fracht. Die Bombe explodiert und hinterlässt einen glühenden Haufen an Platinen und Kabelsalat. Was ist passiert? Unser Roboter R1 hat nicht verstanden, dass, wenn er seine auf dem Wagen befindliche Ersatzbatterie hinaus fährt, auch die Bombe den Raum verlässt.

Die Wissenschaftler können dies nicht fassen und konstruieren einen neuen und besseren Roboter: R1D1. Dieser Roboter soll nicht nur die Auswirkungen seiner Handlungen, sondern auch die Auswirkungen der Nebeneffekte berücksichtigen können. Der Roboter betritt nun den Raum, fährt den Wagen hinaus und bemerkt, dass sich die Raumfarbe während des Hinausschiebens des Wagens nicht ändert. Draußen angekommen dauert es nur wenige Sekunden bis auch R1D1 sich pulverisiert hat.

Die Wissenschaftler sind außer sich vor Wut - sie beschließen, einen neuen, noch besseren Roboter zu bauen: R2D1. Er kann zwischen relevanten und nicht relevanten Auswirkungen unterscheiden. Das neue Wunder der Technik betritt nun den Raum. Bei der Ersatzbatterie angekommen, bleibt er plötzlich wie erstarrt vor dem Wagen stehen. Leise hört man ihn rechnen und rechnen. Die Wissenschaftler, die ob der Untätigkeit ihres

Zöglings in solch einer brenzligen Situation ganz aufgebracht und verängstigt zu dem Roboter hinein blicken, schreien plötzlich: „Tu doch was! Ach tu doch was!“ Der Roboter antwortet daraufhin: „Ich bin so damit beschäftigt, tausende und abertausende von irrelevanten Auswirkungen zu ignorieren, dass ich....“ Weiter kommt er nicht mehr. Eine gewaltige Explosion erschüttert den Raum. R2D1 hat sich binnen einer Sekunde in kleinste Atome „künstlicher Intelligenz“ verwandelt.

3. Die Problematik der Roboter

Was genau ist nun das Problem der Roboter und warum schaffen sie es nicht, kein Opfer der Zeitbombe zu werden und ihre Ersatzbatterie unversehrt zu bergen?

Erstens, sie haben kein Echtzeitplanungssystem. Ein solches würde es ihnen erlauben, sich in jeder Sekunde ihres Daseins an neue Umweltgegebenheiten oder veränderte Situationen anzupassen. Des weiteren fehlen ihnen Repräsentationen, die ausreichende Strategien für weiter gefasste Probleme bieten. Ein weiter gefasstes Problem wäre beispielsweise das nächtliche Schmieren eines Butterbrotes. Dieser, uns sehr banal erscheinende, von Dennett als „Midnightsnack Problem“ deklarierte Sachverhalt, beinhaltet viele zu berücksichtigende Variablen und Stolperfallen wie wir unter **Punkt 6** noch sehen werden. Der wichtigste zu berücksichtigende Punkt ist jedoch, dass sich Roboter nicht wie wir Menschen in Situationen befinden. Auch dazu mehr unter **Punkt 6**. Das Hauptproblem ist also, eine Form der Repräsentation zu finden, die eine komplexe, sich verändernde Welt adäquat abbildet.

4. Der Mensch, ein Intelligentes Wesen

Eines der Hauptmerkmale menschlichen Verhaltens ist, dass der Mensch *„...zuerst denkt bevor er lenkt“*. (Dennett, S. 2) Intelligenz meint auch, das zu nutzen, was man gut kennt, um die Voraussagekraft seiner Handlungen und deren Folgen zu verbessern. Wie macht der Mensch das?

Der Mensch besitzt eine Vielzahl an angeborenen Dispositionen. Neugeborene haben beispielsweise schon die Fähigkeit Schmerz zu empfinden oder unangenehme Gerüche als solche wahrzunehmen. Dies belegen zahlreiche Studien aus der

Entwicklungspsychologie. Auch hat der Mensch von Geburt an das implizite Wissen des *Modus Ponens*:

A bedingt B → Wenn A, dann auch B.

Jmd. umbringen bedingt Gefängnis → Ich bringe jmd. um, also Gefängnis.

Ebenso besitzt der Mensch das implizite Wissen des *Ausgeschlossenen Dritten:*

Dr. Leidlmair ist unser Dozent oder er ist es nicht. Es gibt keinen Status dazwischen.

Der Mensch ist also keine *Tabula Rasa* wie z.B. von Locke angenommen wurde.

5. Der phänomenologische Ansatz

Locke ging bei seinen Annahmen von einem phänomenologischen Ansatz aus. Phänomenologisch meint eine Herangehensweise, die sich ausschließlich mit dem unmittelbar Wahrnehmbaren und als solchem auch mit denjenigen Dingen, die der Introspektion zugänglich sind, beschäftigt. So kann man jedoch nicht den menschlichen Geist verstehen, weil beim Denken etc. sehr wohl Prozesse ablaufen, die man nicht sehen kann.

6. Fragestellung des Heterophänomenologischen Ansatzes

Die Konsequenz der Annahme, dass es nicht genügend sei, die der Beobachtung und Introspektion zugänglichen Informationen zu nutzen, liegt vor allem in der theoretischen Fundierung heterophänomenologischer Annäherungen. Im Falle des „Midnightsnack-Problems“, müsste man sich also fragen, welche Informationen zu dessen Lösung theoretisch notwendig sind. Augenscheinlich ist die Lösung des Problems wenig kompliziert und auf wenige explizite Informationen gestützt. Die tatsächliche Menge an zu beachtenden Dingen, ist jedoch viel größer. So zum Beispiel müssen viele grundlegende Gegebenheiten, wie die Schwerkraft, die Tatsache, dass etwas niemals zur gleichen Zeit an mehr als einem Ort sein kann und viele weitere Details, die die Umgebung betreffen (Haus, Küche, Kühlschrank, Messer, Materialkonsistenzen, etc.), berücksichtigt werden. Tatsächlich handelt es sich dabei aber nicht um eine unendlich lange Liste von Informationen, die nach den Regeln der klassischen Logik, nach und nach durchgearbeitet

wird und in expliziten Handlungsanweisungen mündet, wie es möglicherweise in einer extremen Version der *Language of Thought Theorie* angenommen wird. Die meisten Informationen sind für das seiende Wesen implizit in der Situation enthalten und fließen ohne bewusste Verarbeitung in die Auswahl der richtigen Handlungen ein. Intelligente Wesen sind also intelligent, weil sie, als lebende Wesen, in die Situation eingebettet sowie von dieser betroffen sind und entsprechend implizites Erfahrungswissen über die unterschiedlichsten Situationen angesammelt haben. Die wesentliche Beteiligung der Erfahrung, lässt sich leicht an den Reaktionen eines Handelnden ablesen, wenn etwas nicht funktioniert, also von seiner Erwartung über die Entwicklung der Situation abweicht. Die große Rolle der Trefflichkeit der Erwartungen über Handlungskonsequenzen in bestimmten Situationen, hat außerdem zu einer Diskussion geführt, in der das Frame-Problem auf das Induktionsproblem reduziert werden sollte, welches besagt, dass der Induktionsschluss von einzelnen spezifischen Erfahrungen auf die Allgemeinheit des selben Sachverhalts in der Zukunft, nicht zulässig ist. (http://de.wikipedia.org/wiki/Induktionsproblem)

Es würde also in unserem Kontext bedeuten, dass das Hauptproblem der Künstlichen Intelligenz darin läge, die verschiedenen Einzelwahrscheinlichkeiten der Situation nicht zu kennen. Dennett entgegnet an dieser Stelle, dass auch ein Roboter, der alle denkbaren Wahrscheinlichkeiten des täglichen Lebens gespeichert hätte, nicht handlungsfähig sei, solange er die Wahrscheinlichkeit aller möglichen Ereignisse zunächst explizit prüfen müsste, um eine Aussage über deren situative Relevanz treffen zu können.

7. Umgehungsversuche der Vertreter der Künstlichen Intelligenz

Seit der Formulierung des Frame-Problems als mögliche Ursache der Probleme der Künstlichen Intelligenz, haben Arbeitsgruppen immer wieder versucht selbiges durch neuartige Designs zu umgehen. Einige dieser Versuche sollen im Anschluss kurz vorgestellt werden.

7.1 Frame-Axiome

Hierunter ist grundsätzlich der Versuch zu verstehen, das gesamte relevante Wissen über

Handlungen in bestimmten Umgebungen in Form von Axiomen zu repräsentieren, die sich formell wie folgt darstellen lassen:

Situation 1 + Aktion 1 = Situation 1'

Diese Axiome sollten wiederum von den sogenannten Frame Axiomen umgeben sein, welche die Verwendung der richtigen Handlungsaxiome in der richtigen Situation ermöglichen sollen, indem sie Informationen über allgemeine Situationstypen und Handlungstypen, sowie deren generelle Effekte und Nicht-Effekte enthalten. Im Folgenden soll ein Beispiel für notwendige Frame Axiome der Handlungen „Anmalen“ und „Raustragen“ erläutert werden.

Die Situation sei:

X ist blau, X ist im Haus.
Handlungsbefehl 1: Rot anmalen
Handlungsbefehl 2: Raus tragen

Welcher Schluss ist dem Computer dadurch über den Zustand von X möglich? Zunächst würde man als intelligentes Wesen davon ausgehen, der Computer wisse jetzt, dass X rot ist und sich draußen befindet. Tatsächlich weiß der Computer allerdings nur, dass sich X draußen befindet, weil er nicht davon ausgehen kann, dass die Farbe von X durch die Aktion „raustragen“ nicht beeinflusst wird. Ebenso verhält es sich anders herum, so fern der Befehlszeitpunkt von Handlungsbefehl 1 und 2 vertauscht wird. (http://plato.stanford.edu/entries/frame-problem/)

Notwendige Frame-Axiome wären also:

1) X ist rot und „Raustragen“ = X ist rot
2 X ist draußen und „ rot anmalen“ = X ist draußen

3) X ist rot und „Y blau anmalen“ = X ist rot
4) X ist draußen „Y blau anmalen“ = X ist draußen
. . .

. . .

. . .

Das Problem hierbei ist also eine nicht bewältigbare Menge allein an „Nicht-Effekt-Axiomen“, die letzten Endes zu keiner verbesserten Handlungsfähigkeit führt, wenn man von extrem aspektarmen künstlichen Umgebungen absieht. Selbst ein System, das alle

Regeln einer bestimmten Mikrowelt kennt, wie es von Terry Winograd umgesetzt wurde, hat keinerlei Möglichkeiten erfolgreich in einer anderen Mikrowelt zu agieren.

7.2 Versuch der Relevanztestung

Die Ausgangsüberlegung dieser Annäherung ist, dass sich nur eine minimale Anzahl der möglichen Ereignisse in einer Situation tatsächlich ereignet und dass sich die zu bearbeitende Datenmenge in einer Situation drastisch reduzieren würde, wenn ein Programm in der Lage wäre, die wahrscheinliche Teilmenge der Ereignisse abzustecken. Das Hauptproblem dieses Ansatzes liegt in der formal-logischen Organisation der Schlussprozesse, zu denen ein Computer fähig ist. Grundeigenschaft formal-logischer Systeme ist es nämlich, sich von Schluss zu Schluss durchzuarbeiten und so nach und nach ein richtiges Bild der Informationen zu erstellen. Es ist grundsätzlich nicht möglich bestimmte Teile der Kette je nach Relevanz zu entfernen. Die einzige bleibende Möglichkeit, wäre also ein System, welches zusätzlich zu den repräsentierten Wahrscheinlichkeiten aller möglichen Ereignisse auch noch deren situative Relevanz prüfen könnte. Statt der erwünschten Datenreduktion, explodiert die Datenmenge erneut.

7.3 Die gerichtete Aufmerksamkeit: Der Schlüssel von Minsky und Schank

Der an die Entwicklungspsychologie Piagets erinnernde Ausgangspunkt dieses Lösungsvorschlags, besteht in der Annahme, dass die Anzahl der gesamt gemachten Erfahrungen, auf eine bewältigbare Anzahl von „Situationsstereotypen“ (Frames/Skripts) reduzierbar ist. Entsprechende Stereotypen steuern je nach Situationstyp die Aufmerksamkeit des Handelnden auf bestimmte Aspekte der Situation, bevorzugen bestimmte Schlüsse und stellen situative Abweichungen von den Skripts fest. Ein so konstruiertes System bräuchte dem entsprechend, einige untereinander gut vernetzte Skripts, die gewisse Variationen der Situationstypen beinhalten und eine Möglichkeit der Wahrnehmung. Die Simulation solcher Systeme ist relativ erfolgreich verlaufen, so lange die Situationen nicht zu stark variieren, was zu einem totalen Blackout des Systems führt, so dass es im Anschluss nur noch stumpf die nicht funktionierenden Stereotypen abspult, bis es beendet und neu designed wird. Der entscheidende Mangel liegt also in der völlig fehlenden Fähigkeit dazu zu lernen und alte Frames, je nach Erfahrung, neu zu gestalten oder zu verwerfen. Ein selbst erdachtes Beispiel wäre ein Roboter, der auf einer

Geburtstagsparty weiß, dass mehrere Flaschen Wein auf einem Tisch, wahrscheinlich Geschenke und somit nicht zu trinken sind. Sobald er sich jedoch auf einer Mottoparty befände, würde er fälschlicherweise das Faschingsparty-Skript laden und beginnen wild den Wein zu trinken, wohingegen ein lernfähiges System schlicht den Geburtstagsfeier-Stereotyp um den Begriff Mottoparty ergänzen könnte. Ein Beispiel aus der tatsächlichen Entwicklung von Systemen, die sich solcher Stereotypen bedienen, stammt von einem Geschichtenerzählerprogramm, welches von Roger Schank (s.o.) entwickelt wurde, dass mit Hilfe solcher Stereotypen zwischen den Zeilen lesen können soll. Bekommt es also in einer Geschichte, die den Restaurant-Frame betrifft die Informationen, dass John in ein Restaurant geht, dort einen Hamburger isst und anschließend geht, wobei ihm der Kellner in den Mantel hilft, so kann es die Frage beantworten, ob John bezahlt hat. Es kann von der Höflichkeit des Kellners darauf schließen. Es besteht jedoch weiterhin das Problem der mangelnden Variationsbreite. Angenommen John wäre der Chef des Lokals, so wäre das Programm überfordert.

8. Konklusion

Der zentrale Punkt, um den sich das Frame-Problem dreht, ist die Tatsache, dass sich ein Computer zu keinem Zeitpunkt in einer Situation befindet, die an sich etwas für ihn bedeutet, beziehungsweise ihn betrifft, weil der Computer kein Seiendes ist. Aufgrund dieser Tatsache muss man einem Computer theoretisch alles, was für in Situationen eingebettete Wesen implizit im Kontext enthalten und als kontextueller Stereotyp angelegt ist, explizit einprogrammieren, was aufgrund der unüberschaubaren auftretenden Datenmengen praktisch noch nicht möglich ist. Die Wichtigkeit impliziten Wissens, wird auch bei einer von der NASA durchgeführten Eyetracker Untersuchung deutlich, im Rahmen derer Piloten explizit angeben sollten, wo sie bei der Landung ihrer Maschinen im Cockpit hinsehen. Die Angaben entpupptem sich im Vergleich mit während der Landung erhobenen Eyetracker Daten als falsch. Die Piloten hatten außerdem große Schwierigkeiten ihre Handlungsstrategien bei manuellen Landungen zu erinnern und sie in explizite Regeln zusammen zu fassen. Es zeigte sich sogar, dass diejenigen Piloten, die versucht hatten ihr Wissen in explizite Regeln umzuformen, im Nachhinein schlechter flogen.

9. Diskussion

Daniel C. Dennett, der Autor des von uns bearbeiteten Artikels „Cognitive Wheels: The Frame-Problem of AI“, greift gegen Ende des Artikels eine höhere Ebene der Problematik auf, indem er die im Titel des Artikels erscheinenden Cognitive Wheels näher thematisiert. Es ist eine metaphorische Auseinandersetzung, die die Entwicklung von kognitiven Theorien jenseits der Phänomenologie, mit der Erfindung des Rades vergleicht. Es geht dabei darum, dass das Rad eine unschätzbar große Rolle in der Modernen Fortbewegung einnimmt und somit das Gesicht der Welt entscheidend prägt. Trotzdem bleibt das Rad eine künstliche Erfindung und gibt letztlich wenig Aufschluss über die in der Natur vorkommenden Fortbewegungsstrategien. Springender Punkt dieser Überlegung ist es, dass es ein Trugschluss wäre, davon auszugehen, man könne den menschlichen Geist verstehen, wenn man einen Computer erschaffen hat, der sich so verhält wie ein intelligentes Wesen. Tatsache ist, dass selbst bei einem Roboter, der sich in einer komplexen, sich ständig verändernden Welt zurecht fände und dabei auf phänomenologischer Ebene so agieren würde, wie ein Mensch, nicht zwingend die gleichen basalen kognitiven Vorgänge zu Grunde liegen müssen. Er wäre also eine bahnbrechende Neuerung im Bereich der künstlichen Intelligenz und könnte das Gesicht der Welt auf lange Sicht prägen, würde aber ebenso wie das Rad auf die natürliche Fortbewegung, keine Rückschlüsse auf die natürliche Kognition ermöglichen.

10. Literatur

Printmedien:

Dennett, Cognitive Wheels: The Frame Problem of AI, 2010

Internet:

http://en.wikipedia.org/wiki/Frame_problem
http://plato.stanford.edu/entries/frame-problem/
http://www.iscid.org/encyclopedia/Frame_Problem
http://de.wikipedia.org/wiki/Induktionsproblem