

„Autonomie“ und die Autonomie „autonomer technischer Systeme“

Niels Gottschalk-Mazouz
Institut für Philosophie der Universität Stuttgart,
Seidenstr. 36, 70174 Stuttgart. E-Mail: gottschalk(at)philo.uni-stuttgart.de

Seit den 1990er Jahren sprechen Ingenieure mit Blick auf Roboter, Softwareagenten u.ä. von „autonomen technischen Systemen“. Was heißt dabei „autonom“, und inwieweit ist unsere Autonomie davon betroffen, dass jetzt auch die Technik „autonom“ wird? Die Beantwortung dieser Fragen erfolgt anhand von (I.) der Darlegung des üblichen Autonomieverständnisses, (II.) der Analyse der Autonomieverständnisse, die in der Diskussion um autonome technische Systeme zum Ausdruck kommen, und (III.) Überlegungen dazu, inwieweit unsere Autonomie von autonomen technischen Systemen berührt ist und wie solche Systeme dementsprechend eingerichtet sein sollten.

I.

Historisch gesehen entstammt der Autonomiebegriff der politischen Sphäre (Pohlmann 1971). Die Stadtstaaten im alten Griechenland formulierten damit einen Anspruch auf Selbstbestimmung, gegen Bedrohungen von innen (Tyrannis) und von außen (Fremdherrschaft). Diese Selbstbestimmung war stets eine – u.a. durch übergeordnete Bündnisregeln – begrenzte Selbstbestimmung. Dieses Verständnis hat sich m.E. bis heute erhalten, so etwa in der Kennzeichnung als „autonome Region“, und auch in der Philosophie trifft man ein solches Verständnis an, z.B. in der Diskussion um Freiheit und Determinismus, wenn „Autonomie“ dort diejenige Schwundstufe von Freiheit bezeichnet, die unter den angeblichen Determiniertheiten für uns nur zu haben sein soll (so z.B. bei Walter 1998). Ganz wie im Politischen handelt es sich um eine durch äußere und innere Determinanten begrenzte, aber dadurch teilweise auch erst ermöglichte Selbstbestimmung.

Die in der modernen Philosophie wirkmächtigste Begriffsbestimmung von Autonomie geht auf Immanuel Kant zurück. Autonomie wird nun vom Individuum aus gedacht, wird transzendental verstanden, als Bedingung der Möglichkeit verantwortlichen, und das

heißt: moralischen Handelns überhaupt. Sie ist absolut, unbeschränkt, nicht graduierbar und normativ zu verstehen, drückt also eine bestimmte Form der Anerkennung seiner selbst bzw. eines Gegenübers aus, und das Zukommen von Würde hängt an der Zuschreibung als autonom.

Bei Kant wird Autonomie an den Gesetzesbegriff gebunden. Moralische Qualität bemisst sich ganz allgemein bei Kant an bloßer Form, eben der Gesetzesform, der Maximen: „Autonomie des Willens ist die Beschaffenheit des Willens, dadurch derselbe ihm selbst (unabhängig von aller Beschaffenheit der Gegenstände dieses Wollens) ein Gesetz ist.“ So lautet der bekannte erste Satz des Abschnitts mit dem Titel „Die Autonomie des Willens als oberstes Prinzip der Sittlichkeit“ aus der *Grundlegung zur Metaphysik der Sitten* (vgl. (vgl. GMS BA 87). Daran wurde kritisiert, dies sei ein aufwendiger Apparat und wenig phänomenadäquat, eine rigide technizistische Vorstellung von Soll- und Istwert und von Kontrolle des eigenen Verhaltens per Regel (als Regelung wie im technischen Sinne). Andere moderne Auffassungen nehmen vom Gesetzesbegriff Abstand (Frankfurt 1971; vgl. zusammenfassend auch Buss 2002). Allen diesen Auffassungen sind jedoch so weit ich sehe zwei normative Aspekte von Selbstbestimmung gemeinsam: Einerseits eine begriffliche Bestimmung, d.h. eine Identifikation und Schematisierung der vorgefundenen, eigenen Aktivität (man weiß, was man tut). Andererseits, in distanzierender Reflexion des begrifflich Identifizierten, eine praktische Bestimmung (man verbindet mit seinem Tun eine Werthaftigkeits- oder Richtigkeitsüberzeugung). Material gesehen geht es dabei darum, Unverfügbarkeiten und Verfügbarkeiten situationsadäquat richtig ins Verhältnis zu setzen. Eine vernünftige Handlung, als Ausdruck von Autonomie, ist jedenfalls nicht bloß eine Reaktion auf Vorgefundenes (innen/außen), auch aus Beobachtersicht nicht, sondern erfordert die aktive Ausbildung eines Verhältnisses zu diesem Vorgefundenen.

II.

In der Diskussion um autonome technische Systeme benennt Autonomie teils ebenfalls die Fähigkeit, sich selbst Regeln zu setzen (vgl. Smithers 1997). Doch bleibt diese Rede

völlig metaphorisch, wenn sie nicht *technisch* genauer spezifiziert wird. Bei genauerer Betrachtung der verschiedenen Definitionen der Autonomie technischer Systeme geht es meist um weniger Anspruchsvolles, und zwar um eine Reihe von selbst auferlegten „engineering constraints“ (vgl. Maes 1991):

1. Eine Unabhängigkeit von der Steckdose, allgemein gesagt die Fähigkeit, ohne externe Energie- oder Materialversorgung auszukommen. Das könnte man auch "autark" nennen (z.B. energetisch autark).
2. Die Möglichkeit, sich ohne Führung selbst zu bewegen, d.h. mobil zu sein.
3. Eine Auftrags erledigung ohne Benutzereingriffe, ohne Fernsteuerung: Dafür haben wir m.E. das Wort "automatisch".

Teils werden die Kriterien auch kombiniert, etwa wenn es heißt: "autonomous robots are usually taken to mean free-ranging mobile robots which are not teleoperated but plan and execute their own actions." (Todd 1986, S. 233)

In der Robotik-Szene ist hinsichtlich des dritten Punktes inzwischen die Unterscheidung von ferngesteuerten, halbautonomen und autonomen Systemen üblich. Halbautonome Systeme können bestimmte komplexe Aktionen selbst ausführen, die Steuerung vollzieht sich also auf einer höheren Ebene. Eigentlich bezeichnet diese Einteilung Grade auf einem Kontinuum der Steuerung einzelner Aktoren (wie bei der klassischen Funkfernsteuerung eines Modellautos) über komplexere Befehlsstrukturen (vor, vor, vor, drehen, Hand öffnen, Arm absenken, Hand schließen, arm heben, zurück, zurück, zurück), noch komplexere (heranfahen, aufnehmen, zurückfahren) bzw. schlicht (holen).

Autonomie wird aber auch durch Bezugnahme auf die Systemumwelt beschrieben. Hier fordert man unter dem Begriff der Autonomie:

4. Eine Abhängigkeit des zukünftigen Verhaltens nur von inneren Zuständen des Systems (vgl. McFarland/Bosser 1993, S. 145-147); dieses soll also, würde ich sagen, "umweltunabhängig" sein.
5. Eine Auftrags erledigung in verschiedenen Situationen bzw. eine Erledigung von variierenden Aufträgen; das System soll "adaptiv" sein – und, würde ich sagen, damit

gerade *nicht* umweltunabhängig.

6. Das System soll all dies immer besser können; es soll also "lernend" sein.

Vor allem die Punkte 2 bis 5 berücksichtigt Rammert (2003) in seinem synoptischen Vorschlag, die Autonomie autonomer technischer Systeme als eine Frage des Grades an Eigenaktivität in vier Dimensionen zu charakterisieren:

- Motorik (Antrieb und Bewegung): von unbewegt zu automotiv und automobil
- Aktorik (Arbeit und Ausführung): von fremdbetätigt bis zu eigentätig, automatisch
- Sensorik (Umwelt- und Selbstwahrnehmung): von fremdabgestimmt über umweltsensitiv bis zu selbstanpassend
- Informatik (Steuerung und Regelung): von fest verdrahtet über flexibel programmiert bis zu verteilt problemlösend.

Die nächsten beiden Verwendungen sind etwas komplizierter, weswegen ich ausgiebiger zitieren werde:

7. Autonomie bedeutet, „dass das System in Abhängigkeit von sich ständig wandelnden Umweltsituationen nicht nur – losgelöst von menschlichen Kommandos – zielgerichtet arbeitet, sondern dass ein externer Beobachter Verhaltensweisen des Systems erkennen kann, die nicht explizit bei der Konstruktion des Systems vorgegeben wurden“ (Knoll/Christaller 2000). Das System soll also nicht nur explizit konstruktiv vergebene Verhaltensweisen zeigen; es soll also "innovativ" sein.

8. „Je mehr der kontrollierende Akteur A Kenntnis des inneren Zustands von B hat, und je besser er die Gesetze kennt, mit denen sich der Zustand von B beeinflussen lässt, desto besser kann A B kontrollieren. (...) B's Autonomie gegenüber A ist – qualitativ gesprochen – invers proportional zum Wissen, das A über B hat.“ (Pfeifer 2003). Ein System ist also autonom in dem Maße, wie dessen innere Zustände und Gesetze ihrer Veränderung uns nicht bekannt sind; es wäre dann "opak", "nicht vorhersagbar".

Geht man die einzelnen Definitionsvorschläge durch, lässt sich also zeigen, dass „autonom“ hier im Wesentlichen soviel bedeutet wie „(energetisch) autark“, „mobil“, „automatisch“, „umweltunabhängig“, „adaptiv“, „lernend“, „innovativ“ oder „unvorhersagbar“. Die letzten beiden Bestimmungen sind nun jedoch ausdrücklich aus

Sicht eines Beobachters vorgenommen, d.h. epistemische. Die letzte Bestimmung ist von ihrem Autor ausdrücklich als relationale (relativ zum Beobachter, vgl. für diese Idee bereits Covrigaru/Lindsay 1991) eingeführt, bezeichnet also keine einfache Eigenschaft. Die Grammatik technischer Autonomie folgt, wie man es ausdrücken könnte, der Formel: x ist autonom gemessen am Wissen von y . Diese Beobachtung lässt sich jedoch verallgemeinern, denn bei näherer Betrachtung sind die anderen Kennzeichnungen ebenfalls relative: „Energetisch autark“ und „umweltunabhängig“ z.B. kennzeichnen eine negative Relation zur Umgebung; „automatisch“ kennzeichnet eine negative Relation zum Benutzer; „adaptiv“ eine positive zur Umgebung und eine negative zum Benutzer; „lernend“ schließlich eine positive zur Umgebung und eine negative zum Konstrukteur oder Benutzer.

Auffällig ist schließlich, dass die spezifizierte Hinsicht der Autonomie immer nur unter normalerweise nicht ausgesprochenen Randbedingungen besteht. Energetische Autonomie besteht z.B. nur unter bestimmten Umgebungsbedingungen wie hinreichender Sonneneinstrahlung (wenn Solarzellen verwendet werden).

Allgemein gesagt hat die Zuschreibung technischer Autonomie daher folgende Form: x ist autonom in (positiv oder negativ bestimmten) Hinsichten a unter (normalerweise nicht ausgesprochenen) Randbedingungen b .

Eine solche Zuschreibung setzt das System in u.U. mehrfacher Hinsicht in Relation, sie bestimmt nämlich (i) Austauschrelationen mit der System-Umwelt; (ii) Interaktionsrelationen mit Benutzer und Konstrukteur (Steuerung bzw. Programmierung) - sowie evtl. noch Dritten (Überwachung); (iii) epistemische Relationen bezüglich des Wissens von Benutzer/Konstrukteur/Dritten. Dabei bestimmen die epistemischen Relationen mit über unsere Interaktionen und diese wiederum mit über die Austauschverhältnisse des Systems.

Allgemein kann man das Ergebnis also so zusammenfassen: Autonomie spezifiziert Austausch-, Interaktions- und Wissensverhältnisse von x auf eine oder mehrere Hinsichten a hin, unter Unterstellung von Randbedingungen b .

III.

Auf dieser Basis können nun unsere Autonomie und diejenige technischer Systeme aufeinander bezogen werden: Es besteht eine Asymmetrie insofern, als unsere Autonomie eine moralisch relevante reflektierte Intentionalität *von* Personen ausdrückt, technische Autonomie hingegen eine Relationalität zur Umgebung und *zu* Personen. (Dennoch kann es sinnvoll sein, technischen Systemen Intentionen zuzuschreiben, dazu komme ich gleich noch.) Weiterhin besteht eine Spiegelbildlichkeit der Verfügbarkeiten, denn wir verfügen über das System anhand des für das System Unverfügbaren, und worüber das System verfügt, verfügen wir nicht (mehr). Diese Verfügbarkeiten/Unverfügbarkeiten sind *von uns* gesetzt, in den Dimensionen von Austausch, Interaktion, Wissen. *Unsere* (individuelle wie kollektive) Autonomie ist es, die dabei befördert, wiederhergestellt oder zumindest erhalten werden soll – aber auch unsere Wohlfahrt usw. Das "Unsere" ist dabei natürlich erläuterungsbedürftig, dahinter verbergen sich komplexe Verhältnisse, genauer gesagt: Verhältnisse je individueller und kollektiver Autonomie, die fallbezogen genauer ausgeführt werden müssen. Dies wird dadurch zusätzlich kompliziert, dass sich die Interaktion mit autonomen Systemen ganz unterschiedlich „anfühlt“, d.h. lebensweltlich ganz unterschiedlich erlebt werden kann. Ich möchte das am Beispiel eines (in dieser Art fiktiven) Fahrerassistenzsystems in Automobilen erläutern.

Die Phänomenologie des Handelns unter Einbezug autonomer Systeme umfasst folgende Möglichkeiten: 1. Das System kann mir unbemerkte Unterstützung bieten (Feindosierung von Lenkbewegungen in Extremsituationen); ich erlebe mich dann als gesteigert handlungsfähiges Subjekt. 2. Man kann sich, möglicherweise gleichzeitig, ganz schlicht als Subjekt dem Automobil als Objekt gegenüberstellen. 3. Man kann aber auch das autonome System im Automobil als einen virtuellen Beifahrer erleben, der (wie man selbst) einem Objekt gegenübersteht (dem „restlichen“ Automobil). Wenn autonome Systeme untereinander kommunizieren können (und beim Automobil ist das aus verschiedenen Gründen angestrebt), treten weitere komplexe Verhältnisse auf: 4. Ich erlebe, wie zwei Objekte aufeinander einwirken. 5. Ich erlebe, als Subjekt, wie sich zwei virtuelle Subjekte miteinander austauschen (etwa den günstigsten Abstand einer Kolonnenfahrt miteinander aushandeln). 6. Mein virtueller Beifahrer informiert mich

über ein anderes, als Objekt vorgestelltes Automobil. 7. Der virtuelle Beifahrer eines anderen Automobils spricht zu mir. 8. Mein virtueller Beifahrer informiert mich über einen Output des virtuellen Beifahrers eines anderen Automobils. Wenn sich schließlich ein ein anderes Auto fahrender Mensch an mich wendet bzw. ich eine Interaktion mit ihm verstehen will, und autonome Systeme im Spiel sind, wird es zusätzlich kompliziert, weil bei zwei Subjekten bezüglich der Möglichkeiten 1.-8. keine symmetrische Situation bestehen muss. Bei mehr als zwei Automobilen/Fahrern wird das schnell noch unübersichtlicher (Kolonnen sollen jedoch gewöhnlich mehr als zwei Fahrzeuge umfassen können sollen).

Dass virtuelle Subjekte im Spiel sind, geht u.a. mit dem Einsatz von intentionalem Vokabular einher. Dieses dient vor allem dazu, sich zu sich selbst und anderen gegenüber verhalten zu können und zu interagieren, wenn Determinationen nicht bekannt sind. Ich glaube, es gibt hier eine intrinsische Verbindung zur letzten der oben dargestellten Autonomiedefinitionen. Bei autonomen technischen Systemen können wir nun nicht tatsächliche Intentionalität ansetzen, da es sich stets um von außen zugeschriebene Zwecke handelt. Als solche sind diese Zwecke niemals ethisch relevant (ganz anders als im Kantischen Autonomiebegriff, der gerade eine Achtung der jeweils zugeschriebenen Zwecksetzungskompetenz ausdrückt). Wir sollten also eine Als-Ob-Intentionalität ansetzen. Autonome technische Systeme verarbeiten möglicherweise intentionales Vokabular, auch wenn sie selbst es nicht verstehen, und wir beziehen uns unter einem solchen Vokabular auf sie. Auch hier macht es gelingende Interaktionen möglich, setzt dabei aber bloße Regelmäßigkeit voraus, die implizit bleiben kann.

Wenn wir diskutieren wollen, wie die Autonomie der Technik sich zu unserer eigenen Autonomie verhält, sollte diese Doppelung stets mitgedacht werden. Wir selbst können unsere eigene Autonomie einerseits deskriptiv verstehen, dann diskutieren wir Kriterien für und Konsequenzen von empirischer Autonomie. Dann fungiert sie als Beschreibung, und ist als solche problemlos auch auf autonome technische Systeme anwendbar (wenn diese Systeme komplex genug sind, was sie derzeit häufig noch nicht sind). Andererseits können wir sie transzendental verstehen, als Bedingung der Möglichkeit von Moralität, Verantwortung usw. Dann fungiert sie als Zuschreibung, und ist als solche nicht

problemlos auf technische Systeme anwendbar. Dies ist keine Frage des besseren Gerätedesigns, besserer Software, mehr Arbeitsspeicher oder Rechenleistung. Denn es ist überhaupt keine Frage der empirischen Performanz solcher Systeme. *Wir* sind es, die diese Zuschreibungen vornehmen oder nicht zunehmen, und die dadurch unsere Anerkennung von Gegenübern als moralischen Personen ausdrücken. Natürlich ist das nicht losgelöst von der Frage empirischer Autonomie: Solange autonome technische Systeme so viele Dinge nicht können, die wir von verantwortungsfähigen Personen fordern, wird man ihnen, selbst wenn man es wollte, keine Autonomie zuschreiben können.

Ich glaube, dass diese Konstellation eine Reihe von Herausforderungen mit sich bringt, die ich hier nur kurz benennen will:

Was wir von uns selbst in dieser Hinsicht fordern, lässt sich beispielsweise bioethischen Arbeiten zu Einwilligungsfähigkeit, Behinderungsgraden und anderen Themen entnehmen, insofern sie mit Blick auf oder unter dem Titel der Autonomie geführt werden (vgl. Miller 1995). Ich denke, dass sich hiervon ausgehend unter Hinzunahme von Arbeiten aus der Psychologie und der Kognitionswissenschaft ein graduelles Konzept empirischer Autonomie entwickeln ließe, mit dem auch der jeweilige Grad der Autonomie autonomer technische Systeme beurteilt werden könnte. Anhand einer morphologischen Klassifikation von Handlungskomponenten (vgl. Collins/Kusch 1998) ließen sich zusätzlich diejenigen Komponenten genauer identifizieren, die sich technischen Systemen vergleichsweise einfach übertragen bzw. diesen beibringen lassen. Damit ließen sich diese Systeme auf das hin optimieren, was sie wirklich gut können, bzw. ließen sich Interaktionsverhältnisse gestalten, in denen diese unterstützend genau diejenigen Abläufe übernehmen, die sie genauso gut oder besser erledigen können wie ein Mensch.

Eine besondere Herausforderung, das würde ich mit Blick auf technische Systeme ganz wie Dieter Sturma sehen (vgl. die von ihm geprägten Passagen in Christaller et al., S. 111-134), liegt dabei in der Ankoppelung an den „Raum der Gründe“. Dabei sollte von vornherein klar sein, dass mit technischen Systemen bestenfalls Als-ob-Diskursivität möglich ist: Wir können so tun, als würden wir uns unterhalten, ohne jedoch das

technische Gegenüber wirklich als Autor seiner Sprechakte ansehen zu können. Technisch realisierten Frage-und-Antwortmaschinen können wir vielleicht irgendwann in ihren Fragen und Antworten nicht mehr von einem menschlichen Gegenüber unterscheiden, bis es so weit ist, ist es aber noch ein langer Weg. Insbesondere das Verstehen (im Sinne des adäquat auf sie Reagierens) expressiver und normativer Aspekte von Sprechakten, anders als das von Beschreibungen und Befehlen, ist unterentwickelt. Hier wäre vielleicht auch das menschliche Sprechen noch stärker nach charakteristischen Mustern zu untersuchen.

Das Verstehen und die Modellierung von Interaktionssituationen wie der oben beschriebenen, in der menschliches und nichtmenschliches Tun aufeinandertrifft, kann wie oben angedeutet durchaus gewinnbringend in intentionalen Kategorien beschrieben werden. Die moralische Qualität des Zugeschriebenen allerdings bleibt prinzipiell beschränkt: So kann man einem technischen System nicht Unrecht tun dadurch, dass man es falsch versteht – man kann nur sich selbst oder anderen Personen schaden dadurch, dass man das tut. Die Notwendigkeit, potenziell authentische von rein fremdzugeschriebenen Intentionen zu unterscheiden, macht jedoch eine Art Buchführung über den Autor von Intentionalitätszuschreibungen erforderlich. Intentionen- oder Präferenzzuschreibungen tragen also notwendig eine Art versteckten Index, der auf die Person des Zuschreibenden verweist. Diesen könnte man in agenten- und spieltheoretischen Modellierungen explizit mitführen und so u.a. die Rückwirkungen dieser Zuschreibungen auf das Handeln der menschlichen Agenten erfassen. Oder man könnte zumindest ein Flag setzen, das anzeigt, ob eine Intention oder Präferenz als moralisch relevant gilt, d.h. als solche beispielweise in Nutzenaggregationen eingehen soll. Eventuell würde es sich in manchen Kontexten auch anbieten, das übliche Desire-Belief-Intention-Modell aufzugeben zugunsten eines reinen Belief-Modells, jedenfalls aber die Zuschreibungskomponente explizit zu verbuchen, wie es in einem nicht-technischen Zusammenhang bereits als prinzipielles Kommunikations- und Interaktionsmodell vorgeschlagen wurde (Brandom 1994). Damit ließe sich auch die Ankoppelung an den „Raum der Gründe“ modellieren.

IV.

Die Frage, inwiefern Roboter moralische Subjekte sind bzw. inwiefern ihren Handlungen moralische Qualität zukommt, wird seit einiger Zeit nicht mehr nur in Literatur und Film, sondern auch in der Philosophie diskutiert (siehe z.B. IRIE 2006). Doch autonome technische Systeme sind nicht mit Robotern gleichzusetzen. Alle technischen Systeme, die über Sensoren und Aktoren verfügen, d.h. etwas in der wirklichen Welt bewirken können – und unter dem Stichwort des *ubiquitous computing* oder der *ambient intelligence* wird deren Entwicklung derzeit forciert –, bringen Probleme der Überwachung und der Kontrolle mit sich (Gottschalk-Mazouz 2007). Mit Blick auf Roboter wurde in Christaller et al. (2001) die Frage bereits interdisziplinär diskutiert, was genau nun deren mögliche Autonomie für Maßnahmen erforderlich machen könnte. Die Systeme sollen, so wurde überlegt, möglichst deutlich werden lassen, was sie im Begriff sind zu tun (etwa, indem sie das aktuelle Vorhaben laufend vor sich hersprechen), sollen deutlich werden lassen, dass sie in Benutzungssituationen lernen (um dem Benutzer seine Mitverantwortung klarzumachen), sollen das auch intern protokollieren, sollen teils nur unter ausdrücklicher Zustimmung der Betroffenen eingesetzt werden und sollen als autonom handeln gekennzeichnet werden. Man könnte hinzufügen, dass sie auch Auskunft darüber geben können sollten, wessen Auftrag sie ausführen, wenn sie einen ausführen. Allgemein kann man sagen, dass diese Systeme erkennen lassen sollten, wie sie den Spielraum, den sie haben, nutzen, und (höherstufig), dass ihnen ein solcher Spielraum offensteht, und sie sollten diesen Spielraum zur Disposition stellen können. Dies ließe sich mit Blick auf die oben unterschiedenen acht Lesarten von Autonomie weiter ausführen, worauf hier aus Platzgründen verzichtet wird.

Informationsethische Fragen im engeren Sinne beziehen sich auf die Frage, wie mit den vielen Daten, die ein solches System erfasst, umgegangen werden soll. Dazu müssen die Datenströme zunächst einmal bekannt sein. Hat das System Down- und Uploadfunktionen? Vermutlich wird es solche haben. Aktive Input-Kanäle könnten dann als solche gekennzeichnet sein, Einwilligungen zu Übertragungen könnten situationsbezogen eingeholt werden. Doch die Situation ist noch komplexer als in den herkömmlichen Diskussionen um Datenschutz und Überwachung: Denn vieles der

Informationen, auch des Gelernten, liegt nicht explizit vor. Wenn mit dem Gerät sinnvoll interagiert werden können soll, muss es in hohem Maße implizites Wissen aufnehmen und erwerben können. Dieses äußert sich in einem Können des Systems. Da dieses aber nicht schon automatisch propositional erfasst ist, lässt es sich zwar, in Form von Trainingsdaten oder Algorithmen, technisch u.U. gut übermitteln, aber nicht mehr inhaltlich darlegen oder gar beurteilen. Der Benutzer kann hier dann nur noch die Einwilligung dazu geben, dass eine Speicherung oder Übertragung stattfindet, aber nicht mehr, wozu, d.h. zu dem, was der Inhalt dieser Übertragung ist. Gleichzeitig scheint der Trend hin zu intensiverer, mehr und mehr auch dezentraler Vernetzung ungebrochen (vgl. Gottschalk-Mazouz 2008). Ich glaube, dass genau in dieser Konstellation, der zunehmenden Vernetzung und der zunehmenden Fokussierung auf implizitem Wissen, eine besondere informationsethische Herausforderung in der Gestaltung von und im Umgang mit autonomen technischen Systemen liegt.

Literatur:

- Brandom, R. (1994): *Making it explicit. Reasoning, Representing, and Discursive Commitment*. Cambridge, MA
- Buss, S. (2002): "Personal Autonomy", *The Stanford Encyclopedia of Philosophy (Winter 2002 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2002/entries/personal-autonomy/>.
- Christaller, T.; Decker, M.; Gilsbach, J.-M.; Hirzinger, G.; Lauterbach, K.; Schweighofer, E.; Schweitzer, G.; Sturma, D. (2001): *Robotik. Perspektiven für menschliches Handeln in der zukünftigen Gesellschaft*. Heidelberg
- Collins, H. M.; Kusch, M. (1998): *The Shape of Actions. What Humans and Machines Can Do*. Cambridge, MA.
- Covrigaru, A. A.; Lindsay, R. K. (1991): "Deterministic autonomous systems". In: *Artificial Intelligence Magazine*, S. 110-117.
- Frankfurt, H. (1971): "Freedom of the Will and the Concept of a Person". In: *The Journal of Philosophy* 68 (1), S. 5-20.
- Gottschalk-Mazouz, N. (2007): „Überwachung und Macht & die Spezifik technisierter Überwachung“, in: Gaycken, S.; Kurz, C. (Hg.): *1984.exe - Gesellschaftliche, politische und juristische Aspekte moderner Überwachungstechnologien*, Bielefeld, S. 161-183.
- Gottschalk-Mazouz, N. (2008): „Internet and the flow of knowledge: which ethical challenges will we face?“ in: Hrachovec, H.; Pichler, A. (Hg.): *Philosophy of the*

Information Society. Proceedings of the 30. International Ludwig Wittgenstein Symposium Kirchberg am Wechsel, Austria 2007. Volume 2, Frankfurt, Paris, Lancaster, New Brunswick: Ontos, S. 215-232.

IRIE (Hg.) (2006): *International Review of Information Ethics*, Bd. 6 (Ethics in Robotics), URL = < <http://www.i-r-i-e.net/issue6.htm>>

Kant, I. (1785/1968): *Grundlegung zur Metaphysik der Sitten*, zit nach der Weischedel-Werkausgabe, Bd. 7, Frankfurt 1968.

Knoll, A.; Christaller, T. (2000): „Selbstrepräsentation, Selbstwahrnehmung und Verhaltenssteuerung von Robotern“. In: Sandkühler, H.-J. (Hg.): *Selbstrepräsentation in Natur und Kultur*, Frankfurt, S. 109-132
(<http://www.iais.fraunhofer.de/fileadmin/images/pics/Abteilungen/IL/TC/Publikationen/Christaller2000.1.ps.gz>).

Maes, P. (Hg.) (1991): *Designing Autonomous Agents*. Cambridge, MA.

McFarland, D./Bossler, T. (1993): *Intelligent behaviour in animals and robots*. Cambridge, MA.

Miller, B. (1995): "Autonomy", in Reich, W. T. (Hg.): *Encyclopedia of Bioethics*, Revised Edition. New York, S. 215-220

Pfeifer, R. (2003): „Körper, Intelligenz, Autonomie“. In: Christaller, T.; Wehner, J. (Hg.): *Autonome Maschinen. Maschinen werden selbständig - was kommt auf uns zu?* Wiesbaden: Westdeutscher Verlag, S. 137-159
([http://www.ifi.uzh.ch/ailab/people/gomez/PfeiferPublications/Koerper-Intelligenz-Autonomie\(GermanPaperFinalOne\).pdf](http://www.ifi.uzh.ch/ailab/people/gomez/PfeiferPublications/Koerper-Intelligenz-Autonomie(GermanPaperFinalOne).pdf)).

Pohlmann, R. (1971): "Autonomie". In: Ritter, J. (Hg): *Historisches Wörterbuch der Philosophie*, Bd. 1, S. 702-719.

Rammert, W. (2003): *Technik in Aktion. Verteiltes Handeln in soziotechnischen Konstellationen*. Technical University Technology Studies Working Papers TUTS-WP-2-2003, Berlin.

Smithers, T. (1997): "Autonomy in Robots and Other Agents", In: *Brain and Cognition* 34, S. 88-106.

Todd, D. J. (1986): *Fundamentals of robot technology*. London.

Walter, H. (1998): *Neurophilosophie der Willensfreiheit. Von libertarischen Illusionen zum Konzept natürlicher Autonomie*. Paderborn 1998.