Maschinenautonomie – Überlegungen zum „Selbst“ eines Roboters

Markus Christen
Institut für Neuroinformatik
Universität Zürich / ETH Zürich

Übersicht

1. Kurze Erläuterung des Vorgehens.
2. Welche Aspekte bestimmen das „Selbst“ und warum ist Autonomie der relevanteste Aspekt bei Robotern?
3. Der Körper von Robotern.
4. Das „Bewusstsein“ von Robotern.
5. Autonome Roboter und die Frage nach der „Maschinenverantwortung“.
6. Beispiel für die Diskussion.

Vorgehen

Im Zentrum des Vortrags steht die Klärung von Begriffen im Zusammenhang mit dem „Selbst“ eines Roboters. Dazu möchte ich jeweils:

...zuerst kurz zusammenfassen, was die Philosophie zu diesem Begriff sagt.

...dies dann in Beziehung zum Gebrauch des Begriffs in der Robotik setzen

...und daraus jeweils Folgerungen im Hinblick auf die Kernfrage Ihres Proseminars – *der Mensch geht, der Roboter kommt?* – ziehen.

Das Selbst: Philosophie

Der Begriff „Selbst“ tritt erst in der neuzeitlichen Philosopie und dabei insbesondere im englischen Sprachraum auf:

- Locke: Selbst ~ Person (Essay concerning human understanding).
- Berkeley: Selbst ~ Seele, Geist (Philos. comment.)
- Hume: Selbst ~ Erfahrungen (Treaties of human nature).
- Fichte: Selbst ~ Selbstbestimmung (Wissenschaftslehre).
- Nietzsche: Selbst ~ Leib (Also sprach Zarathustra)

Psychologie: Selbst ~ empirische Eigensicht des Individuums, Gesamtheit der subjektiven Sicht der eigenen Person.

Drei Aspekte des „Selbst“

Ich denke, es lassen sich drei Aspekte des „Selbst“ unterscheiden:

1) Das Selbst in Bezug auf die materielle Existenz, den Körper, den Leib; kurz: **Körper**

2) Das Selbst in Bezug auf Personsein, inneres Erleben, Selbstbild; kurz: **Bewusstsein**

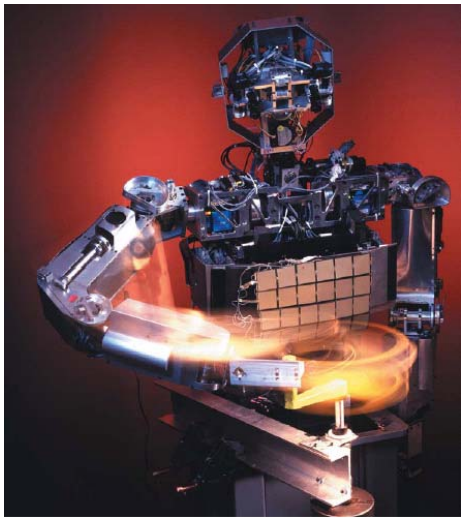
3) Das Selbst in Bezug auf Interagieren mit der Welt, freiheitliches Handeln und Übernehmen von Verantwortung; kurz: **Autonomie**

Warum ist Autonomie zentral?

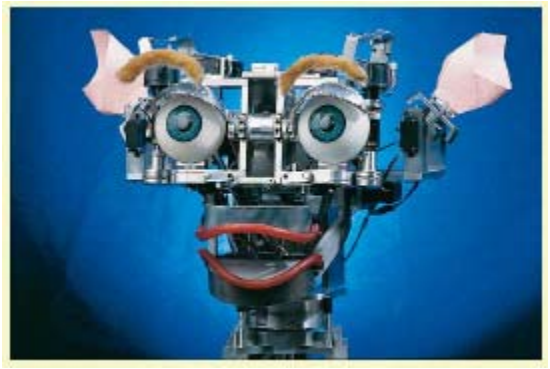
Bei der Frage nach dem Selbst eines Roboters ist die Autonomie der relevanteste Aspekt. Dies ergibt sich im Vergleich mit den anderen beiden Aspekten:

- Die Frage nach dem **Körper** des Roboters ist kein philosophisches, sondern primär ein technisches Problem. Ontologisch lässt sich kein Unterschied feststellen, will man sich nicht des Vitalismus schuldig machen.
- Die Frage nach dem künstlichen **Bewusstsein** eines Roboters ist derzeit rein spekulativ und führt in ein Problemfeld (Philosophie des Geistes), wo selbst Kernkonzepte (Bewusstsein) kontrovers diskutiert werden.
- Die Frage nach der **Autonomie** ist ein aktuelles wissenschaftlich-technisches wie philosophisch-ethisches Problem: Was sind sinnvolle Definitionen von Maschinenautonomie? In welchem Bezug stehen diese zu philosophischen Definitionen? Welchen Grad an Autonomie will man technischen Systemen generell zugestehen?

Körper: Was sind Roboter?



COG, MIT

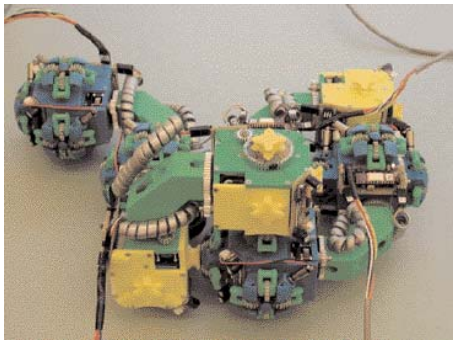


Kismet, MIT

Aibo,Sony

Khepera, EPFL

Ein Roboter ist eine Maschine, die über Sensoren Informationen aus der Umwelt aufnimmt sowie, basierend auf diesen Input, über Aktoren auf die Umwelt einwirkt



Modulroboter, Dartmouth/USA

Kennzeichen moderner Roboter

Folgendes wird in der modernen Robotik angestrebt:

- Direkte senso-motorische Kopplung: Kein Handeln aufgrund eines extern gegebenen „Weltmodells“, sondern im Sinn von „Reflexen“
- Integration autonomer Subsysteme: Keine zentrale Steuereinheit.
- Lernfähigkeit: Gewinnung neuer Kenntnisse durch Interaktion mit der Umwelt.
- Kommunikation: Auf (menschliche) Benutzer abgestimmtes Kommunikationsinterface sowie Kommunikation mit anderen Systemen.
- Selbstreparatur: Modularer Aufbau, Erkennung und Elimination von defekten Teilen.

Roboter: physikalische Modelle

Doktrin des „embodiment“: Das Körperhafte und die Situierung kognitiver Systeme ist zentral. Roboter werden dadurch zum Instrument der Erkenntnisgewinnung.

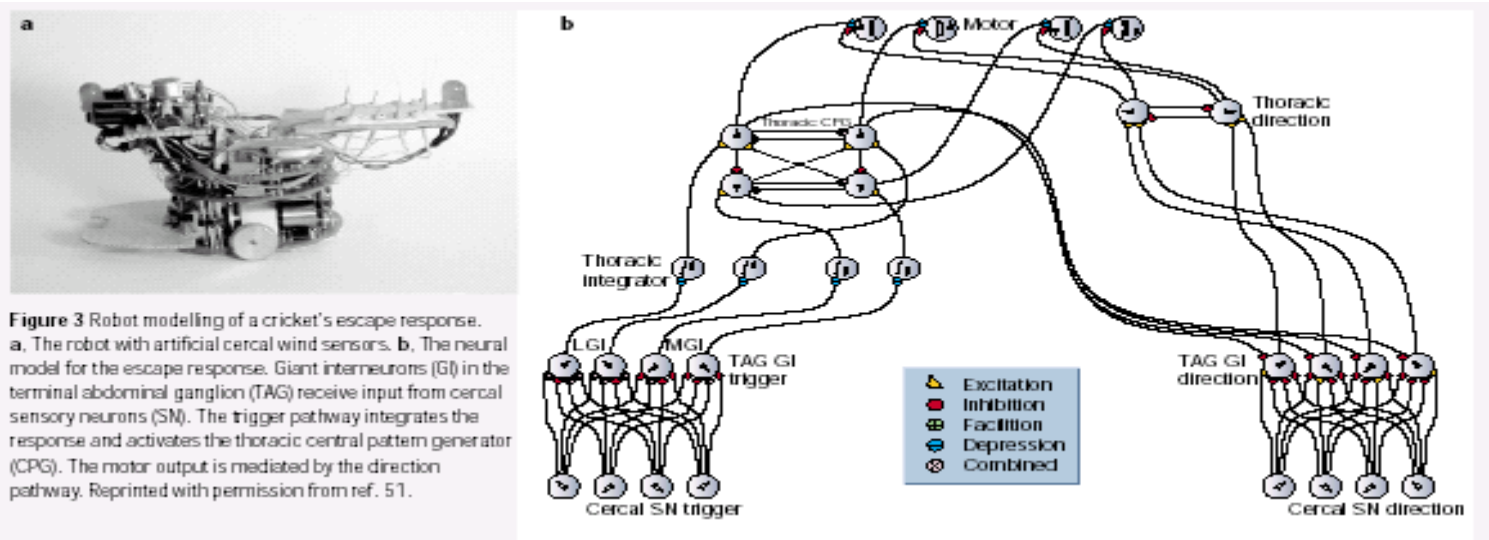


Figure 3 Robot modelling of a cricket's escape response. **a**, The robot with artificial cercal wind sensors. **b**, The neural model for the escape response. Giant interneurons (GI) in the terminal abdominal ganglion (TAG) receive input from cercal sensory neurons (SN). The trigger pathway integrates the response and activates the thoracic central pattern generator (CPG). The motor output is mediated by the direction pathway. Reprinted with permission from ref. 51.

Nature 417 (2002), pp 359-363

Roboter: Was ist (bald) möglich?

Strukturierung des Materials: Vorstoss auf die Ebene der Moleküle (von Mikrometern zu Nanometern)

Rechner: Transistordichte: 1 pro ❍m^2 ✸ 1 pro nm^2. Einsatz von organischen Molekülen für den Bau von informationsverarbeitenden Systemen.

Energie: Vom theoretischen Limit (Anzahl Additionen pro Joule) sind wir noch 12 Grössenordnungen entfernt.

Verstärkter Einbezug von neurowissenschaftlichen Kenntnissen (Sensorik, Informationsverarbeitung)

Roboter-Körper und Philosophie

Zwei Bemerkungen zu ontologischen und erkenntnis-theoretischen Aspekten von Roboter-Körpern:

1) Kein kategorialer Unterschied zwischen Roboter- und Menschen-Körper im Sinn der Ontologie.

2) Der wissenschaftliche Einsatz von Robotern folgt dem Erkenntnisideal von Giambattista Vico, wonach man versteht, was man bauen kann.

Bewusstsein: Philosophie

Es lassen sich eine Reihe verschiedener Phänomene unterscheiden, die Philosophen wie Kognitionswissenschaftler als kennzeichnend für Bewusstsein ansehen:

Aufmerksamkeit

Informationsverarbeitung

Sprache

Qualia-Bewusstsein

...

Als philosophische Herausforderung gilt Qualia-Bewusstsein. Was genau das ist, sowie ob und wie man solches feststellen kann, ist kontrovers. Die Frage nach dem Qualia-Bewusstsein bei Robotern ist demnach spekulativ.

Bewusstsein: Innerer Zustand

Eine Annäherung an das „Bewusstsein“ eines Roboters ist die Frage nach seinem „inneren Zustand“. Je nach Ansatz ergibt sich ein unterschiedliches Bild:

- Perspektive der „finite state machine“: Zustände und Regeln für das Wechseln der Zustände sind gegeben.
- Dynamisches-System-Perspektive: Keine Vorgabe von Zuständen, sondern Entwicklung dieser im Zeitverlauf (abhängig von Randbedingungen, z.B. neuronales Netz).

Inwiefern ein Roboter „Einsicht“ in diese Zustände hat, kann nur spekuliert werden bzw. ist eine Frage der Definition von „Einsicht“. Wichtig werden diese Ansätze bei der Frage nach der Maschinenautonomie.

Künstliches Bewusstsein

Vorgeschlagene Kriterien zur Ermittlung von „künstlichem Bewusstsein“:

- Die Komplexität des informationsverarbeitenden Systems übersteigt einen bestimmten Schwellenwert (von Neumann).
- Das System besteht den Turing-Test (Turing).
- Das System kann innere Zustände kommunizieren, welche die Erbauer sonst nicht erfahren können (Dennett).

Roboter-Bewusstsein und Philosophie

Die Frage nach dem Bewusstsein von Robotern ist jetzt und bis auf weiteres reine Spekulation, da man...

... derzeit keine allgemein akzeptierte Definition von Bewusstsein hat

... man keinen allgemein akzeptierten Test für die Ermittlung von Fremdbewusstsein hat (das bei Menschen gebrauchte Argument der Strukturanalogie funktioniert bei Robotern nicht).

Die Möglichkeit von „künstlichem Bewusstsein“ stellt zudem philosophisch-ethische Fragen: „künstliches Leid“ (Birnbacher), Frage nach „Maschinen-Rechten“ (Putnam), etc.

Autonomie: Philosophie

Der Autonomiebegriff hat in der Philosophie unterschiedliche Ausprägungen:

Politische Philosophie: Autonomie als politische Selbstständigkeit. In diesem Sinn wird der Begriff Autonomie bereits seit der griechischen Antike gebraucht.

Rechtsphilosophie: Die erste Ausprägung des Autonomiebegriffs in der Neuzeit findet sich im Recht. Hier geht es um das Verhältnis der persönlichen Selbstbestimmung im Rahmen einer rechtlich vorgegebenen Ordnung.

Immanuel Kant: In der Philosophie von Kant wird Autonomie zu einem zentralen philosophischen Begriff. Sie charakterisiert die Freiheit des Menschen als Vernunftwesens, gebunden durch den kategorischen Imperativ.

Autonomie und Maschinen (1)

Autonomie und Maschinen scheinen auf den ersten Blick unvereinbar, denn Maschinen...

... sind für einen bestimmten Zweck gebaut

... basieren auf einem Konstruktionsplan eines Erfinders

... und sollten (deterministisch) funktionieren

Insbesondere scheint undenkbar, dass Maschinen freie Handlungen im Sinn einer eigenständigen Zielsetzung realisieren können. Die Frage nach der Vereinbarkeit von Autonomie und Maschine kann als Variante der Determinismusdebatte aufgefasst werden.

Autonomie und Maschinen (2)

Dennoch gibt es Gründe, welche die Möglichkeit einer Maschinenautonomie nicht ausschliessen:

- Auch Menschen „funktionieren“ nur dann, wenn ihr physikalischer Aufbau gewissen „Konstruktionsgesetzen“ folgt.
- Maschinen können so konstruiert werden, dass sie „nichtdeterministisch“ funktionieren. Es gibt durchaus Gründe, Maschinen derart bauen zu wollen.
- Wenn es dereinst „Maschinenbewusstsein“ geben könnte, dann vielleicht auch „Maschinenfreiheit“.

Idee der Maschinenautonomie

Vom technologischen Standpunkt besteht der zentrale Aspekt der Maschinenautonomie darin, dass man einem Roboter nicht sagen sollte, *wie* er ein Problem lösen soll, sondern das sich aus den Randbedingungen des Problems ergeben soll.

Man will eine Autonomie bezüglich der Mittel und nicht bezüglich der Ziele.

Deshalb sollte ein autonomes System nicht von der Perspektive der „finite state machine“, sondern vom dynamischen-System-Ansatz aus beschrieben werden. Ein regelbasiertes System ist nicht autonom.

Konzepte von Autonomie

Maschinenautonomie – allgemeinste Definition:

Ein System agiert aufgrund der inneren Dynamik und der Interaktion mit der Umwelt.

Maschinenautonomie – spezifischere Definition:

- Das System entscheidet aufgrund sensorischem Input, Planen, Schlussfolgern und Abschätzen von Konsequenzen.
- Das System handelt aufgrund externer Zielvorgaben selbstständig durch die Kombination von Planungs- und Überwachungsschritten.
- Das System kann Lernen und Fehler beseitigen.
- Das System kann mit anderen autonomen Systemen kommunizieren.

Maschinenautonomie: Beispiel

S5, G. Miller, www.snakerobots.com

CONRO,university of southern california

Ein autonomer snakebot optimiert den Energieverbrauch und wählt die Gangart je nach Oberflächenbeschaffenheit (ermittelbar durch ein Sensor), ohne regelbasierte (= programmierte) Kontrolle. Das „wie“ der Bewegung bleibt der Maschine überlassen. Durch diesen Ansatz sind grundsätzlich Lösungen möglich, welche die Konstrukteure nicht vorausgesehen haben.

„Handlungen“ von Maschinen?

Philosophisch gesehen ist eine Handlung die **Umsetzung eines** gewollten (oder gesollten) **Zwecks** in die Realität. Das Ergebnis einer Handlung ist die **Tat**. Kriterium für die **Zurechenbarkeit** von Handlungen ist die **Bewusstheit** des Handlungsvollzugs und die **Freiwilligkeit** der Handlung (Platon).

Dieser Begriff von Handlung ist auf autonome Maschinen offenbar nicht zurechenbar, da man im heutigen Verständnis bei einer Maschine weder von Bewusstheit des Handlungsvollzugs noch von Freiwilligkeit der Handlung sprechen kann.

Dennoch kann ein autonomes System Dinge tun, welche von den Konstrukteuren nicht vorhergesehen sind. Dies ist bewusst gewollt, denn Maschinenautonomie ist ein Instrument, um neue Lösungen für ein Problem zu gewinnen.

Wer oder was handelt also?

Maschinenverantwortung? (1)

Bei (problematischen) Handlungen stellt sich die ethische Frage nach der Verantwortung. Der Verantwortungsbegriff ist ein sieben-stelliges Prädikat (Ropohl):

a *verantwortet* b *für* c *wegen* d
***vor* e *zum Zeitpunkt* f *im Sinn von* g**

a: Subjekt der Verantwortung
b: Objekt der Verantwortung
c: Ethisch relevanter Aspekt des Objekts der Verantwortung
d: Kriterium der Verantwortungsbewertung
e: Instanz der Verantwortungsbeurteilung
f: Zeitspanne der Zurechenbarkeit von Verantwortung
g: Psychologischer Aspekt der Betroffenheit von Verantwortung

Maschinenverantwortung? (2)

(Mindestens?) vier Variablen des Prädikats Verantwortung sind im Fall von „Handlungen“ autonomer Systeme problematisch:

a: Das Subjekt der Verantwortung ist offenbar nicht die Maschine allein. Ist es nur der Konstrukteur oder die Maschine und der Konstrukteur?

e: Instanz der Verantwortungsbeurteilung: Wenn Autonomie von der Instanz (z.B. Auftraggeber) gewollt wurde, inwiefern ist diese mitverantwortlich?

f: Die Zeitspanne der Zurechenbarkeit von Verantwortung ist insbesondere ein Problem bei lernenden Systemen.

g: Es ist fraglich, inwiefern Verantwortung beim Subjekt handlungsleitend sein kann, wenn das Systemverhalten inhärent unsicher ist.

Ein mögliches Kernproblem könnte die Erkennung innerer Zustände autonomer Systeme sein, welche zu potenziell ungewollten Aktionen des Systems führen.

Beispiel für Diskussion

A buyer accesses an autonomous computer controlled by a seller - a widget (Gerät) merchant - and asks the price of widgets. The buyer has never had any dealings with the seller or the seller's computer before. Having checked that there are widgets in stock, the computer uses knowledge that it has acquired itself to calculate a price by means of a complex formula that it has evolved for itself. The computer then notifies the buyer of the price at which it is prepared to sell the widgets. The buyer responds by ordering a quantity of widgets from the computer at the price quoted. The computer informs the buyer that it accepts his order and then causes the widgets to be despatched to the buyer and an appropriate debit to be made from his bank account. The seller never knows that this transaction has occurred. Does the transaction constitute a valid contract? If so, between whom?

(Allen/Widison: Can computers make contracts?)