



Leopold-Franzens-Universität Innsbruck  
Institut für Psychologie

Ao. Univ.-Prof. Dr. Karl Leidlmair

## **Symbol Grounding Problem**

Semesterarbeit

im Rahmen des Seminars

Forschungsseminar: Neuere psychologische Fachliteratur

VO-Nr.: 720 584

SS 2006

Studienkennzahl: C 298

Vorgelegt von

Kiebacher Georg, 0217145

Petter Markus, 0317384

Sprenger Michael, 0315578

Innsbruck, Mai 2006

## Inhaltverzeichnis

1	Die Herausforderungen der realen Welt und die Grenzen des Computermodells .....	3
2	Symbol Grounding Problem.....	4
2.1	Einführung.....	4
2.2	Menschliche Symbole .....	4
2.3	Symbol Grounding Problem.....	4
2.4	Zitate zum Symbol Grounding Problem .....	5
2.4.1	Searle .....	5
2.4.2	Harnad .....	6
2.4.3	Dreyfus .....	6
3	Lösungsansätze zum „symbol grounding problem“.....	6
3.1	Stevan Harnad und Searles chinesisches Zimmer.....	6
3.2	Systematische Symbolmanipulation vs. Intrinsische Bedeutung.....	7
3.3	Exkurs: Autonome Agenten .....	8
3.4	Harnads „totaler Turingtest“ .....	9
3.5	Steve Austin versus the Symbol Grounding Problem.....	10
3.6	Schlusspunkte und Kritik .....	13

# 1 Die Herausforderungen der realen Welt und die Grenzen des Computermodells

Das Computermodell des Geistes steht für die erste, die »klassische « Phase der Kognitionswissenschaft. Seine Stärken liegen in der Modellierung gut formalisierbarer kognitiver Leistungen wie Rechnen oder Schach spielen. Die Grenzen des Modells wurden und werden jedoch umso deutlicher, je öfter man Computerprogramme beziehungsweise Roboter, die von ihnen gesteuert werden, aus ihren simulierten Welten entlässt und sie mit den Unwägbarkeiten des täglichen Lebens konfrontiert.

Computer erledigen schnell und zuverlässig eben diejenigen intellektuellen Leistungen, die Menschen am schwersten fallen: abstraktes Rasonieren und präzises Rechnen. Eben weil diese abstrakten Leistungen den Menschen schwer fallen, gilt ihre Beherrschung als Zeichen hoher Intelligenz.

Doch bei Computern verhält es sich gerade nicht sowie bei Menschen: Ein Computer, der die abstraktesten Formen des Rasonierens beherrscht, kann keineswegs auch die scheinbar einfacheren Dinge, wie Schuhe zubinden, den Schirm mitnehmen, wenn es nach Regen aussieht.

Anders als Schachprogramme oder Expertensysteme sind Menschen nicht auf eine Tätigkeit festgelegt. Sie können nicht nur Schach spielen oder nur Theoreme beweisen oder nur Krankheiten diagnostizieren. Sie können auch Pizza backen und Socken stricken, eine Party geben, sich in einer fremden Stadt zurechtfinden und vieles mehr. Dies bezeichnet man als allgemeine Intelligenz.

Auch wenn Psychologen heute verstärkt davon ausgehen, dass sich die menschliche Intelligenz aus Modulen zusammensetzt, ist es allem Anschein nach nicht damit getan, unterschiedlich spezialisierte Systeme einfach zusammenzuschalten. Es führt kein Weg vom Schachcomputer Deep Blue zu den typischen Leistungen allgemeiner menschlicher Intelligenz.

Zu den praktischen Problemen gehören vor allem die geringe Fehlertoleranz und die mangelnde Robustheit der künstlichen Systeme. Während Menschen sich anhand von Informationsfragmenten orientieren können und geringe Probleme mit dem Verstehen schlampiger Sprechweisen, dem Lesen teilweise überklebter Buchstaben oder dem Erkennen halbverdeckter Gegenstände haben, goutiert ein Computerprogramm solche Dinge überhaupt nicht.

.Ein weiteres Problem schließlich ist die so genannte Echtzeit-Performance. Ein intelligenter Organismus muss den Kopf in dem Moment einziehen, wenn der Stein geflogen kommt, nicht erst zehn Minuten später, wenn er mit seiner Datenverarbeitung fertig ist. Es spricht alles dafür, dass die zentralisierte Informationsverarbeitung in klassischen seriellen Computern dafür zu langsam und unflexibel ist.

Eines der größten Probleme für Systeme der klassischen Künstlichen Intelligenz und hier in unserem Referat auch das zu behandelnde Problem ist das Symbol Grounding Problem. Dabei handelt es sich um die Frage, wie abstrakte Symbole zu ihrer Bedeutung in der realen Welt kommen.

## **2 *Symbol Grounding Problem***

### ***2.1 Einführung***

Die grundlegende Frage bei der Herstellung von KI ist, wie die Symbole, die ein Computer verstehen soll, ihre überhaupt Bedeutung bekommen. Es geht also um das Prinzip der Symbolverankerung. Für uns Menschen ist der Umgang mit Begriffen normal, unsere ganze Sprache baut auf diesen wenigen Symbolen des Alphabetes auf. Allerdings sind unsere Begriffe und Symbole für uns nicht einfach eine logische Konstruktion aus bestimmten Bausteinen, sondern unsere Begriffe sind über eine vielfältige Erfahrung in der Wirklichkeit, unserer Umwelt, verankert. Dies hat zur Folge, dass eine Symbolverankerung nach einem Konzept einer Identität verlangt.

### ***2.2 Menschliche Symbole***

Menschliche Symbole haben immer einen Bezug zu der für uns Menschen zugänglichen Umwelt. Sie sind wie oben genannt in unserer Umwelt verankert. Als Beispiel: Wird von einem Ethnologen eine Sprache erlernt, für die es kein Wörterbuch gibt, so kann er den Umweg über die Umwelt gehen. Zeigt eine Person auf ein Ding, so kann sich der Ethnologe zwar nicht 100%ig sicher sein was die Person meint, aber er hat eine Vermutung darüber.

### ***2.3 Symbol Grounding Problem***

In den verschiedenen Computerprogrammen sind die verwendeten Symbole immer syntaktisch definiert. Das heißt, dass sie über Kausalrelationen, in denen sie zu anderen

Symbolen stehen, verankert sind. Es handelt sich bei der syntaktischen Ebene nur um die Regeln des Satzbaus und der Wortgruppen. Die Symbole sind somit also nur definiert über das was passiert wenn sie aktiviert sind, und über die Art wie sie zu verarbeiten sind. Die semantische Komponente, das heißt die Bedeutung eines Satzes ergibt sich aus den Einzelbedeutungen der Symbole, wird außer acht gelassen. Die Symbole selbst sind nicht definiert. Somit beziehen sich die syntaktischen Prozesse die in einem Computerprogramm ablaufen auf gar nichts. Sie sind nur deshalb Symbole, weil wir Menschen sie als diese interpretieren. Die Bedeutung der Symbole wird also wieder vom Menschen zugeschrieben und nicht von Computerprogramm. Ein bezeichnendes Beispiel für die Arbeitsweise eines Computers ist Searles Konzept des chinesischen Zimmers. Der Computer hat die Anweisungen wie er mit den Symbolen umzugehen hat, aber die Bedeutung der Symbole an sich bleibt ihm unbekannt.

Dies bedeutet also, dass die Definition der Symbole nicht intrinsisch, also aus dem System heraus erkennbar ist, sondern nur parasitär, das heißt durch die Interpretation des Benutzers. Der Computer selbst ist nur ein geschlossenes System, das keine Außenwelt kennt und diese nicht erfahren kann. Wird also so einem Computer oder Programm vom Menschen Intelligenz zugesprochen, so liegt dies nur daran, dass wir Menschen uns in diesen Dingen sehr leicht täuschen lassen.

Interessant ist, dass frühere intelligente Systeme, wie etwa alte Schachcomputer nicht dem Symbol Grounding Problem unterlagen. Diese Programme brauchten nicht zu verstehen was sie tun, wie in diesem Fall dass sie Schach spielen. Hier reicht es völlig aus, wenn der Benutzer weiß, dass das Programm dazu in der Lage ist. Anders ist es hingegen bei neueren Robotern und Computern. Für sie ist es wichtig sich in der Welt zurecht zu finden und zu wissen und zu verstehen, welche Symbole für welche Objekte und Gegenstände stehen.

## ***2.4 Zitate zum Symbol Grounding Problem***

### **2.4.1 Searle**

Searle meinte zu diesem Thema: „Intelligenz ist ein mentales Phänomen. Dieses ist nicht unabhängig von der Hardware (unser Gehirn) – sie braucht unseren Körper. Somit kann ein Computer nicht intelligent sein.“

Wichtig ist hier aber festzuhalten, dass Searle nicht der Meinung ist, dass es nicht irgendwann eine Maschine gibt, die die Sprache versteht. Dass Computer dies bis jetzt noch

nicht können liegt nicht daran, dass sie anstatt wie das menschliche Gehirn aus Neuronen, sondern aus Siliziumchips besteht. Er kritisiert nur jene Maschinen, deren Verhalten auf rein formalen Prozessen basiert. Searle ist der Meinung, dass Computer die die Sprache verstehen können über ein Bewusstsein und über eine Intentionalität verfügen müssten. Solche Systeme müssen also eine uns ähnliche Wahrnehmung besitzen und ihre eigenen Erfahrungen in der Umwelt machen.

#### **2.4.2 Harnad**

Auch Harnad ist der Meinung, dass eine Maschine wie beispielsweise ein Roboter intelligent sein kann und Sprache verstehen kann. Harnad versteht unter Intelligenz ein beobachtbares Verhalten. Damit ein Programm oder eine Maschine aber intelligentes Verhalten zeigen kann, muss dieses über einen Körper in seine Umgebung eingebettet sein. Auch er ist somit der Meinung, dass ein Programm seine eigenen Erfahrungen machen muss, damit man von Intelligenz sprechen kann.

#### **2.4.3 Dreyfus**

Dreyfus veröffentlichte zu diesem Thema in den 70er Jahren des letzten Jahrhunderts, dass dieses Problem nicht gelöst werden kann, da es nicht möglich ist das Alltagswissen eines Menschen explizit darzustellen, um es einem Computer verständlich zu machen. Dieses Wissen wird nämlich nicht nach bestimmten Regeln und in bestimmten Sätzen gespeichert, sondern auch in nicht-expliziter Form. Unter Wissen in nicht-expliziter Form versteht er zum Beispiel körperlich eingeübte Fähigkeiten und Fertigkeiten, die keine regelhaften Abstraktionen sind. Er vertritt somit die Meinung, dass nur Wesen mit einem Körper dazu in der Lage sind, nicht explizite Wissensformen zu verstehen.

### **3 Lösungsansätze zum „symbol grounding problem“**

#### ***3.1 Stevan Harnad und Searles chinesisches Zimmer***

Auf der Suche nach einer Lösung zum so genannten symbol grounding problem stößt man unweigerlich auf den Namen Stevan Harnad, dessen Arbeiten den obigen Begriff und die damit verbundenen Problematik entscheidend geprägt haben:



Stevan Harnad, (geb. 02.06.45 in Budapest) ist ungarischer Kognitionswissenschaftler mit Professuren in Cognitive Science an der University of Southampton sowie der Université du Québec in Montreal. Seine Forschungsschwerpunkte liegen im Bereich Kategorisierung, Kommunikation und Kognition. Harnad erklärt nun das so genannte symbol grounding Problem anhand von 2 Versionen von Searles Modell des chinesischen Zimmers:

Beim ursprünglichem Modell muss die Versuchsperson, welche sozusagen in die Rolle des Computers in dem Zimmer sitzt, eigentlich gar nicht intelligent/selbstständig handeln, sondern lediglich die in der eigenen Muttersprache abgefassten formalen Regeln zur Neuordnung der chinesischen Symbole befolgen. Dies entspricht in etwa der Arbeitsweise eines normalen „Programms“. Harnad geht aber einen Schritt weiter und zeigt auf, dass der Computer unfähig ist, die Bedeutung der Symbole zu erkennen. Also darf auch das Regelwerk der Versuchsperson für diese nicht verständlich sein.

In der 1. Version soll die Versuchsperson also Chinesisch als zweite Sprache allein mit einem chinesisch-chinesisch Wörterbuch lernen. Dies ist an sich schon ein äußerst schwieriges, doch immerhin nicht vollkommen unlösbares Unterfangen, da die Person ja über eine vorherige Kenntnis über grundlegende Regeln wie z.B. Satzbau, Grammatik usw. verfügt und deshalb die chinesischen Symbolfolgen zumindest als „Worte“ und „Sätze“ (intrinsische Bedeutung) erkennen kann.

Die 2. Version stellt nun die Konsequenz dieses Dilemmas dar. Hierbei soll Chinesisch allein mit Hilfe des chinesisch-chinesisch Wörterbuchs als erste Sprache gelernt werden, d.h. ohne vorherige Kenntnis einer Sprache und deren formaler Regeln, ja sogar ohne die Kenntnis, was „Sprache“ eigentlich bedeutet! Diese Version ist damit unmöglich lösbar, da die Versuchsperson einfach nicht in der Lage ist, die Symbole als Buchstaben, Worte und Sätze als solche zu deuten und einfach planlos mit ihnen spielt. Somit lässt sich allein aus der symbolischen Syntax keine direkte semantische Bedeutung gewinnen.

### ***3.2 Systematische Symbolmanipulation vs. Intrinsische Bedeutung***

Anhand genau dieser unlösbaren Situation stellt sich für Harnad das symbol grounding problem anschaulich dar: In der Unfähigkeit des Computers, die Bedeutung der eingegebenen Symbolfolgen inhaltlich zu verstehen und sie selbstständig anhand eines intrinsischen Wissens zu bearbeiten. Wie kann aber der Übergang von der systematischen Manipulation von Symbolen durch ein Computerprogramm zur intrinsischen Bedeutung gelingen? Harnads Meinung dazu: Damit ein Programm versteht, was die Symbole bedeuten, muss es in der Lage sein, sie mit der Umwelt selbstständig in Beziehung zu setzen. Dies funktioniert aber nur,

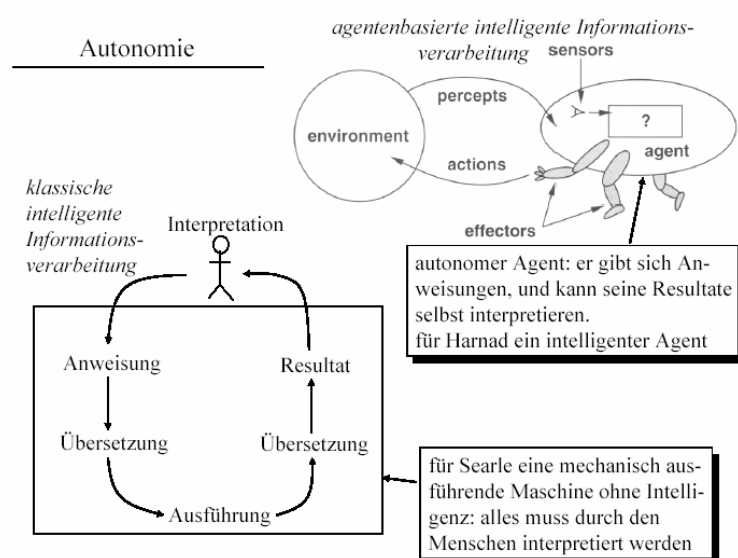
wenn das Programm über einen Körper und dessen Sensoren in seine Umgebung eingebettet („situated“) ist.

Anders ausgedrückt: Das symbol grounding problem handelt von der Schwierigkeit, verschiedene Symbole mit ihren entsprechenden Bedeutungen in der Umwelt zu verknüpfen, ohne auf externe Interpretationsprozesse zurückgreifen zu müssen. Hierbei stellt sich die Frage, wie die Symbole des Programms mit Objekten der Umgebung, die sie darstellen, verbunden werden können. Eine mögliche Idee zur Lösung bietet sich im Bereich der „embodied“ KI mit der Entwicklung der „autonomen Agenten“.

### 3.3 Exkurs: Autonome Agenten

Nach Turing zeigt sich Intelligenz im beobachtbaren Verhalten. Harnad geht ebenfalls davon aus, dass ein Programm, welches über einen Körper in seine Umgebung eingebettet ist, lernen kann, die Bedeutung von Objekten der Umwelt zu begreifen, Situationen interpretieren und entsprechend reagieren. Prinzipiell kann somit ein Roboter Intelligenz zeigen.

Autonome Agenten sind nun Roboter, welche in einem künstlichen Körper (Gehäuse) stecken, mittels unterschiedlicher Sensoren Informationen aus der Umwelt aufnehmen können und dank eingebauter Instrumente darauf reagieren können. So verfügen diese Agenten meist über irgendeine Art von Fortbewegungsmittel wie z.B.



Räder, Raupen oder auch Beine, um sich durch ihre Umwelt bewegen zu können. Teilweise haben sie auch Greifarme oder anderes Gerät, um mit anderen Objekten noch zusätzlich „interagieren“ zu können.

Das System der autonomen Agenten ist insofern autonom, als dass es nicht durch einen menschlichen Überwacher gesteuert werden muss. Stattdessen setzt es sich aus verschiedenen Programmen zusammen, die auf unterschiedlicher Ebene einzelne Funktionsweisen des Roboters kontrollieren und dabei ununterbrochen Informationen austauschen. So registriert beispielsweise das Kameraauge des Roboters verschiedene Umweltobjekte, welche aufgrund ihrer erfassten Eigenschaften (wie z.B. Größe, Form, in



Bewegung oder nicht) verschiedene Unterprogramme oder auch Verhaltensweisen aktivieren, um in adäquater Weise auf die jeweilige Situation reagieren zu können. So werden beispielsweise große Hindernisse als solche erkannt und umfahren, kleinere Objekte aus dem Weg geräumt oder auch gesammelt, und gegenüber beweglichen Objekten wie Menschen wird der Rückzug angetreten, um mögliche Zusammenstöße zu vermeiden.

Beispiele für solche autonome (d.h. zumindest ansatzweise selbständig handelnde) Agenten finden sich in den unterschiedlichsten Versionen und Ausführungen in den verschiedensten Einsatzbereichen. Größtenteils



finden sie im Bereich der technischen Weiterentwicklung und praktischen Erprobung Verwendung, werden aber zunehmend auch in Bereichen des täglichen Lebens, der Wirtschaft oder der Forschung eingesetzt. Die Bandbreite ihrer Nutzung reicht von „relativ“ banalen, spielerischen Aufgabenstellungen wie „Roboter-Fußball“ bis hin zum Einsatz in speziellen Bereichen der Raumfahrt (z.B. Pathfinder-Vehicle Marsmission 96).

### 3.4 Harnads „totaler Turingtest“

Um eine eingehende Überprüfung des Intelligenzgrades eines Programms zu gewährleisten, schlägt Harnad im Folgenden eine Erweiterung des Turing-Testverfahrens und die Einteilung in 4 Versionen vor:

Der „Partial Turing Test“ (T1) erfordert lediglich, dass ein System in der Lage ist, innerhalb eines einzelnen Bereichs des symbolischen Levels wie z.B. Gesichtserkennung oder Schachspiel exakten und sinnvollen Input und Output zu liefern. Viele der heutigen Expertensysteme erfüllen die T1-Anforderungen für eine einzelne kognitive Aufgabe, beispielsweise der Schachcomputer „Deep Blue“. Allerdings genügen die Kriterien des T1 bei weitem nicht, um einem System Intelligenz zuzuschreiben.

Der „Standard Turing Test“ (T2) ist das originale Verfahren, welches Turing selbst entwickelt hat. Hierbei soll das System in der Lage sein, in jedem kognitiven Bereich sinnvolle Input- und Outputinformationen zu liefern. Aber auch der Standard Turing Test ist nicht ausreichend, um Intelligenz nachzuweisen. Dies wurde schon weiter oben durch Harnads Argumentation bezüglich der Unlösbarkeit des chinesischen Zimmer-Modells klar

gestellt. Harnad meint dazu: „Ein virtueller Flieger fliegt nicht wirklich, eine virtuelle Kochplatte heizt nicht wirklich, (...) ein virtueller Brieffreund denkt und versteht auch nicht wirklich, sondern er ist nur ein Symbolsystem, welches lediglich aufgrund unserer Interpretation zu denken und zu verstehen vorgibt.

Beim „Robotic Turing Test“ (T3) soll neben den Input-Output-Infos aller Kognitionsbereiche zusätzlich das vollständige Spektrum der menschlichen sensomotorischen Leistung gegeben sein. Dazu muss das betreffende System mit einem Körper ausgestattet sein, über den es mit der Umwelt in Austausch treten kann. Harnad zufolge stellt der T3 die Mindestanforderungen dar, unter denen ein Mensch alltägliche Situationen erfassen und darauf reagieren kann. Doch um die Bedingungen des T3 wirklich zu erfüllen, müsste vorerst das besagte symbol grounding problem gelöst werden.

Der „Total Turing Test“ (T4) schließlich erfordert, dass ein System alle Voraussetzungen für T3 erfüllt und zusätzlich in seinem neuromolekularen Aufbau und Funktionsweise keine Unterschiede zu einem menschlichen Gehirn mehr aufweist.

Wie aus der obigen Einteilung ersichtlich wird, zeigt sich, dass bereits die Bewältigung des T3 für heutige Systeme ein geradezu gigantisches Unterfangen sein würde und dass sich folglich der totale Turingtest (T4) mit keinerlei System mehr verwirklichen lässt. Damit kommt Harnad zu dem Schluss, dass sich das gesamte Konzept des „symbol grounding problems“ insofern bewahrheitet hat, als dass es für diese Problematik keine greifbare Lösung zu geben scheint.

Zum Abschluss soll hier noch kurz auf einen weiteren Artikel von John L. Taylor und Scott A. Burgess von der Humboldt State University eingegangen werden, um einen weiteren zusätzlichen Blickwinkel auf die vorliegende Problematik zu erhalten.

### ***3.5 Steve Austin versus the Symbol Grounding Problem***

Nachdem zuvor einige Ansätze zur Lösung des symbol grounding problems anhand der Sichtweise Stevan Harnads und des Konzepts der „embodied“ KI vorgestellt wurden, folgt nun ein Artikel von zwei Vertretern des „semantischen“ Ansatzes der Kognitionsforschung, welche in gewisser Hinsicht den Gegenpol zur embodied KI darstellt.

In ihrem Artikel setzen sich die Autoren Taylor und Burgess kritisch mit Harnads Argumenten über Searles chinesisches Zimmer und den totalen Turing Test auseinander. Harnad geht dabei ja davon aus, dass es zwischen der einfachen systematischen Manipulation von Symbolen eines Computerprogramms und der Verknüpfung dieser Symbole mit einer

intrinsischen Bedeutung. Die zentrale Fragestellung lautet hierbei: Wie funktioniert die Übersetzung der Bedeutung von wahrgenommenen Umweltobjekten in ein Symbolsystem? Und welche Rolle spielen diese Übersetzungsprozesse dabei genau?

Taylor und Burgess bewegen sich auf ihrer Suche nach diesen Übersetzungsprozessen nun in die gegenteilige Richtung: Ausgehend vom natürlichen Wahrnehmungsprozess des Menschen werden dessen einzelne sensomotorische Funktionen des Körpers schrittweise reduziert, bis hin zu einem (fiktiven) vollständigem Abkoppeln des Gehirns vom Körper. Auf diese Weise wollen die Autoren herausfinden, inwieweit der Prozess der Verknüpfung von Symbolen mit Umweltobjekten unabhängig von der Einbettung des Systems in einen Körper ist. Anhand der Figur von Steve Austin, dem Protagonisten der amerikanischen TV-Serie „The six million dollar man“ aus den Jahren 1974-78 erstellen Taylor und Bruggess 4 Szenarios, in welchen sie die Figur von Steve Austin schrittweise depriviert wird, der Körper entfernt und schließlich das Gehirn durch ein synthetisches Präparat ersetzt wird:



Szenario 1: „Normal Steve/Bionic Steve“ – Zu Beginn ist Steve ein normaler Mensch mit natürlichen Wahrnehmungsfähigkeiten, d.h. mit der natürlichen menschlichen Fähigkeit, seine Umwelt anhand intrinsischer Prozesse entsprechend zu bewerten und auch mit anderen Objekten in seiner Umwelt zu interagieren. Dazu verfügt Steve über organische Übersetzer wie z.B. die Retina, das Gehör oder die Hautrezeptoren. Nach seinem folgenschweren Absturz werden in einer 6 Millionen Dollar Operation Teile seines Körpers (Beine, rechter Arm, linkes Auge) durch cybernetische Teile ersetzt. Damit verfügt Steve neben seinen natürlichen kognitiven Funktionen auch über verschiedene elektronische Übersetzer wie z.B. Mikrophone, visuelle Linsen, Positionssensoren usw., welche ihm dabei helfen, seine Umwelt wahrzunehmen und ihm den Austausch mit Umweltobjekten ermöglichen.

Taylor und Burgess gehen nun davon aus, dass sich die Bewertungs- und Handlungsfähigkeit von Steves Gehirn durch einen schrittweisen Austausch von natürlichen Übersetzern durch künstliche Übersetzer nicht wesentlich ändern dürfte. Sie glauben, dass Steves Gehirn zwischen den übermittelten Informationen seiner natürlichen Organe und den Informationen der künstlichen Übersetzer keinen Unterschied machen würde, wenn nur die Qualität der technischen Ersatzteile hoch genug wäre. Die Autoren glauben sogar, dass sich

ohne weiteres auch verschiedene Teile des Gehirns austauschen lassen könnten, was in dem Szenario 4 noch näher ausgeführt werden wird.

Szenario 2: „Shut in/Sensory deprivation Steve“ – Das zweite Szenario sieht vor, dass Steve aufgrund seines Unfalls eine akute Agoraphobie entwickelt. Folglich zieht er sich ins Haus zurück und schränkt den Kontakt zu seinem Umfeld völlig ein, doch er ist immer noch ohne weiteres in der Lage, seine Gedanken mit einer intrinsischen Bedeutung zu versehen. Selbst unter Bedingungen völliger sensorischer Deprivation (wie z.B. in einer Isolations-Kammer) dürfte seine Denkfähigkeit nicht wesentlich eingeschränkt werden. Der Grund dafür ist, dass Steve nicht nur mittels direkter Übersetzung aus der Umwelt inhaltlich sinnvolle Gedanken generieren kann, sondern auch dank rein symbolischen Übersetzern bedeutungsvolle Gedanken hat. Also gehen Taylor und Burgess noch einen Schritt weiter und eliminieren jegliche Art der direkten Übersetzung, indem sie im folgenden Szenario Steves Gehirn von seinem Körper abkoppeln.

Szenario 3: „Disembodied Steve“ – Aufgrund der Annahme, dass Steve fähig ist, inhaltlich sinnvolle Gedanken auch ohne direkten Kontakt zur Umwelt zu produzieren, soll sein Gehirn in Folge vollständig von seinem Körper getrennt und in einem speziellen externen Speicher mit Nährflüssigkeit („brain in a vat“) transferiert werden. Wenn diese Operation theoretisch möglich wäre und das Gehirn daraufhin mittels direkten Anschluss an nun rein künstliche Übersetzer sämtliche benötigte Informationen über seine Umwelt erhalten würde, glauben die Autoren beweisen zu können, dass intrinsische Bedeutung auch ohne die Zuhilfenahme eines Körpers entstehen kann!

Damit widersprechen Taylor und Burgess der Meinung von Stevan Harnad, dass ein System nicht ohne einen Körper Intelligenz entwickeln kann. Im Gegenteil: Sie sind der Meinung, dass Steves Gehirn selbst im 4. Szenario vollständig durch eine molekulare Kopie ersetzt werden könnte, welche allein durch künstliche Übersetzer symbolischer Art inhaltlich sinnvolle Gedanken produziert.

Szenario 4: „Synthetic Steve“ – Durch die obige Annahme, dass Steves Gehirn durch eine molekulare Kopie, den so genannten „brain in a vat“, völlig ersetzt werden könnte, wird hier nun suggeriert, dass es für Harnads totalen Turing Test (T4) eine Lösungsmöglichkeit geben würde! Es stellt sich hierbei die Frage: Könnte man die Vision von einem rein synthetischem

Gehirn und damit einem selbstständig denkenden System tatsächlich Wirklichkeit werden lassen?

### **3.6 *Schlusspunkte und Kritik***

Die Autoren John L. Taylor und Scott A. Burgess kommen trotz ihrer Kritik an Harnads Konzepten selbst nicht umhin einzuräumen, dass die genaue Rolle der Übersetzungsprozesse in Bezug auf die intrinsische Bedeutung noch immer unklar bleibt. Sie beschränken sich vielmehr auf die Frage nach dem Unterschied zwischen dem Gehalt von intrinsischer und extrinsischer Bedeutung.

Anhand der im Seminar vorgestellten Beispiele der Filme „Matrix“ und „Matrix Reloaded“ wird ersichtlich, dass selbst ein hoch entwickeltes Computersystem mit nahezu perfekter Abbildung einer künstlichen Umwelt sich letzten Endes als unzureichend herausstellt, da es lediglich eine in sich selbst geschlossene Welt in sich birgt.

Die Problematik bei der eben vorgestellten Geschichte: Selbst wenn es in Zukunft möglich sein sollte, ein synthetisches Äquivalent unseres Gehirns zu schaffen, welches selbstständig inhaltlich sinnvolle Gedankengänge produzieren kann, ist die Art der Informationen, welche das synthetische Gehirn erhält, immer ein künstlich gefiltertes Abbild der Umwelt. Um es auf den Punkt zu bringen ist dies gar keine wirkliche Umwelt mehr, sondern lediglich eine weitere künstliche Realität, in der das System gefangen ist!

Gleichzeitig wurde anfangs auch aufgezeigt, dass auch der Ansatz der „embodied“ KI vor bis jetzt noch immer unüberwindlichen Problemen steht, wenn es um die Fähigkeit zu wirklich selbstständigem Denken und Handeln geht. Es bleibt noch abzuwarten, inwiefern sich die momentane Situation betreffend dieser Problematik noch weiter entwickeln wird, doch ist sicher, dass sich für das Konzept des „symbol grounding problem“ wohl so bald keine endgültige Lösung finden lassen wird.