

Dipl. Ing. Dr. rer. pol. habil. Gero Zimmermann (München)  
**Wollen autonome Fußballroboter Tore schießen?**  
**Ein Beitrag zur Diskussion um den freien Willen**

---

Die Diskussion um den freien Willen ist in den letzten Jahren im wesentlichen ein Disput zwischen Hirnforschern und ihren Erkenntnissen einerseits und Philosophen andererseits gewesen. Eine ganz andere Sicht, nämlich die der Informatik und Robotik, blieb dabei außen vor. Man hat sich meist mit dem Hinweis auf die Grenzen der künstlichen Intelligenz und der neuronalen Netze abgefunden, denen übereinstimmend die Fähigkeit des Wissens, Denkens, freien Entscheidens und Bewusstseins abgesprochen wird. Stellvertretend stellt Wolfgang Lenzen die Frage: „Ist ein Wesen, das mit Hilfe gewisser Wahrnehmungsorgane ... (Sinnes-)eindrücke zu einer Stelle in seinem Inneren leitet, wo sie alsdann identifiziert, kategorisiert und interpretiert werden, ist ein solches Wesen wirklich möglich, wenn man zugleich darauf besteht, dass in seinem ‚Geist‘ bzw. in seinem Inneren keinerlei Spur von Bewusstsein, Empfinden und Erleben stattfindet?“ (Lenzen, S.270).

Die zunehmenden Fähigkeiten künstlicher Systeme, insbesondere in der Robotik, zeigen, dass solche Fragen noch nicht abschließend geklärt sind und deshalb nicht zu den Akten gelegt werden dürfen. Deutlich wird dies bei den Fähigkeiten autonomer Fußballroboter, die ihren aktuellen Entwicklungsstand erst jüngst bei der RoboCup Weltmeisterschaft vom 14. bis 20. Juni 2006 in Bremen eindrucksvoll demonstrieren konnten. Diese Fußballroboter sind autonome Systeme, die im Team versuchen müssen, möglichst viele

Tore zu schießen und umgekehrt Tore des Gegners zu verhindern.

In einem Faltblatt der Deutschen Forschungsgemeinschaft DFG (ohne Jahresangabe, vermutlich von 2006) mit dem Titel: »Roboter mit Fingerspitzengefühl«, werden mögliche Einsatzgebiete für intelligente Systeme wie folgt beschrieben: „Mobile Roboter erweitern zunehmend die Einsatzbereiche der klassischen Robotik. Sie sollen dabei nicht nur beweglich sein, sondern auch eigenständig handeln können. Die Forschung auf dem Gebiet der mobilen Robotik versucht daher, lernende mobile Systeme zu entwickeln, die ihre Umgebung wahrnehmen, Wissen aufbauen und sinnvolle Verhaltensentscheidungen treffen und umsetzen können. Fragen zur Sensorik oder Bildverarbeitung spielen dabei eine ebenso wichtige Rolle wie das Entstehen von Verhaltensmustern oder Lernprozesse.“

Die aktuellen Entwicklungen beschäftigen sich mit vielen Problemstellungen, mit denen auch Lebewesen permanent konfrontiert werden. Die Roboter müssen ihre Umgebung (ihr Habitat?) wahrnehmen, sich selbst lokalisieren (wo bin ich?) und eine Route berechnen (wohin will bzw. soll ich?). Und natürlich müssen sie schnelle Entscheidungen treffen, (nicht nur) wenn sie im Ballbesitz sind (was will bzw. was soll ich jetzt tun?). Die Roboterentwickler sprechen hier vom Weltmodell des Roboters, aus dem sich u.a. sein Verhalten ableitet. Roth schreibt: „Kognition bezieht

sich auf komplexe, für den Organismus bedeutungsvolle, d.h. für Leben und Überleben ... relevante und deshalb meist erfahrungsabhängige Wahrnehmungs- und Erkenntnisleistungen. Diese arbeiten in der Regel mit Repräsentationen im Sinne einer »Stellvertretung« sowie mit rein internen »Modellen« der Welt und der Handlungsplanung, *gleichgültig ob diese bewusst oder unbewusst sind*“ (Roth 1997, S.31). Und deshalb sind unsere Maschinenfußballer auch kognitive Systeme.

Wichtig ist hierbei, dass auf solche Systeme kontinuierlich eine Vielzahl von Informationen einströmt und sie auf die permanenten Änderungen in ihrer Umgebung jeweils sofort angemessen reagieren müssen, sie müssen ihre Umwelt also in Echtzeit wahrnehmen (Wo ist der Ball? Wer hat den Ball? Wo ist das Tor? Wo sind die eigenen Mitspieler und wo die des Gegners usw.?). Da Fußball ein Mannschaftsspiel ist, dürfen die Spieler keine „einsamen Entscheidungen“ treffen, sondern müssen kooperieren, das heißt Spielzüge „planen“ und gemeinsam durchführen. Sie müssen dafür durch spontane Vernetzung Teams bilden und anstehende Aufgaben gemeinsam erledigen, indem sie Aufgaben auf die Spieler verteilen (wer macht was?). Sie müssen deshalb „soziale Wesen“ sein. Ein Fußballroboter vor dem gegnerischen Tor muss beispielsweise entscheiden, ob er selbst aufs Tor schießen soll, oder ob er den Ball einem Mitspieler zuspiesen soll, der besser positioniert ist. Und natürlich muss er seine Entscheidung mit dem Mitspieler „absprechen“.

Wir wollen diese Anforderungen an unsere Fußballroboter zunächst an den Kriterien spiegeln, die in Philosophenkreisen angeblich den „freien Willen“ charakteri-

sieren und danach fragen, wie diese Fähigkeiten technisch realisiert sind.

Das gängigste Argument, das für den freien Willen ins Feld geführt wird, ist, dass unsere Entscheidungen durch Gründe bestimmt werden. „Gründe sind Antworten in Reaktion auf zweifelnde Fragen, was man für wahr halten oder was man tun soll. Die Beantwortung solcher Fragen ist die Funktion von Gründen. Zur Erfüllung dieser Funktion müssen sie Wünsche, Absichten und Überzeugungen in Verbindung setzen mit deren Wahrheits- und Erfüllungsbedingungen in der Welt. Ohne die Eigenschaft, einen Bezug auf solche Gültigkeitsbedingungen herzustellen, würden Gründe ihre Aufgabe nicht erfüllen. Neuronale Zustände haben diesen Gültigkeitsbezug aber nicht. Sie können ihresgleichen nicht zu etwas von ihnen Verschiedenem in semantische Beziehungen bringen“ (Wingert, S. 200/201).

Stellen wir uns nun vor, unser Fußballroboter steht vor dem gegnerischen Tor und muss die Entscheidung treffen, ob er in die linke oder in die rechte Ecke schießen soll. Wenn der Torwart in der linken Ecke steht, wird er die rechte Ecke wählen und umgekehrt. Ist nun die Stellung des Torwarts die Ursache oder der Grund für die Entscheidung des Roboters? Wir meinen, beide Termini werden dem wahren Kern der Problematik nicht gerecht, man sollte besser von Bedingungen reden. Der Einfachheit halber unterstellen wir hier aber, dass sich die Roboterentscheidung mit der Stellung des Torwarts begründen lässt, denn laut Wingert leitet sich aus der Stellung des Torwarts die Antwort auf die Frage ab, in welche Ecke der Roboter schießen soll und „die Beantwortung solcher Fragen ist die Funktion von Gründen.“ Hier wird die „Absicht“, ein Tor zu

schießen, mit den dazu erforderlichen Erfüllungsbedingungen (dem Weg zum Tor darf kein Torwart im Wege stehen) in Verbindung gebracht. Reicht es aber schon aus, wenn sich das Verhalten des Roboters aus Gründen bzw. Bedingungen (die z.B. in einem Programmcode fixiert sind) ableiten lässt, oder muss der Roboter selbst diese Gründe kennen (er selbst, oder der Computer, der ihn steuert, oder wer eventuell sonst noch?) und eventuell sogar explizit benennen können? Zwar hat unser Roboter keine „neuronalen Zustände“, dafür können wir aber im Arbeitsspeicher seines „Computergehirns“ bestimmte Bitmuster in Form von elektrischen Zuständen ausmachen. Und diese Zustände repräsentieren genau den Gültigkeitsbezug zwischen der Stellung des Torwarts und der Entscheidung für die Ecke, in die der Schuss erfolgen soll. Dabei bringen das Rechen- und Steuerwerk des Computers die elektrischen Erregungsmuster im Chip in eine semantische Beziehung zu etwas von ihnen Verschiedenem, nämlich zur Situation vor dem Tor. Man sagt, die Erregungsmuster im Chip *repräsentieren* die Situation vor dem gegnerischen Tor. Und aus dieser Repräsentation, sie entspricht einer „mental-ur-sache“, leitet sich die Entscheidung des Fußballroboters ab.

Nach Ansgar Beckermann „... drängt sich die Frage auf, wie es denn möglich sein soll, dass die Gehirnzustände, auf die wir (angeblich) Wünsche und Überzeugungen zurückführen können, in der Regel genau die Handlungen verursachen, die im Hinblick auf diese Wünsche und Überzeugungen rational sind“ (Beckermann, S.16). Die Antwort lautet: Die Gehirnzustände sind hier Bitmuster im Arbeitsspeicher des Computergehirns des Roboters, die die

aktuelle Situation vor dem Tor repräsentieren und das Computerprogramm leitet aus ihnen eine rationale Handlung ab, nämlich den Torschuss in die richtige Ecke. Natürlich muss unser Roboter auch wissen, wo er sich befindet, also dass er vor dem gegnerischen Tor steht (Selbstlokalisierung), und ob er besser selbst (Ich?) schießen soll, oder sein Mitspieler (Du?). „Mit diesem »Ich«-Sagen ... wird ... ausgedrückt, dass der Sprecher eine Reihe von ziemlich anspruchsvollen Dingen tut: Er setzt sein Urteil und seine Bewertung zu Wahrheitsbedingungen bzw. zu Normen und Wertmaßstäben in Beziehung. Und er unterscheidet zwischen dem Erfülltsein dieser Bedingungen und Anforderungen einerseits und seinem Urteilen, Entscheiden und Tun andererseits; er unterscheidet zwischen sich und der sozialen und natürlichen Welt“ (Wingert, S. 198). Sicherlich kann unser Roboter zwischen sich und seiner natürlichen Welt (Spielfeld, Tore, Ball) und seiner sozialen Umwelt (die anderen Spieler) unterscheiden. Auch kann er natürlich unterscheiden zwischen den Bedingungen für seine Entscheidungen und deren Erfülltsein einerseits (ist die Bedingung, dass der Torwart in der rechten Ecke steht, erfüllt?) und seinem Urteil, also der daraus abgeleiteten Konklusion andererseits (also schieße ich, falls die Bedingung erfüllt ist, in die linke Ecke). Sollte der Torwart ziemlich in der Mitte stehen, muss unser Roboter die alternativen Wahrheitsbedingungen (der Torwart steht links bzw. der Torwart steht rechts) bewerten und entscheiden, welche Ecke bei einem Torschuss größere Erfolgsaussichten verheißt. Hat unser Roboterfußballer aber deshalb auch schon Intentionen? Die bekannte Position von John Searle lautet ja: syntaktische

Repräsentationen auf Basis formaler syntaktischer Regeln reichen nicht aus, um Intentionalität zu erzeugen. Kann der Roboter also aufgrund rein syntaktischer Programmanweisungen auf das Tor schießen, ganz ohne begleitende Intentionen, denn Syntax reicht für Intentionen ja angeblich nicht aus? Kann mit anderen Worten ein Torschuss abgeleitet werden, der nicht (z.B. auf das Tor) gerichtet ist? Wir brauchen nur den Programmcode zu studieren, in dem festgelegt ist, wie Entscheidungen getroffen werden, dann können wir diese Frage mit nein beantworten.

Eine weitere Forderung von Wingert lautet: „Wenn man, wie Neurobiologen es oft tun, sagt: »Das Gehirn urteilt und entscheidet«, dann müsste im Fall der Herstellung eines Wahrheitsbezugs gezeigt werden können, wie neuronale Prozesse einen fehlbaren Anspruch erheben, das heißt wie sie irren können“ (Wingert, S. 198). Ersetzen wir wieder die neuronalen Prozesse durch Abläufe im Computer und fragen, ob sich ein solches System auch irren kann. Natürlich kann es sich irren, indem es beispielsweise eine bestimmte Situation falsch einschätzt. Genau genommen irrt sich dabei nicht der Computer, sondern das Gesamtsystem aus Wahrnehmungssensoren, daraus abgeleiteten internen Repräsentationen und daraus wiederum abgeleiteten Entscheidungen. Hier kann die Wahrnehmung unscharf sein, oder die daraus abgeleitete Repräsentation auf einer Missinterpretation beruhen oder aus der Repräsentation eine nicht adäquate Schlussfolgerung gezogen werden. So identifizieren unsere Fußballroboter beispielsweise den Ball anhand seiner Farbe. Befindet sich nun versehentlich und unerlaubterweise ein anderer Gegenstand mit derselben Farbe ebenfalls im Gesichtsfeld

des Roboters, kann dieser diesen Gegenstand fälschlicherweise für den Ball halten und damit einem Irrtum unterliegen. Und wir erweitern unsere Frage gleich noch dahingehend, ob unsere Fußballroboter auch etwas „glauben“ können? Beispielsweise versucht jeder Maschinenfußballer den Ort des Balles aus seinem Blickwinkel heraus zu lokalisieren. Dabei kann es durchaus voneinander abweichende Ergebnisse bei der Standortbestimmung des Balles geben (Roboterhunde in der Four Legged Liga sehen beispielsweise nur sehr unscharf, Kameraauflösung 208 x 160 Pixel). Daher stimmen die Fußballer die Ergebnisse ihrer Ortsbestimmungen untereinander ab und ordnen dem Ball dann gemeinsam die Position zu, die dessen Aufenthaltswahrscheinlichkeit maximiert. Diese Lokalisation ist aber mit einer bestimmten Ungenauigkeit bzw. Unsicherheit behaftet. Deshalb „wissen“ die Roboter nicht genau, wo sich der Ball befindet, sondern „glauben“, dass er dort ist, wo ihn die gemeinsame Ortsbestimmung lokalisiert. Idealerweise sollte ihr Verhalten dann aber auch berücksichtigen, dass es sich hier um ein „unscharfes Wissen“ handelt. Derartige Problemstellungen lassen sich mit „Fuzzy Logik“ behandeln. Die Informatiker arbeiten hier mit sog. Gewissheitsfaktoren (certainty factors) die fragen, wie sich die Gewissheit im dann-Teil ändert, wenn sich im wenn-Teil Wahrscheinlichkeiten ändern (vgl. z.B. Spies).

Fragen wir noch einmal: kann ein Fußballroboter etwas glauben? Zumindest gibt es für seinen Glauben einen propositionalen Gehalt, der durch einen dass-Satz ausgedrückt wird (*dass* der Ball am angenommenen Punkt liegt) und eine Erfülltheitsbedingung (*dass* er nicht woanders liegt).

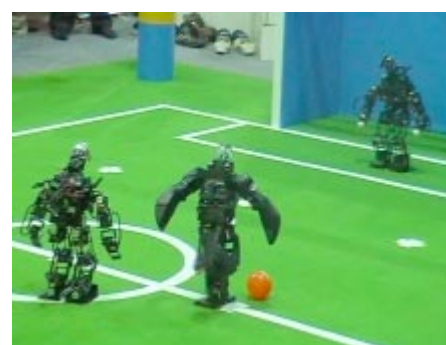
Es geht um die Frage, ob dies hinreichende oder nur notwendige Bedingungen für ein Glauben und ganz allgemein für einen mentalen Zustand sind. Reicht es, wenn sein Entwickler (der den Programmcode und die Art der internen Repräsentation der Balllokalisierung kennt) weiß, dass der Roboter etwas glaubt, oder muss er selbst es wissen? Dies entspricht der Frage, ob auch all die unbewussten Annahmen und Prämissen, also das was wir gemeinhin für richtig halten und glauben und was unserem alltäglichen Tun und Denken unbewusst zugrunde liegt, als intentional bezeichnen wollen oder nur bewusste Intentionen. Was konkret bedeutet in unserem Fall die Forderung, dass der Roboter selbst wissen muss, dass er etwas glaubt, oder anders gesagt, nicht genau weiß? Es bedeutet, dass er die Quelle seines Wissens und deren Zuverlässigkeit kennen muss. Und dieses Wissen muss ihm explizit zugänglich sein, damit er es bei seiner Entscheidungsfindung mit berücksichtigen kann. Dieser Anforderung werden die Fußballroboter (heute noch) nicht gerecht. Wenn die Four Legged Fußballer nicht wissen, wo der Ball ist, wenn sie ihn also noch nicht lokalisiert haben, drehen sie den Kopf in alle Richtungen, bis sie ihn erspäht haben. Haben sie also eine Vorstel-

lung von einem nicht sichtbaren Objekt, also ein *Meta-Wissen*, dass es einen solchen Ball geben muss, den es nur noch zu finden gilt? Ein solches Meta-Wissen haben sie nicht, sondern nur Programm-anweisungen die festlegen, was zu tun ist, wenn der Ball noch nicht lokalisiert ist (*if* Ball bereits lokalisiert, *then* Aktion xyz, *else* Kopf drehen, *until* Ball lokalisiert). Man darf also aus der Beobachtung ihres Verhaltens keine falschen Schlüsse ableiten. Aber könnte es nicht sein, dass dies analog auch für unser Bewusstsein gilt, das unser eigenes Verhalten (genauer: die Repräsentation des Verhaltens unseres Körpers und unserer Körperzustände im Gehirn) beobachtet?

An dieser Stelle wollen wir jetzt kurz auf die Frage eingehen, wie denn das Verhalten unserer Maschinenfußballer konkret gesteuert wird. Grundprinzip der Entscheidungsfindung bei derartigen Systemen ist die Aufteilung der Verarbeitungs-module auf verschiedene Kompetenzschichten (näheres hierzu bei Pfeifer/Scheier). Wir sprechen hier auch von Abstraktionsebenen (levels of abstraction). Am Beispiel unserer Fußballroboter wollen wir zunächst zeigen, was wir uns unter verschiedenen Abstraktionseben konkret vorzustellen haben. Nehmen wir ein



Four legged Liga (AIBO-Hunde)



Liga der humanoiden Roboter

(Fotos: Zimmermann)

praktisches Beispiel. In der Liga der autonom agierenden vierbeinigen Fußballroboter (getunte AIBO Roboterhunde) lautet eine Anweisung für einen Roboter beispielsweise: „wenn nah (Ball) und geradeaus (Tor) dann führe einen Headkick aus“ (vgl. z.B. [www.lh2006.com](http://www.lh2006.com)). Der nahe Ball und der Standort vor dem Tor sind dann die Gründe (besser: die Bedingungen bzw. mentalen Ursachen) für einen Schuss mit dem Kopf aufs Tor. Andere Verhaltensanweisungen können lauten: „gehe zum Ball“, „gehe ins eigene Tor“ oder „schieße den Ball zum Mitspieler xy“. Für alle diese Einzelaktionen gibt es jeweils Gründe (Bedingungen), die diese Aktionen auslösen. Welche Gründe in einem konkreten Fall gerade relevant sind, leitet sich aus der Bewertung der Gesamtsituation auf der nächst höheren Ebene ab (wo ist der Ball, wo sind die eigenen Spieler und wo die des Gegners?). Unterhalb der „Navigationsebene“ wird dann beispielsweise eine „Bewegungsebene“ angesiedelt, auf der die einzelnen Bewegungselemente im Detail geplant werden (vorwärts/rückwärts gehen, schnell/langsam gehen, geradeaus gehen oder nach links/rechts drehen usw.). Auch hier gibt es wieder Gründe für die jeweils angemessene Bewegung (z.B. genaue Position relativ zum Ball, Kollisionsvermeidung u.ä.). Für die unterste Ebene wird hieraus dann (etwa in der Klasse der humanoiden Fußballroboter) abgeleitet, welche Gelenkcontroller wie anzusteuern sind, um die beabsichtigten Bewegungen auszuführen. Hier bewegen wir uns auf der Ebene der Physik, also der Ebene von (physikalischen) Ursachen und Wirkungen (vgl. z.B. [www.roboerectus.org](http://www.roboerectus.org)). Wenn also Ansgar Beckermann schreibt: „Wir haben gesehen, dass der Versuch, mentale Ausdrücke in

physikalischer Sprache zu definieren, scheitert und scheitern muss“ (Beckermann, S. 98), dann wird hier nicht berücksichtigt, dass natürlich die Sprache der verschiedenen levels of abstraction eine unterschiedliche ist. Das Computerprogramm beschreibt das Roboterverhalten unabhängig davon, wie es physikalisch implementiert wird. Deshalb kann physikalisches Verhalten nicht retrograd das Programm beschreiben, ebenso wenig wie uns die Natur die Gesetze der Physik und das Leben auf unserem Planeten die Gesetze der Evolution erklärt. Diese müssen wir schon selbst erarbeiten und was wir dabei erhalten, sind allenfalls Hypothesen. Haben wir sie aber erst gefunden, können wir auch künstliche Systeme schaffen, die genau diesen Gesetzen folgen. Deshalb schreibt John Searle auch ganz richtig: „Es gibt nicht zwei Mengen von Ursachen: das Bewusstsein und die Neuronen. Es gibt nur eine Menge, die auf verschiedenen Ebenen beschrieben wird“ (Searle, S. 47). Hier landen wir noch bei einem weiteren Problem, das in Zusammenhang mit den vielfach zitierten und kommentierten Experimenten von Libet, Haggard und Eimer zum „freien Willen“ diskutiert wird. Auf welcher Ebene wird hier im „Roboterhirn“ eigentlich entschieden? Es gibt, vereinfacht gesagt, Entscheidungen zum *was* man tun möchte, zum *wie* man es tun möchte und zum *wann* man es tun möchte. Man löst die Problematik des freien Willens nicht dadurch, dass man einfach die Entscheidungen darüber, *wann* eine Aktion konkret gestartet wird zu „Nicht-Entscheidungen“ umdeklariert. Beispielsweise bezeichnet Mynarek in den o.g. Experimenten „das bewusste Fingerdrücken, den letzten Willensruck ...“ nur noch als Exekutivakt (Mynarek, S.121).

Das hieße für unseren Roboter, die Auslösung eines „Schusses“ ist nur noch der Vollzug einer vorausgegangenen Planung. Wie sieht also diese Planung aus, was würde Libet im Microchip des „Roboterhirns“ an Aktionspotentialen, oder sagen wir besser an Aktionen, messen? Die Antwort lautet: das hängt davon ab, auf welche Kompetenzschicht sich seine Messungen beziehen. Auf der obersten Ebene würde er Aktivitäten erkennen, wenn Grundsatzentscheidungen zu treffen sind (Direktschuss aufs Tor oder Ballweitergabe an einen Mitspieler). Auf der darunter liegenden Ebene gibt es Aktionspotentiale, wenn die Annäherung zum Ball und die Stellung relativ zum Ball geplant wird. Am *Ende*(!) jeder dieser Planungen steht eine Entscheidung, dann ist das Aktionspotential aber bereits wieder abgeklungen! Wir dürfen also nicht den Entscheidungsvorgang mit der getroffenen Entscheidung verwechseln! Wenn unser Bewusstsein erst die getroffene Entscheidung registriert (was sonst?), brauchen wir uns also nicht zu wundern, wenn ihr das Aktionspotential vorausgeht. Hier stellt sich die weitere Frage: Was soll eigentlich dem freien Willen unterworfen sein? Der Input (die Bedingungen bzw. Gründe) für den Entscheidungsprozess, oder kann der Prozess selbst willentlich gesteuert werden, oder besteht unsere Freiheit nur darin, die getroffene Entscheidung nachträglich noch zu korrigieren? Fragen wir also unseren Fußballroboter. Die Basis (Input) für den Entscheidungsprozess ist auf der obersten Ebene die Gesamtsituation, also die eigene Position (wo bin ich) relativ zum Ball, den anderen Spielern und zum Tor. Daraus wird die Entscheidung (E1) abgeleitet, was jetzt zu tun ist. Ist diese Entscheidung getroffen, wird die nächst niedere

Entscheidungsebene mit E1 als Input versorgt. Dort läuft wieder ein Entscheidungsprozess ab über die jetzt einzuleitenden konkreten Bewegungen. Dieser Prozess liefert seinerseits eine oder mehrere simultane Entscheidungen (E2). Und diese Entscheidungen lösen dann u.a. die Entscheidung dafür aus, dass jetzt (und nicht erst in zwei Sekunden!) der Schussmechanismus zu aktivieren ist (das „auf den Knopf drücken“ in den Experimenten von Libet), weil sich der Roboter jetzt genau in der richtigen Schussposition befindet. Es gibt also nicht eine Entscheidung die frei oder unfrei ist, sondern es gibt eine Hierarchie miteinander wechselwirkender Entscheidungsprozesse! Auf der obersten Ebene geht es um das langfristige (nicht explizit definierte, sondern nur in den Verhaltensanweisungen implizit verankerte) Ziel, ein Tor zu schießen, auf den darunter liegenden Ebenen um konkrete Maßnahmen zur Zielerreichung. Der Unterschied liegt also nur in der *Art* der Entscheidungen. Wenn Mynarek jetzt argumentiert, dass das Bereitschaftspotential vom Wollen noch umgelenkt werden kann „in ganz analogen und für jeden nachvollziehbaren Situationen des Alltags“ (Mynarek, S.122), dann gilt dies natürlich analog auch für unsere Fußballroboter. Wenn der Torwart beispielsweise von der rechten zur linken Ecke wechselt, muss der Roboterschütze auch seine Entscheidung, den Schuss auf die linke Ecke zu positionieren, rückgängig machen. Die Frage ist dann nur, wie entscheidet er, dass er sich umentscheiden will? Die Antwort lautet: er muss permanent seine Entscheidungen an der aktuellen Situation auf dem Spielfeld spiegeln und sie bei Bedarf korrigieren (überdenken?). Dafür gibt es verschiedene Alternativen. Er

kann beispielsweise mehrmals pro Sekunde eine Neuplanung durchführen (das Planungsintervall sollte besser auf den verschiedenen Ebenen unterschiedlich groß sein) und seine vorausgegangenen Entscheidungen „vergessen“ (sog. Neuaufwurfsprinzip). Er kann aber auch diese Entscheidungen nur korrigieren (sog. Fortschreibungsprinzip), also permanent überprüfen, ob sie noch den aktuellen Anforderungen entsprechen. Dabei gibt es wiederum zwei Alternativen. Er kann entweder bereits getroffene Entscheidungen verwerfen oder er kann den Prozess der Entscheidungsfindung selbst unterbrechen (Interrupt). Unsere These für das menschliche Gehirn lautet, dass wir zumindest bei kurzfristigen Entscheidungen, die sich im Entscheidungspotential spiegeln, nicht den Entscheidungsprozess selbst unterbrechen, sondern uns allenfalls nachträglich neu bzw. umentscheiden. Die *nachträgliche Korrektur* einer Entscheidung ist dann relevant, wenn ich aus der Entscheidung selbst die Erkenntnis gewinne, dass die Entscheidung falsch war (feed back), die *Unterbrechung* des Entscheidungsprozesses selbst (noch ehe das Ergebnis feststeht) ist dann relevant, wenn ich aus veränderten Inputbedingungen erkenne, dass das Ergebnis des laufenden Entscheidungsvorganges nicht zielführend sein kann (feed forward) (der Torwart ist in die andere Ecke gegangen) oder ein hoher Entscheidungsdruck für eine sofortige Umorientierung besteht. Ein Beispiel aus der maschinellen Prozesssteuerung und -überwachung sind sog. Alarmmeldungen. Geht ein Alarm ein, sind alle laufenden Programme sofort zu unterbrechen und der Alarm vorrangig zu behandeln, ehe es zu spät ist. Dies gilt analog auch für das menschliche Gehirn (ein Kind läuft

vor meinem Auto auf die Straße). Außerdem lernen wir von unseren Robotern, dass die Änderung von Entscheidungen auf unterschiedlichen Ebenen ausgelöst werden kann. Wurde die Ballposition falsch eingeschätzt (ein Roboter kann irren), muss auf der Zwischenebene die Stellung zum Ball korrigiert werden. Die neue Position wird dann nach oben gemeldet (bottom up) und ist dort bei den weiteren Planungen mit zu berücksichtigen. Hat sich die Position des Torwarts dagegen geändert, muss die Planung auf der obersten Ebene über den Haufen geworfen und dies nach unten (top down) weitergemeldet werden. Die Probleme des Roboters sind also vergleichbar mit denen des Menschen und er löst sie, ohne dass es eines freien Willens bedarf. Könnte man jetzt den freien Willen vielleicht noch daran festmachen, dass der Fußballroboter gar nicht anders kann, als auf dem Spielfeld Fußball spielen, denn diese Aufgabe hat ihm sein Entwickler zugedacht. Sein Kollege, der im Wettbewerb Penalty Kick (entspricht dem Elfmeterschießen) teilnimmt, braucht eine andere Strategie, um Erfolg zu haben, und muss deshalb auch andere Entscheidungen treffen. Wenn er aufs Tor schießt, steht der Torwart in der Mitte. Für welche Ecke soll er sich also „entscheiden“? Diese Frage lässt sich aus der Umweltanalyse allein nicht mehr eindeutig beantworten, denn die Bedingungen für die Alternativen, und damit die „Gründe“ für die Auswahl einer bestimmten Ecke, sind gleichwertig (das typische Dilemma von Buridans Esel). Bieri postuliert: „Unsere Freiheit ist die Freiheit, uns für oder gegen etwas entscheiden zu können“ (Bieri, S.54). Unser Roboter kann sich hier einer „grenzenlosen“ Freiheit erfreuen. Folgt aus der *Möglichkeit* der frei-



en Entscheidung aber auch schon, dass der Entscheidungsprozess selbst ein freier ist, und ggf. frei wovon? Unser Torhüter muss sich hier schon etwas „überlegen“. Und ist er so programmiert, dass seine Schüsse „vorhersagbar“ sind, dann hat ein „schlaues“ Torwart, der ihn „durchschaut“, ein leichtes Spiel. Unser Kommentar: Jeder muss sehen, wie er in seiner (Mikro-)Welt zurechtkommt, und dafür ist er optimiert. Das hat aber nichts mit Willensfreiheit zu tun. Ich kann mich auch nicht entscheiden, mit den Fischen im Meer zu schwimmen, weil ich dafür nicht genetisch programmiert bin.

Man wird hier an Jerry A. Fodors Ansatz der 3 Ebenen erinnert. Die oberste Ebene bildet dabei intentionale Zustände mit gegenseitigen Kausalwirkungen ab, die zweite Ebene realisiert diese Zustände durch Symbolverarbeitungsprozesse und die dritte Ebene ist dann die Implementations-ebene, die Ebene der physischen Realisierung. Eine analoge Aufgabenteilung finden wir auch im menschlichen Gehirn zwischen präfrontalem Cortex (Handlungsplanung und -vorbereitung), prä- und supplementär motorischen Arealen (Bewegungsablaufplanung und Handlungsauslösung) und dem primären motorischen Cortex (Feinsteuerung der Bewegungen). Und ergänzend lassen sich Handlungen aus Gründen durch Lernalgorithmen verbessern. Beim autonomen Lernen (ohne Lehrer) wird der Erfolg einzelner Aktionen bewertet (wurde ein Tor erzielt? Hat der angespielte Mitspieler den Ball erreicht?). Daraus lässt sich ableiten, ob Aktionen erfolgreich waren oder nicht. Diese Entscheidung trifft bei Lebewesen die Evolution (Überleben oder Aussterben) oder das menschliche Bewusstsein (Überleben oder Aussterben unserer Hy-

pothesen). Beim nächsten mal wird dann in einer vergleichbaren Situation, das heißt, beim Vorliegen derselben Gründe (Bedingungen), ein erneuter Versuch mit einer anderen Aktion unternommen, oder dieselbe Aktion anhand zusätzlicher Kriterien vorab noch einmal kritisch hinterfragt (liegt der Ball tatsächlich exakt dort, wo er vermutet wird) (vgl. z.B. [www.brainstormes.uos.de](http://www.brainstormes.uos.de) mit weiterführenden Literaturangaben).

Diese Beispiele zeigen, dass die Frage, ob sich Handlungen auf Ursachen (der physikalische Schuss hat einen physikalischen Auslöser) oder Gründe (warum wird der Schuss gerade jetzt ausgelöst und warum gerade in diese Richtung?) zurückführen lassen, und ob einer Entscheidung ein Entscheidungsfindungsprozess vorausgeht, für die Frage, ob den daraus abgeleiteten Aktionen „freie Willensentscheidungen“ zugrunde liegen, völlig irrelevant ist! Unser Roboter schießt in die linke Ecke, weil der Torwart in der rechten Ecke steht. Er könnte aber auch anders handeln und in die rechte Ecke schießen, wenn es dafür einen Grund gäbe, zum Beispiel wenn der Torwart in der linken Ecke steht. Seine Entscheidung trifft er, weil es so im Programmcode steht und nicht, weil er einen freien Willen hat. In unseren entscheidungsfreudigen Fußballrobotern werden wir jedenfalls erfolglos nach einem „Modul für den freien Willen“ oder irgendeinem „Geist“ suchen. Natürlich gibt es für den Fußballroboter einen Grund, in die linke Torecke zu schießen (s.o.), dies bedeutet aber noch lange nicht, dass wir im Programmcode (oder in der Zentraleinheit des Computers) irgendwo Wünsche, Absichten oder Überzeugungen ausmachen können. Dies gilt übrigens analog für instinktgeleitete Tiere. Wir finden diese

Wünsche nur in den Gehirnen derjenigen, die den Roboter programmiert haben. Im Roboter selbst sind sie nur noch implizit verankert, sie bestimmen sein Verhalten in Abhängigkeit von gewissen Randbedingungen. Trotzdem können wir den Roboter natürlich so konzipieren, dass er uns die Gründe für seine Entscheidungen (sowohl vor als auch nach seinem Schuss) nennen kann. Und wir können ihn sogar so programmieren, dass er seine Gründe mit den beiden Worten „ich will“ oder „ich wollte“ verknüpft. Dies setzt allerdings voraus, dass diese Gründe aus ihrer impliziten Verankerung im Programmcode herausgelöst und explizit z.B. als Daten in Tabellen gespeichert werden. Diese (Entscheidungs-)Tabellen geben dann an, unter welchen Bedingungen welche Aktionen auszuführen sind. Diese Tabellen bzw. allgemeiner gesagt, eine Wissensbasis, können dann nicht nur herangezogen werden, um die Robotermechanik anzusteuern, sondern auch, um einen „Erklärungsmechanismus“ zu steuern, der schriftliche oder verbale Erklärungen für die von den Philosophen geforderten Begründungen für die Roboterentscheidungen liefert. Die Künstliche Intelligenz (KI) hat hierfür die Rückwärtsverkettung (backward reasoning) bzw. ganz allgemein eine sog. Erklärungs-komponente entwickelt. Schon vor 35 Jahren (!) hat Terry Winograd mit seinem Programm SHRDLU demonstriert, dass künstliche Systeme ihre Entscheidungen und Aktionen in natürlicher Sprache begründen können. Sein Programm konnte in einer simulierten Klötzchenwelt mit einem simulierten Roboterarm menschliche Anweisungen befolgen (find a block which is taller than the one you are holding and put it into the box), Fragen beantworten (what does the box contain?) und

seine Aktionen begründen (Person: why did you clear off that cube? Computer: to put it on a large green cube. Person: why did you do that? Computer: to stack up the red cube, a large red block and a large green cube. Person: why did you do that? Computer: because you asked me to) (vgl. <http://hci.stanford.edu/winograd/shrdlu/index.html>). Umwelt- und Körperrepräsentationen liefern hier die Bedingungen bzw. Prämissen, aus denen ein Roboter-gehirn Konklusionen ableitet, die dann in Roboterhandlungen umgesetzt werden. Ist das System in der Lage, seine Schlussfolgerungskette rückwärts nachzuvollziehen, kann es seine Aktionen auf Umweltsituationen und interne Körperzustände zurückführen und so durch rein syntaktische Algorithmen semantische Beziehungen zwischen Ereignissen und Zuständen in der realen Welt und seinen eigenen Reaktionen erkennen (weil der eigene Energievorrat gering ist und der Ball näher beim Mitspieler M liegt, bewegt sich der Mitspieler M zum Ball). Eines Geistes bedarf es zur Herstellung derartiger semantischer Beziehungen nicht, und es wird dabei auch nicht gegen die Gesetze der Physik verstoßen.

Die Umweltrepräsentation im Computer-gehirn des Roboters bedeutet dem Computer selbst überhaupt nichts. Aber das Roboter-gehirn erkennt, dass sie für den Roboter (genauer: für den vom Computer überwachten und gesteuerten Roboter-körper) von Bedeutung sind und dieser deshalb angemessen auf sie reagieren muss. Es ist sogar so, dass der Computer von sich selbst überhaupt nichts weiß, sondern nur von dem, was an Repräsentationen der realen Welt in seinem Arbeitsspeicher existiert. Aber gilt dies nicht analog auch für den Menschen? Das ist die

entscheidende Frage, die von der Philosophie bisher überhaupt noch nicht thematisiert wurde.

Jetzt stellt sich noch die weitere Frage, wer trifft denn hier letztlich die Entscheidungen? Der Roboter, oder sein „Gehirn“, also der Computer, der ihn steuert, oder die Entwickler, die in den Tabellen die Entscheidungskriterien hinterlegt haben? Die naheliegende Antwort lautet: die Entwickler entscheiden und der Fußballroboter ist nur ihr „verlängerter Arm“. Problematisch wird es aber, wenn die Entwickler gar nicht diese Regeln vorgeben, sondern nur einen Algorithmus, nach dem der Roboter durch Versuch und Irrtum bzw. Erfolg oder Misserfolg selbst lernt, wie er sich am besten in charakteristischen Situationen verhält. Zumindest auf den unteren Kompetenzschichten werden solche Algorithmen heute schon eingesetzt. Jetzt kontert die Philosophie: Aber der Lehrmeister des Roboters ist der Mensch (sog. Lernen durch Unterweisung = supervised learning) und die Kriterien, was als Erfolg bzw. Misserfolg beim Lernen zu gelten hat, die werden ebenfalls vom Entwickler vorgegeben. Darauf lautet die Antwort, dass in diesem Bereich heute bevorzugt das sog. reinforcement learning eingesetzt wird, ein Lernen *ohne* Unterweisung. Dabei werden Verhaltensweisen danach klassifiziert, ob sie richtig oder falsch sind, ohne dass angegeben wird, *warum* sie richtig oder falsch sind. Schüsse, die im Tor landen gelten dann als richtig, solche die daneben gehen, als falsch, ohne dass danach gefragt wird, warum der Schuss daneben ging. Das System soll dann einfach lernen, die Verhaltensweisen anzuwenden, die erfolgreich waren und die weniger erfolgreichen verlernen. Daraus entwickeln sich dann „die persönli-

chen Präferenzen“ eines Roboters. Und „frei ... handelt eine Person, die in einer bestimmten Situation eine Option x statt einer Option y wählt, genau dann, wenn sich die Entscheidung für x und gegen y auf die personalen Präferenzen der Person zurückführen lässt“ (Pauen, S. 96). Handelt unser Roboter, der seine Schüsse auf Basis eigener Erfahrungen platziert, also frei? Unsere Antwort lautet: nein! Er handelt lediglich auf Basis vergangener Erfahrungen. Pauens Kommentar: „Voraussetzung dafür ist, dass die eigentliche Entscheidung am Ende einer Kausalkette auf den Akteur und nicht ausschließlich auf äußere Umstände zurückzuführen ist“ (Pauen, S. 97). Auch diese Bedingung erfüllt der Roboter, denn je nachdem, was er gelernt hat, wird er in einer bestimmten Situation so oder so entscheiden. Allerdings betont Pauen: „Zentral ist dabei die Vorstellung, dass eine Person die Möglichkeit haben muss, sich wirksam gegen eine Präferenz zu entscheiden“ (Pauen, S. 97). Bleibt nur zu fragen, warum die Person sich gegen eine bestimmte Präferenz entscheidet. Wenn es dafür einen Grund gibt, dann steht diese Alternative auch unserem Fußballroboter offen, ansonsten müssen wir noch einen (Pseudo-)Zufallsgenerator einbauen.

Hier kommt also das *Prinzip der operanten Konditionierung* zur Anwendung. Und jetzt können wir näher hinterfragen, woher eigentlich Tiere und Menschen wissen, was für sie gut oder weniger gut ist, woher also die Ziele stammen, die wir verfolgen. Die Antwort hierauf liefert die Evolution oder anders gesagt, „der blinde Uhrmacher“ in der Terminologie von Richard Dawkins. Er hat einfach nach demselben Prinzip (des reinforcement learning) diejenigen Verhaltensweisen aussterben

lassen, die weniger erfolgreich waren (und zwar ohne explizite Erklärung) und so haben wir gelernt, was für uns gut und erfolgreich ist. Und in der Sprache der Philosophie leiten sich daraus unsere Wünsche und Ziele ab. Bleibt noch zu fragen: wozu bedarf es eigentlich dieser Wünsche? Wir könnten doch, wie die Tiere und unsere Roboter auch, gleich in die Tat umsetzen, was in einer bestimmten Situation das jeweils Beste für uns ist. Wozu bedarf es also der Gefühle, Wünsche, Ziele, Pläne, des Willens und der sog. Entscheidungsfreiheit? Dies entspricht der Frage, ob gut optimierte Instinkthandlungen (und genau das sind die Programmroutinen, die im Roboter unter vorgegebenen Bedingungen aktiviert werden) nicht auch für unseren Überlebenskampf ausreichen würden. Die Antwort ist bekannt. Die Fähigkeit, die Konsequenzen unserer Handlungen, insbesondere in neuartigen Situationen, im Geiste vorab durchspielen zu können, hat den Vorteil, dass mit den Worten von Karl Popper „nicht wir, sondern unsere Hypothesen sterben“, falls unsere Handlungen nicht erfolgreich sind. Deshalb bringt es Vorteile mit sich, wenn wir erst im Geist handeln und die Konsequenzen abschätzen und nicht sofort zur Tat schreiten. Wir lernen dann nicht durch Versuch und Irrtum, sondern durch Einsicht!

Das haben inzwischen auch die Entwickler der Roboterfußballer erkannt. In der Sprache der Robotik arbeiten ihre Systeme bisher wie niedere Tiere, rein *reaktiv*, d.h. sie reagieren unmittelbar auf die auf sie einströmenden „Sinneswahrnehmungen“. Jeder Informationsinput wird analysiert und daraus abgeleitet, wie zweckmäßigerweise auf ihn reagiert werden sollte.

Die nächste Entwicklungsstufe lässt aber die Roboter *kognitiv* agieren, d.h. sie sollen erst „überlegen“, was eine angemessene Reaktion sein könnte, eventuell mehrere Alternativen „durchdenken“ und in einer weiteren Ausbaustufe auch die mögliche Reaktion der Gegenspieler in ihr Kalkül einbeziehen, die sie dazu am besten vorab beobachten sollten, um ihr Verhalten besser einschätzen zu können. Wie das praktisch geht, haben uns schon der Schachcomputer „Deep Blue“ und seine Nachfolger vorgeführt. Auch sie analysieren die möglichen Züge, bewerten sie und treffen dann eine Entscheidung. Bei einer derartigen Vorabplanung von Aktionen müssen die Alternativen bewertet und die Fehlentscheidungsrisiken (Fehler der 1. und 2. Art in der Terminologie der Statistik) abgeschätzt werden. Der nächste Schritt zur Implantierung von „Gefühlen“ wird also sicher nicht mehr allzu lange auf sich warten lassen. Peter Bieri sagt: „Die Freiheit des Willens liegt darin, dass er auf ganz bestimmte Weise bedingt ist: durch unser Denken und Urteilen“ (Bieri, S. 80). Können wir unseren Fußballrobotern also einen freien Willen zugestehen, wenn sie ihre Entscheidungen aus „Gedanken“ und „Urteilen“ ableiten? Immerhin behauptet Bieri ja: „Die Gewissheit des freien Willens rührt von der Gewissheit des Überlegens her: Ich überlege, also bin ich frei“ (Bieri, S. 182). Allerdings äußert er sich an anderer Stelle etwas skeptischer. „Der Wille, der eine Handlung anstößt, ist selbst etwas, das angestoßen wird, und das macht ihn zu etwas, das er nur auf den ersten Blick nicht zu sein schien: zu einem bloßen Geschehnis. ... (Der Wille) ist, wenn man ihn als etwas Bedingtes betrachtet, gar nicht mehr wirklich ein Wille, und deshalb vermag er auch keine Urhe-

berschaft zu garantieren“ (Bieri, S. 195). Wir meinen, dass auch die Kriterien von Bieri aus unserem Maschinenfußballer kein Artefakt mit einem freien Willen machen.

Bleibt jetzt noch die Frage, wie wir zu der Behauptung von Hans Jonas stehen: „Ein Computer hat erst gedacht, wenn ein Denker das Resultat in Empfang genommen hat“ (Jonas, S. 14 in: Beiersdörfer). Unser Kommentar hierzu lautet: Bei unseren Roboterfußballern gibt es keinen Denkenden, der ein Resultat in Empfang nehmen könnte. Das Resultat „der Überlegungen des Roboterhirns“ nehmen vielmehr die niedrigeren Kompetenzschichten entgegen und wandeln es in angemessenes Verhalten um – ganz wie auch beim Menschen! Ob das Roboterhirn „verstanden“ hat, worauf es ankommt, ob es also richtig gedacht hat, zeigt sich dann spätestens bei der nächsten Aktion des Roboters. Und deren Konsequenzen kann das Roboterhirn selbst erkennen und daraus eigenständig ableiten, ob es richtig gedacht, d.h. ob es die Situation richtig eingeschätzt und daraus die richtigen Schlussfolgerungen gezogen hat. „Gedankenketten“ die nicht zielführend sind, werden dann einfach durch reinforcement learning ausgebootet. Wir können natürlich auch mit Ludwig Wittgenstein sagen: „Aber eine Maschine kann doch nicht denken! – Ist das ein Erfahrungssatz? Nein. Wir sagen nur vom Menschen, und was ihm ähnlich ist, es denke ... es habe Empfindungen; es sehe; sei blind; höre; sei taub; sei bei Bewusstsein, oder bewusstlos“ (Wittgenstein zitiert in Metzinger, S. 166).

Die Überlegenheit des Menschen können wir aber vorerst noch an seinem Selbstbewusstsein festmachen. Wir haben ein Ich-Bewusstsein, was den Robotern sicher noch fehlt. Dies bringt uns zur näch-

sten Frage. Wer trifft eigentlich die Roboterentscheidungen? Trifft sie der Roboter, oder trifft sie sein Gehirn (ein Computer!) oder doch sein Entwickler, oder gibt es womöglich doch irgendwo ein verborgenes Ich, das die letzte Entscheidung trifft? Der Einfluss des Entwicklers tritt, wie bereits gezeigt, immer mehr in den Hintergrund. Er legt „nur noch“ den (genetischen!) „Programmcode“ fest, der, wie beim Menschen, in immer höherem Maße durch Lerneffekte (Lebenserfahrung und „Erziehung“) des Roboters sowie dessen fallspezifische „subjektive Lagebeurteilungen“ zurückgedrängt wird. Wenn er dann auch noch seinen aktuellen Energievorrat (die Batterie schwächelt) und seinen Körperzustand (die Mechanik ist heißgelaufen und wird demnächst wegen Überhitzung abgeschaltet) erfasst und in seine Entscheidungsfindung einbezieht, gewinnt er zunehmend an Autonomie und sein Verhalten wird für einen außenstehenden Beobachter, seinen Entwickler eingeschlossen, immer weniger vorhersehbar (warum hat er jetzt bloß wieder ...?). Und jetzt können wir auch die Frage stellen, wer eigentlich verantwortlich ist, wenn unser Roboter einen Fehler begeht und beispielsweise irgendeinen Gegenstand am Spielfeldrand mit dem Ball verwechselt und ihn als Schussobjekt missbraucht. Wer muss für den Schaden aufkommen den ein solcher Roboter verursacht? Ist das sein Entwickler, sein Trainer (Erzieher), oder die Umgebung in der er trainiert hat und von der er falsch „gepol“ worden ist?

Die strittige Frage, wer denn nun beim Menschen die Entscheidungen trifft, der Mensch oder sein Gehirn, wollen wir am Beispiel der Fußballroboter näher diskutieren. Zuvor einige Stimmen hierzu. Der



Small Size Liga



Middle Size Liga

(Fotos: Zimmermann)

Hirnforscher Gerhard Roth sagt: „Mir scheint der Satz »Nicht das Ich, sondern das Gehirn hat entschieden!« korrekt zu sein, denn »eine Entscheidung treffen« ist ein Vorgang, dessen Auftreten objektiv überprüfbar ist“ (Roth 2004, S. 77). Dem schließt sich der Biokybernetiker Holk Kruse an: „Die Hauptaufgabe, die Gehirne zu erledigen haben, ist die Entscheidung darüber, was jeweils als nächstes zu tun ist. Dabei wägt dieses System die Argumente ab und trifft eine Entscheidung“ (Kruse, S.224). Analog schreibt der Psychiater Spitzer: „Das Gehirn ist – gerade aus der Sicht der empirisch-naturwissenschaftlichen Forschung – nicht Spielball, sondern Spieler!“ und „Das Gehirn bestimmt sich selbst aktiv“ (Spitzer, S. 283) sowie: „Die Unterscheidung von „wir“ und „unser Gehirn“ ... (ist) wenig hilfreich ..., denn das Gehirn macht die Person (und damit das Subjekt von Entscheidungen und Handlungen) ja gerade aus“ (Spitzer, S. 307). Der Protest von Seiten der Philosophie lässt nicht lange auf sich warten. Wir führen zunächst noch einmal die bereits oben kommentierte Aussage von Wingert an: „Wenn man, wie Neurobiologen es oft tun, sagt: »Das Gehirn urteilt und entscheidet«, dann müsste im Fall der Herstellung eines Wahrheitsbezugs gezeigt werden können, wie neuronale Prozesse einen fehlbaren

Anspruch erheben, das heißt wie sie irren können“ (Wingert S. 198). Dies haben wir oben bereits gezeigt. Thomas Buchheim behauptet: „Dadurch dass ich mit dem Gehirn denke, denkt aber noch nicht das Gehirn statt meiner“ (Buchheim, S. 161). Denkt also bei einem kognitiv konzipierten Maschinenfußballer der Computer (sein Gehirn) oder denkt der gesamte Roboter? Wir fragen noch provozierender, gibt es bei diesem Roboter vielleicht sogar ein Ich, das denkt? Wir nehmen die Antwort vorweg: Im Programmcode des Roboters finden wir kein Ich! Der Lektor Helmut Mayer meint: „Es ist nicht etwa empirisch falsch, vom denkenden, fühlenden, wahrnehmenden etc. Hirn zu sprechen: Es ist ein Kategorienfehler. Dieser »mereologische Trugschluss« vom Subjekt auf einen seiner Teile, auf das Gehirn nämlich, ist grundlegend für die meisten anderen neurowissenschaftlichen Verzeichnungen“ (Mayer, S. 211). Ist also das Gehirn des Menschen und das des Fußballroboters nur irgendeines der vielen Teile des Gesamtsystems? Und ist dieses Gesamtsystem ein Subjekt?

Fragen wir jetzt noch einmal, entscheidet der Roboter als Gesamtheit oder entscheidet der Computer „in seinem Kopf“ oder womöglich ein „out of body“ implantierter Computer?

Beim RoboCup treten u.a. Fußballroboter der Middle Size Liga und der Small Size Liga an. Zur Verdeutlichung dienen obenstehende Bilder.

In der Middle Size Liga hat jeder Roboter sein (Computer-)Gehirn „an Bord“ und ebenfalls sein „Wahrnehmungsorgan“, eine Rundumkamera mit einem 360 Grad Blickwinkel, die einen Überblick über das gesamte 8 x 12 Meter große Spielfeld erlaubt. Die Small Size Liga dagegen operiert mit Robotern, die aussehen wie umgestülpte Kochtöpfe auf vier Rädern. Eine Kamera über dem Spielfeld ermittelt die Position der einzelnen Roboter, der Tore und des Balls aus einer Art von „out of body“ Sicht, d.h. aus der Vogelperspektive (auf ähnliche Erscheinungen stoßen wir auch bei Nahtodeserlebnissen des Menschen). Sie erfasst „mit einem Blick“ das gesamte Spielgeschehen und liefert pro Sekunde etwa 30 Bilder, die jeweils von einem zentralen Computerprogramm (je Mannschaft) ausgewertet und in Befehle für die Aktionen der einzelnen Roboter umgewandelt werden. Die Roboter werden dann von der Zentrale über Funk (drahtloses Local Area Network, WLAN) angesteuert. Interpretieren wir die gesamte Mannschaft als eine Einheit, dann haben wir hier ein „Ich“ vor uns, das aber keine körperliche Einheit mehr bildet. Die Situation ist vergleichbar mit meinem Klavierspiel, wo ich meine Finger und die Tastatur sozusagen „von oben“ beobachte und mein Gehirn dann die einzelnen Finger ansteuert. Die einzelnen Roboter haben hier also überhaupt kein „Gehirn“ und sie sehen und entscheiden auch nichts. Sie sind „willenlose“ Werkzeuge, die von einem zentralen Computer befehligt werden, der alle eingehenden Informationen auswertet und dann zentral eine Entscheidung

für alle Roboter (wer hat was zu tun) trifft und diese Entscheidung 30 mal pro Sekunde aktualisiert. Wir erinnern hier an Gerhard Roths Aussagen: „Das Gehirn generiert mit der Ausbildung eines Ich einen »virtuellen Akteur«, dem ein Körperschema und ein Ort im Raum zugeschrieben wird und der zum scheinbaren Träger der Willkürhandlungen wird“ (Roth S. 204 in: Pauen/Roth 2001) und „nicht nur die von mir wahrgenommenen Dinge sind Konstrukte in der Wirklichkeit, *ich selbst bin ein Konstrukt*. Ich komme unabweisbar in dieser Wirklichkeit vor. Dies bedeutet, dass das reale Gehirn eine Wirklichkeit hervorbringt, in der ein Ich existiert, das sich als *Subjekt* seiner mentalen Akte, Wahrnehmungen und Handlungen erlebt, einen Körper besitzt und einer Außenwelt gegenübersteht“ (Roth 1997, S. 329). Im vorliegenden Fall (Small Size Liga) besteht der Akteur aus mehreren Einheiten. Wer, so fragen wir jetzt, ist hier im Sinne von Mayers Argumentation (s.o.) das Subjekt? Wir verweisen auf eine weitere Aussage von Roth. „So gibt es Patienten, deren Identitätsbewusstsein gestört ist und die glauben, mehr als eine Person zu sein, oder die nicht wissen, wer sie sind. Andere glauben, an mehreren Orten gleichzeitig zu sein, und finden nichts Merkwürdiges dabei. Wieder andere haben das Gefühl, dass sie nicht im »richtigen« Körper stecken, und noch andere glauben, dass ihre Gedanken und Handlungen von fremden Mächten gelenkt werden“ (Roth 2003, S. 128).

In der Middle Size Liga liegen die Verhältnisse klarer. Jeder Roboter ist hier eine Einheit mit einem eigenen (Computer-)Gehirn und das bestimmt, was zu tun ist und die Robotermechanik gehorcht „willenlos“. Allerdings hatten wir bereits erwähnt, dass wir es hier mit „sozialen Wesen“ zu

tun haben, denn unsere Fußballagenten stimmen sich vor jeder Aktion untereinander ab und entscheiden gemeinsam, wer was zu tun hat. Sie kommen dabei, wie das Gehirn, ohne eine zentrale letzte Entscheidungsinstanz und ohne ein „Ich“ aus. Sie können uns also demonstrieren, wie man das Bindungsproblem löst und auch dass es weder eines Ichs noch eines freien Willens bedarf, um in komplexen Umwelten zielführend agieren zu können. In keinem Programm findet sich eine Anweisung oder eine Zielvorgabe der Art „Ihr müsst versuchen, möglichst viele Tore zu schießen“. Und trotzdem schießen sie aufs Tor, ohne dass damit ein Wille verbunden ist! Dafür reichen simple (oder auch etwas komplexere) »wenn ... dann Beziehungen« aus, genauso wie im Gehirn. Alles andere ist „Begleitmusik“. Und die Gründe für ihre Aktionen finden wir in den Bitmustern des Arbeitsspeichers im Computergehirn. Dabei kann die Belegung eines einzigen Feldes den Ausschlag geben oder aber ein sehr komplexes Belegungsmuster, das zum Beispiel eine gesamte aktuelle Situation auf dem Spielfeld repräsentiert. Und gleiches gilt für das Gehirn. In manchen Fällen kann dieses Muster direkt eine Aktion auslösen, in anderen Fällen muss eine Situationsanalyse und Bewertung (ein Denkprozess) vorgeschaltet werden. Jetzt schließt sich noch die Frage an, ob unsere Roboterfußballer eventuell doch über eine rudimentäre Form von Bewusstsein verfügen. Nach Meinung von Michael Tye „... ist ein mentaler Zustand genau dann phänomenal bewusst, wenn er einen für kognitive Funktionen bereitstehenden abstrakten, nicht-begrifflichen, intentionalen Gehalt hat“ (Tye, S. 95f). Nach seiner PANIC Theorie (poised, abstract, non-conceptual, intentional concept) sind phä-

nomenale Zustände einfach nur sensorische Repräsentationen und deren Inhalte werden zu Erlebnisqualitäten. Interpretieren wir die Repräsentation des Spielgeschehens im Arbeitsspeicher des Computers, der den Roboter steuert, als mentalen Zustand mit intentionalem Gehalt, müssten wir auch dem Roboter phänomenales Bewusstsein zugestehen. Bei den humanoiden Robotern ist im Computer nicht nur sein jeweils aktueller Standort auf dem Spielfeld, sondern darüber hinaus auch noch sein eigener aktueller Zustand repräsentiert, also ob er gerade steht, geht, sich dreht oder einen Schuss abgibt. Und falls er das Gleichgewicht verliert und umfällt, melden ihm seine Propriozeptoren, ob er auf dem Bauch, auf der Seite oder auf dem Rücken liegt. Dieses Wissen benötigt er, damit er wieder von alleine aufstehen kann, denn wer das nicht ohne fremde Hilfe schafft, darf im Wettbewerb gar nicht mitspielen. Haben die humanoiden Fußballspieler also womöglich sogar ein Selbstbewusstsein? Die Definitionen der Philosophie für Bewusstsein geben hierauf keine eindeutigen Antworten. Aus Sicht des Verfassers haben die (heutigen) Roboter weder ein phänomenales noch ein Selbstbewusstsein aus Gründen, deren Diskussion den Rahmen dieser Abhandlung sprengen würde. Was unsere Fußballroboter (und andere autonome Agenten) von reinen Computern bzw. Computerprogrammen oder KI-Methoden unterscheidet ist, dass sie über eine *Körperhaftigkeit* verfügen und die „Robotergerirne“ „ihren Körper“ mit seinen Aktionen „erfahren“, also wissen, was er tut, in welchem Zustand er sich befindet und ihren Input von diesem Körper bzw. dessen „Wahrnehmungsorganen“ erhalten und ihren Output wieder an die-



sen Körper als „Handlungsanweisungen“ abgeben. Dieser Körper steht in einer dynamischen Beziehung mit seiner Umwelt und wird auch in dieser Beziehung vom Computergehirn „gesehen“. Diese *Situiertheit* des Roboters in seiner Umwelt führt dazu, dass sich sein Computergehirn nicht mehr mit abstrakten, von Menschen vorgegebenen Problemen beschäftigt, sondern mit der Interaktion des Roboterkörpers mit seiner Umwelt. Und deshalb werden derartige Systeme immer menschlicher und wir können von ihnen manches lernen, auch über die Wechselbeziehung von Körper und Geist. Die Entwickler der Roboterfußballer jedenfalls haben sich das ehrgeizige Ziel gesetzt, im Jahr 2050 mit ihren Schützlingen den dann amtierenden Fußballweltmeister (der Menschenfußballer) in einem 90 Minuten Spiel auf grünem Rasen nach den Regeln der FIFA zu schlagen. Bis dahin ist sicher noch viel Entwicklungsarbeit zu leisten. Eines scheint aber heute schon sicher: An der erforderlichen „Brain-Power“ wird es diesen Artefakten sicher nicht fehlen. Ausgehend vom sog. Mooreschen Gesetz (nach Gordon Moore, dem Erfinder des integrierten Schaltkreises und Gründer der Firma Intel), wonach sich in den letzten 40 Jahren alle 12 bis 24 Monate die Zahl der Transistoren auf einem integrierten Schaltkreis verdoppelt hat, prognostiziert Ray Kurzweil für das Jahr 2023 die Verfügbarkeit von 1.000 Dollar-Computern mit einem Siliziumäquivalent zum menschlichen Gehirn. Für 100 Milliarden Neuronen mit jeweils 1000 Verzweigungen im Gehirn benötigt er dafür 1 Million Milliarden Bits. Und ein solcher Computer wird zudem etwa 1 Milliarde mal schneller arbeiten als das menschliche Gehirn. Wenn wir das Mooresche Gesetz kritiklos wei-

ter in die Zukunft projizieren, können wir schließlich im Jahr 2060 für 1000 Dollar einen Computer bekommen, der die Leistung aller menschlichen Gehirne zusammen übertrifft. Hans Moravec ist mit seinen Prognosen etwas vorsichtiger und meint, wir erhalten um 2040 für 1000 Dollar Computersysteme mit der Leistung und Fähigkeit des menschlichen Gehirns (4. Roboter-Generation). Bleibt noch die Frage: Und wer programmiert ihnen die erforderliche Intelligenz ein? Die Antwort lautet: Man wird ihnen allenfalls noch eine gewisse philogenetisch/programmierte Grundausstattung mit auf den Weg geben. Alles Weitere müssen sie ontogenetisch selbst erwerben. Und was sie gelernt haben, kann man dann einfach herunterladen und an andere Roboter überspielen. Bleibt die Frage, ob damit auch ihr „Ich“ überspielt wird, und das „Ich“ als Summe der Erfahrungen eines Individuums auf diese Weise vervielfältigt und vielleicht sogar die Basis für ewiges Leben gelegt werden kann. Mit den in der vorliegenden Abhandlung aufgeworfenen Fragen wird zunehmend auch die Philosophie konfrontiert. Wie sagte Dennett doch so schön: „Ist einmal die Katze aus dem Sack, dann ist es vorbei mit dem schönen Zustand, in dem man eine Sache, die einen so gefesselt hat, nach eigenem Belieben verklären kann“ (Dennett, S. 38).

#### **Literatur:**

Beckermann, Ansgar: Analytische Einführung in die Philosophie des Geistes, Berlin, New York 1999.

Bieri, Peter: Das Handwerk der Freiheit. Über die Entdeckung des eigenen Willens, München 2001.

Buchheim, Thomas: Wer kann, der kann auch anders, S. 158-165 in: Christian

Geyer (Hrsg.): Hirnforschung und Willensfreiheit. Zur Deutung der neuesten Experimente. Frankfurt 2004.

Beiersdörfer, Kurt (Hrsg.): Was ist Denken? Gehirn-Computer-Roboter Paderborn, München, Wien, Zürich 2003.

Dennett, Daniel C.: Philosophie des menschlichen Bewusstseins, Hamburg 1994.

Mayer, Helmut: Ach das Gehirn S. 205-217. in: Christian Geyer (Hrsg.): Hirnforschung und Willensfreiheit. A.a.O.

Kruse, Holk: Ich bin mein Gehirn. Nichts spricht gegen den materialistischen Monismus. S. 223-228 in: Geyer Christian (Hrsg.): Hirnforschung und Willensfreiheit. A.a.O.

Kurzweil, R.: Homo S@piens. Leben im 21. Jahrhundert – Was bleibt vom Menschen? München 2000.

Lenzen, Wolfgang: Zombies, Zimbos und das „Schwierige Problem“ des Bewusstseins. S. 255-281 in: Esken, Frank; Heckmann, Dieter (Hrsg.): Bewusstsein und Repräsentation. Paderborn 1998.

Metzinger, Thomas (Hrsg.): „Bewusstsein. Beiträge aus der Gegenwartsphilosophie. Paderborn, München, Wien, Zürich 1996.

Moravec, H.: Mind children. Der Wettlauf zwischen menschlicher und künstlicher Intelligenz. Hamburg 1990.

Moravec, H.: Computer übernehmen die Macht. Vom Siegeszug der Künstlichen Intelligenz. Hamburg 1999.

Mynarek, Hubertus: „Materie und Geist: Gehirn – Bewusstsein – Willensfreiheit. Argumente und Gegenargumente. Aufklärung und Kritik Heft 1/2006, S. 118-132.

Pauen, Michael: Illusion Freiheit? Mögliche und unmögliche Konsequenzen der Hirnforschung. Frankfurt a.M. 2004.

Pfeifer, Rolf; Scheier, Christian: Understanding Intelligence. Cambridge/Massachusetts, London 2000.

Pauen, Michael; Roth, Gerhard (Hrsg.):

Neurowissenschaften und Philosophie. München 2001.

Ratey, John J.: Das menschliche Gehirn. Eine Gebrauchsanweisung. München 2003.

Roth, Gerhard: Das Gehirn und seine Wirklichkeit 1997.

Roth, Gerhard: Aus Sicht des Gehirns, Frankfurt a.M. 2003.

Roth, Gerhard: Worüber dürfen Hirnforscher reden – und in welcher Weise. S. 66-85 in: Geyer Christian (Hrsg.): Hirnforschung und Willensfreiheit. A.a.O.

Searle, John R.: Freiheit und Neurobiologie. Frankfurt a.M. 2004.

Spies, Markus: Unsicheres Wissen, Wahrscheinlichkeit, Fuzzy-Logik, neuronale Netze und menschliches Denken. Heidelberg, Berlin, Oxford 1993.

Spitzer, Manfred: Selbstbestimmen. Gehirnforschung und die Frage: Was sollen wir tun? Heidelberg 2004.

Tye, Michael: Das Problem primitiver Bewusstseinsformen: Haben Bienen Empfindungen? S. 91-122 in: Esken, Frank; Heckmann, Dieter (Hrsg.): Bewusstsein und Repräsentation. A.a.O.

Wingert, Lutz: Gründe zählen. Über einige Schwierigkeiten des Bionaturalismus S. 194-204 in: Geyer Christian (Hrsg.): Hirnforschung und Willensfreiheit. A.a.O.

*Zum Autor:*

**Gero Zimmermann**, Jahrgang 1942. Studium der Elektrotechnik mit Schwerpunkt Datenverarbeitung/Kybernetik an der TU Stuttgart sowie des Wirtschaftsingenieurwesens mit Promotion an der Universität Karlsruhe. Habilitation in Wirtschaftsinformatik und Privatdozent (2 Jahre) an der Universität Würzburg. Seit 10 Jahren Auseinandersetzung mit Themen der Kognition aus Sicht der Hirnforschung, Soziobiologie, Evolution, Philosophie, Chaos- und Komplexitätstheorie, Synergetik, Informatik u.a.