

Visualität als Lebens- und Erkenntnisform des Menschen

11. Vorlesung: (Do 18.10.)

Maschinen ohne Leben und Intelligenz: Warum Computer keine Personen sind und wie das mit dem Mangel an Selbstkontrolle und visuellen Fähigkeiten zusammenhängt

In den letzten Vorlesungsstunden hatte ich A. Noes etwas engeres Verständnis des Visuellen gewarnt, das Visualität und die Bedeutung der visuellen Welt allein von den Handlungs- und d.h. hier Bewegungsabläufen aus verstehen will, die sich den wahrnehmenden Tieren und Menschen anbieten. Selbst wenn wir Noes These als grundlegend akzeptieren und ihm darin zustimmen, dass es ein praktisches, körperlich angeeignetes Handlungswissen gibt, wie eine sichtbare Gestalt z.B. eines Würfels im Gegensatz zu einer Kugel ganz bestimmte, charakteristische Bewegungen von uns fordert, so ist dies doch keine Erklärung für die anderen Bedeutungsebenen, die auch alltägliche visuelle Erfahrungen für uns haben können, gegeben. Wir können manchmal schon von weiten *sehen*, dass ein Mensch tot ist - um John Bergers Beispiel auszunehmen - weil wir die Starre eines Körpers direkt visuell erfassen. Noch andere, nämlich existentiell, lebensentscheidende und ethische Bedeutungen vermitteln uns Bilder.

Unser Verständnis gerade dieser existentiellen Bedeutung von Bildern hat hängt aber von unserem Selbstverständnis ab und der Rolle, die wir der Reflexion und Selbstkontrolle zu billigen. Mit dieser These und Ankündigung

schloß die letzten Vorlesungsstunde. Wir wollen also heute durch den Kontrast zwischen maschinellen Fähigkeiten von PCs die Fähigkeiten, besser verständlich machen, die den Menschen auszeichnen.

Computer – nicht Tiere, aber auch nicht Engel oder Gott – sollen uns also als das Gegenmodell dienen, mit dem wir die geistigen Fähigkeiten des Menschen zum reflexiven Denken kontrastieren wollen. Computer sind, da sie zur abstraktiv basierten Selbstkontrolle nicht und zur logischen Selbstkritik nur beschränkt fähig sind, nicht intelligent. Sie sind deshalb ebensowenig „dumm“ wie sie „intelligent“ genannt werden können: Da sie nichts anderes können als, auf wie komplizierte Weise auch immer rechnen, sind sie mit der reflexiven Intelligenz des Menschen nicht vergleichbar. Doch reproduziere ich da nur ein Vorurteil? Nein, denn ich habe einen guten Grund für diese Behauptung: Nur der menschliche Geistes ist durch unbegrenzt selbstreflexive Prozesse so kontrollierbar, dass auch noch die *moralische Wertungen*, die in die Selbstkontrolle eingehen, kontrolliert werden können. Das bedeutet aber: Er unterscheidet sich aufgrund seiner reflexiven Fähigkeiten – logische Selbstkritik, Selbstbestimmung und moralische Selbstkontrolle – dadurch vom programmgesteuerten Rechnen der Computer, dass er prinzipiell nicht auf ein Abarbeiten feststehender Programme eingeschränkt sein kann. Vielmehr kann er selbstkontrolliert und abstraktiv auf seine Umgebung antworten, und diese enaktive Fähigkeit zur kontrollierten Selbst- und Fremdbestimmung ist nicht fixierbar. Ich werde deshalb, die kritische Flexibilität und Autonomie des Geistes gegenüber Programmen und äußeren Vorgaben nicht nur als einen wichtigen

Unterschied, sondern als eine Grenze für das verstehen, was die künstliche Intelligenz von Computern im Gegensatz zum menschlichen Geist zu leisten vermag, um die visuelle Umgebung bewertend zu erfassen und in die Gestaltung menschlichen Lebens einzubeziehen.

Früher gab es einmal Ansätze, die für die Erforschung der künstlichen Intelligenz einen sehr hohen Anspruch erhoben. Diese sog. starke These der AI behauptet, dass ausnahmslos alle kognitiven Prozesse des menschlichen Geistes durch Algorithmen nicht nur darstellbar, sondern im Kern auf sie reduzierbar sind. Danach wäre der menschliche Geist prinzipiell, nämlich vom Typ her, mit den Rechenprozessen eines Computers identisch. Zum einen sind Algorithmen ein mathematisches Darstellungsmittel, das jeden Prozess beschreibt, der gemäß eines formal entscheidbaren Verfahrens verlässlich zu einem Ergebnis führt. Alles, was im Ablauf von Prozessen an allgemeinen, funktionalen Bedingungen für einen formal entscheidbaren Ablauf erfüllt wird, kann also auch formal durch einen Algorithmus beschreibbar sein – oder aber auch nicht. Dabei ist völlig gleichgültig, ob ein Prozess im Geist, im Rechner, im Reagenzglas, in einem Bakterium oder etwa in einer Fabrik abläuft. Ich denke, dass nichts dagegen spricht, dass fast jeder geistige Prozess unter dem Aspekt seiner algorithmischen Entscheidbarkeit beschrieben werden kann oder eine Ebene des material-neurophysiologischen Ablaufs aufweist, die algorithmisch darstellbar ist. Wir können also der starken AI zugestehen: Es gibt eine universale Eigenschaft der Berechenbarkeit, auch wenn damit nicht alle und insbesondere nicht alle zeit- und entwicklungsabhängigen Aspekte geistiger Prozesse fassbar sind.

Damit ist aber nicht das gewonnen, was die traditionellen AI-Theoretiker wünschen, die die starke These vertreten: Nämlich Reduzierbarkeit. Denn damit ist noch nicht gesagt, dass alle kognitiven Leistungen z.B. der kritisch bewertenden und strukturierenden Denkprozesse, vollständig algorithmisch fassbar sind. Vielmehr werde ich zeigen, dass das Gegenteil richtig ist: Die Eigenschaft der Berechenbarkeit macht nur Sinn, weil sie auf dem Hintergrund nicht berechenbarer selbstkontrollierter Leistungen geistiger Prozesse steht. Doch was bedeutet und welche Konsequenzen hat die Existenz nicht-algorithmischer Eigenschaften kognitiver Prozesse in reflexiver Position?

Um diese Frage zu beantworten, ist eine praktisch-ethische Betrachtung und Kontextualisierung geistiger Prozesse erforderlich. Der früher in Augsburg lehrende Philosoph Klaus Mainzer hat in einem Vortrag zum Thema „Künstliche Intelligenz - eine Problemskizze“ gesagt, dass es von unserer Entscheidung abhängt, „was wir an künstlicher Intelligenz neben uns brauchen und dulden wollen“. Damit sind die handlungs- und Umgebungsbeziehungen geistiger Prozesse thematisiert. Wer entscheidet, was wir brauchen und wollen? Es kann sich nicht allein um individuelle Entscheidungen handeln. Es geht um gesellschaftliche Entscheidungen und Entwicklungen und ihre ethische, an unseren geteilten Zwecken und Idealen ausgerichtete Orientierung. Ihr Ergebnis wird die meisten Mitglieder unserer Gesellschaft und ihre Lebensweise prägen und mitbestimmen.

Die folgenden zwei Punkte sollen Mainzers These untermauern, weiterführen und einige Probleme mit Computer und AI in den Vordergrund

rücken:

(1.) Computer sind keineswegs - obwohl dies häufig behauptet wird - universelle Maschinen. Es wäre sogar in einer Hinsicht irreführend zu sagen, dass sie Annäherungen an derartige universelle Maschinen sind. Es gibt nämlich einen prinzipiellen Grund, warum Computer niemals universelle Maschinen sein *dürfen*. Damit Computer und die in ihnen laufenden Programme in der Welt verwendbar werden und für uns nützlich sein können, ist es nötig, dass sie die zufälligen empirischen Bedingungen und Umstände einbeziehen, die in unserer Welt jeweils lokal bestehen. Doch kein Computerprogramm kann alle zufälligen faktischen Bedingungen und Umstände seiner Anwendung enthalten und allgemein vorwegnehmen. Deshalb ist der nicht-algorithmisierbare Eingriff des aktiv eingreifenden Benutzers eine unverzichtbare Bedingung der Anwendbarkeit jedes Programms. Computer werden deshalb niemals ohne ständige aktive Betreuung auskommen, die eingreift, um ihr sachgerechtes Funktionieren für uns in menschendienlicher Weise zu gewährleisten. Wir, die Verwender der Computer, liefern diese notwendige aktive Betreuung und gehören zu dem ökologischen System Mensch-Maschine notwendig dazu. Unser Handeln im Sorgeverhältnis um den Computer muß dessen Operationsbedingungen immer wieder und immer wieder erst herstellen, damit wir ihn überhaupt sinnvoll verwenden können. Dies geschieht nicht nur dadurch, dass wir die Programmierung des Computers verändern. Vielmehr sind zielgerichtete und lokale Eingriffe im Ablauf nötig, um den erfolgreichen Einsatz von Programmen zu ermöglichen. Anders als Tier, Pflanze und Mensch können

Computer keine sich selbst anwendenden und für sich selbst sorgenden Systeme sein.

(2.) Damit hängt das Fehlen einer weiteren reflexiven Fähigkeit von PCs zusammen: Anders als Tiere, Pflanzen und Menschen sind Computer keine sich selbst reproduzierenden Systeme. Menschen sind nicht nur biologisch selbsterhaltende, sondern sich sozial selbstreproduzierende Wesen. Das Menschen nur als soziale Wesen selbsterhaltend sind bedeutet: Ohne Menschen im Plural gibt es den Menschen in der Einzahl als Person und Individuum nicht. Die biologische Tatsache, dass es mindestens einen Mann und eine Frau braucht, um einen neuen Menschen zu zeugen, weist in die richtige Richtung. Doch zu Menschen im vollen kulturellen und sozialen Sinne werden wir nur durch das Miteinander, den wechselseitigen Austausch in einer Gemeinschaft mit anderen Menschen, werden. Die Fähigkeit zur sozialen Selbsterhaltung und -reproduktion hat ihre Entsprechung im visuellen Miteinander des offenen Austauschs von Angesicht zu Angesicht. Nur dadurch können Menschen ein Selbst werden, zu individuellen Personen werden, die zur Selbstkritik und Selbstkontrolle befähigt sind. Personen sind selbsthaft gewordene Menschen. D.h. sie haben als Individuen einen geistig-sozialen, autonomen Status für sich selbst erobert und gewonnen. Sie sind fähig, in kulturellen und gesellschaftlichen Prozessen und Entwicklungen ihren Weg zu finden und prägend Individuelles für alle anderen, die ihre Zeitgenossen sind, beizutragen. Aber nur weil sie einander als zwecksetzende Subjekte anerkennen, konnten sie auch ihre individuelle Persönlichkeit ausbilden. Doch Computer sind weder selbsterhaltend noch

selbstreproduktiv, und sie können auch nicht zu selbstbestimmten Personen werden. Das bedeutet aber: Computer besitzen keinen Geist, der dadurch bestimmt ist, dass sie in Beziehungen eintreten müssen, um „ihr Selbst als Person“ erwerben zu können.

Doch warum sollten wir es wollen, warum sollten wir es brauchen, dass nun auch Computer dank AI ein Selbst entwickeln? Ich denke, dass Klaus Mainzer recht hat: Wir sollten in eine Debatte eintreten, welche Art von künstlicher Intelligenz wir haben wollen. Es spricht viel dafür, dass Computer keinen Geist haben und keine Personen werden sollten. Wir sollten auch nicht anstreben, dass Computer die vollständige Selbstkontrolle möglich ist. Denn würde dies geschehen, so wären sie nicht mehr als Computer von uns nutzbar: Wir müßten sie als Personen und damit als moralische Subjekte anerkennen, denen wir dann auch die selben Rechte zubilligen, die wir für uns selbst als Menschen in Anspruch nehmen.

Doch damit sind wir schon wieder, fast zwangsläufig, von der logischen Selbstkritik zum Thema der umfassenderen, auch moralischen Selbstkontrolle übergegangen. Doch genau dazu sollte uns der konkrete Kontrast zwischen menschlicher und maschineller Intelligenz dienen. Gleichwohl war auch dies ein Vorgriff, und wir werden erst in der nächsten Vorlesung sowohl die Konzeption der Selbstkontrolle und -kritik in ihrer Beziehung zur distanzierend, abstraktiven Leistung des Ausbildens von Zwecken, Zielen und Werten beschreiben. Dann wird es uns auch möglich sein, zusätzlich zum menschlichen Selbstverständnis, die Rolle von Weltbildern zu berücksichtigen. Sie sind es nämlich, die es uns abstraktiv und

erklärend erlauben, die Beziehung zwischen Lebensform und Visualität genauer zu fassen.

Wir wollen uns in dieser Vorlesung vielmehr von einer ganz anderen Seite her mit der Problematik des Lebensformbegriffs beschäftigen, indem wir den in der letzten Vorlesung bereits beschriebenen grundlegenden Unterschied zwischen Computer und Menschen näher untersuchen. In der letzten Vorlesung hatte ich ja die Unterschiede zwischen den Rechenoperationen von Maschinen und menschlicher Intelligenz herausgearbeitet. Ja, ich hatte bestritten, dass die mehr als metaphorische Bezeichnung von Computern als intelligent oder dumm Sinn macht. Meine Begründung argumentierte mit den Graden der logischen Selbstkritik und moralischen Selbstkontrolle, die Menschen ausüben können, die jedoch für PCs prinzipiell nicht zugänglich ist. Diese Behauptung bedeutet jedoch nicht, dass es nicht möglich wäre, Computer und Menschen miteinander zu vergleichen. Denn natürlich können auch Menschen kognitive Leistungen erbringen, indem sie Algorithmen ausführen. Computer und Menschen sind also dann auf dieselbe Weise bewertbar, wenn es um Leistungen geht, die wie z.B. im Rechnen oder Schachspielen ausführbar sind. Ob Mensch oder Computer: Geht es z.B. um das Ableiten logisch notwendiger Folgerungen, so kommt es allein auf die formallogische Beziehung zwischen Prämissen und Konklusion an. Sie ist entweder wahrheitserhaltend oder nicht und deshalb berechenbar. In diese Richtung denkt schon 1902 der amerikanische Philosoph und Logiker C. S. Peirce, wenn er behauptet, dass

„...ein Mensch als eine Maschine betrachtet werden kann, die, sagen

wir, einen geschriebenen Satz produziert, der eine Folgerung ausdrückt, wobei der Mensch-Maschine als Prämisse eine Tatsachenbeschreibung gegeben worden war. Da diese Leistung nicht mehr verlangt als eine Maschine ausführen könnte, weist sie keine wesentliche Beziehung zu dem Umstand auf, dass die Maschine mit gesteuerten Zahnrädern arbeitet, während der Mensch sich einer unzureichend verstandenen Anordnung von Gehirnzellen bedient; ... logische Kritik ... ist gleichermaßen auf die Leistung der Maschine und die des Menschen anwendbar.“ (CP 2.59,¹ 1902)

An dieser Überlegung ist für die Fragestellung unserer Vorlesung der folgende Aspekt wichtig: Wenn wir den Unterschied in der Rolle und der Reichweite verstehen wollen, den intelligente geistige Prozesse für Menschen und für Computer spielen, so ist die Entscheidung für ein enges Verständnis, das nur berechenbare geistige Leistung berücksichtigt, gleichbedeutend damit, dass wir den Unterschied zwischen Computer und Mensch ignorieren. Es reicht nicht aus, dass wir die Leistung allein der wahrheitsfunktionalen bzw. der effektiv berechenbaren mathematischen Funktionen durch diese beiden Arten der Intelligenz betrachten. Vielmehr ist eine Betrachtungsperspektive angemessen, die die Rechenleistung logischer Intelligenz im Zusammenhang ihrer Verwendungsbedingungen berücksichtigt.

Damit sind nun nicht alle logisch notwendigen, berechenbaren

1

Das Sigel „CP“ bezieht sich auf die „Collected Papers of Charles Sanders Peirce“, Bde 1-6 hg. von Charles Hartshorne und Paul Weiss, Cambridge, Mass. 1933-35; Bde. 7 & 8, hg. von Arthur Burks, Cambridge, Mass. 1958. Die CP werden folgendermaßen zitiert: in CP 2.59 bezeichnet 2 die Nummer des Bandes, 59 die Nummer des Abschnitts.

Folgerungen bedeutungslos geworden. Irrelevant sind nur logische Beziehungen zwischen einer kleinen Zahl bereits festliegender Prämissen. Denn alle realistischen Umstände der Anwendung logischer Beziehungen setzen stets die Selektion von Prämissen aus einer großen Zahl von Sachverhalten voraus. Auch bei notwendigen, deduktiven Schlüssen lassen sich aus manchen Zusammenstellungen von Prämissen, deren Zahl größer als zwei ist, mehrere, manchmal unbegrenzt viele Konklusionen ableiten.² Schon daraus, dass unter einer großen Zahl von Prämissen ausgewählt werden muss, ergibt sich eine Möglichkeit, einen Unterschied zwischen künstlicher und natürlicher Intelligenz zu beschreiben.

Denn wie entscheiden wir, welche Prämissen wichtig und relevant sind? Nun können wir uns rational, aus guten Gründen, für die Relevanz bestimmter Themen oder Zwecke und gegen andere entscheiden, wenn wir Sachverhalte als aussagekräftige und fruchtbare Prämissen relativ zu einem Zweck identifizieren und andere nicht. Doch woher kommen die Zwecke? Es ist von Hume und etlichen anderen Philosophen behauptet worden, dass die Auswahl von Zwecken nicht rational erfolgen kann. Ich denke, dass die Humesche Tradition irrt: Was Menschen zu Personen macht, was ihre Klugheit und ihren Charakter ausmacht, ist ihre Fähigkeit, selbstkritisch, rational und konsistent (d.h. für diese Person integrierend) über die Zwecke und die Maßstäbe von Entscheidungen urteilen zu können. In dieser Vorlesung werde ich zeigen, dass es eine konstitutive Beziehung zwischen

2

„... all deductive reasoning, except that kind which is so childishly simple that acute minds have doubted whether there was any reasoning there -- I mean non-relative syllogism -- requires an act of choice; because from a given premiss, several conclusions -- in some cases an infinite number -- can be drawn.” (CP 6.595, 1893)

zweckbestimmten Entscheidungen in intelligent-selbstkritischen geistigen Prozessen und dem Abwägen von konkurrierenden Zwecken gibt. Die Bedeutung dieser Beziehung erklärt dann den Unterschied, der zu jeder künstlichen Intelligenz besteht. Meine Ausgangspunkte bilden zwei Einwände, die Peirce gegen die Möglichkeit von allgemeinen Denkmaschinen anführt. Sie knüpfen beide an den folgenden Befund an: Es ist nicht klar, was es heißt, dass eine künstliche Intelligenz mit „Originalität und Initiative“ ihren Einsatz und ihre Daten nach Relevanzgesichtspunkten selbst festlegen könnte. Denn:

„Ein Akt der ersten und zufälligen Bestimmung wäre erforderlich; und es scheint fast evident, dass keine Maschine einen derartigen Akt ausführen kann, außer in engen Grenzen, die zuvor entworfen wurden und in ihre Konstruktion eingebaut ist. ... Außerdem ist ein besonderer Akt üblicherweise erforderlich, der angemessen als Abstraktion bezeichnet werden kann, der darin besteht, dass die flüchtigen Elemente des Denkens erfaßt und als „substantielle“ Objekte festgehalten werden, um eine Formulierung von William James zu verwenden.“ (6.595, 1893) Zur Auswahl

von relevanten Daten müsste bei einer KI die selbstkontrollierte Einführung von Abstraktionen hinzutreten. Erst auf dieser Basis können effektiv berechenbare logische Beziehungen, können Algorithmen in einem stärkeren Sinne, „intelligent“ angewendet werden. Wie das Ausbilden von Abstraktionen mit der menschlichen Fähigkeit verbunden ist, rational und selbstkontrolliert für bestimmte Zweck- und Relevanzordnungen so zu gestalten, dass dadurch letztlich die Lebensform auf immer wieder neue

Weise gestaltet. Doch wie das geschieht, wie Lebensform und Selbstkontrolle ineinander greifen wird noch genauer zu klären sein. Versuchen wir dazu zunächst eine erste These zu formulieren. Sie soll die Verschiedenheit und das einander ergänzende von moralischer Praxis der Lebensform und der eher theoretischen abstraktiven und selbstreflexiven Leistung der kognitiven Fähigkeiten zu verbinden.

Eine These über die Differenz

Wir hatten Selbstkritik und kontrollierte Abstraktion als jene kognitiven Fähigkeiten eingeführt, durch die sich künstliche von menschlicher Intelligenz unterscheidet. Wie noch genauer zu zeigen sein wird, ist es Menschen durch ihre Abstraktionsfähigkeit möglich, sich von den Wahrnehmungen und damit von den engen sensomotorischen Wissen zu lösen. In einer ersten Hypothese, Differenzthese genannt, behaupte ich folgendes:

1. Differenzthese: Weil Menschen fähig sind, durch kontrollierte Abstraktionen zu Zwecken und Werten ihrer logischen Selbstkritik und Selbstkontrolle zu machen und ein Selbstverständnis aufzubauen, ist es ihnen möglich, sich zu Personen zu entwickeln, die auch reflexiv ihre Lebensform im Austausch miteinander und mit ihrer visuellen Umgebung auszugestalten. Es ist diese Fähigkeit zu einer reflexiven, selbstkontrollierten Ausgestaltung ihrer Lebensform, die menschliche von maschineller Intelligenz unterscheidet. Welche Gründe sprechen für diese Behauptung, die ich Differenzthese genannt habe? Für ihre Begründung sind mehrere Argumentationsschritte erforderlich, die ich in Betracht der fortgeschrittenen

Zeit nur zum Teil Ihnen vorführen kann. Doch verschaffen wir uns eine Übersicht: Im ersten Schritt ist der Begriff der menschlichen Person näher zu bestimmen. Zweitens wäre genauer zu zeigen, wie Selbstkontrolle und Abstraktion aus den Wahrnehmungskontexten in der biographischen Entwicklung nicht nur einer Person, sondern menschlichen Lebensform wirksam werden können. Drittens wäre im Zusammenhang damit zu klären, wie Abstraktionen den Prozeß der Zweck- und Relevanzbildung visuell und lebenswirklichkeitsbezogen begründen können. Schließlich müsste es viertens darum gehen, noch einmal konkret zu verdeutlichen, warum menschliche Personen, jedoch nicht Computer in der Lage sind, selbstkritisch Abstraktionen für ihre Entwicklung einzusetzen, um sich zu Personen zu entwickeln.

Das Selbstverständnis menschlicher Personen: Charakter und Lebensform

Es ist Teil der menschlichen Selbstverständigung, die zum Person-Sein oder besser Person-Werden gehört, Abstraktionen auszubilden, durch die Personen sich selbst begrifflich näher bestimmen. Dass menschliche Personen weder Engel, Tiere, weder Götter noch Computer sind, sind solche Selbstvergewisserung, die durch Abgrenzung erfolgen. Denn zu verstehen, welche Gründe hier für Unterschiede sprechen, ist das Herstellen genau jenes menschlichen Selbstverständnisses.

Einer der zentralen Begriffe menschlicher Selbstverständigung ist der Begriff der Person selbst. Doch welchen Begriff der Person erhalten wir, wenn wir Personen als Menschen verstehen, die sich durch selbstkontrollierte

Abstraktionen entwickeln? Der Bremer Philosoph Georg Mohr hat eine Liste von 18 begrifflichen Merkmalen aufgeführt, die den Begriff der Person spezifizieren sollen. Sie reichen von den Eigenschaften der Individualität und Leiblichkeit bis zur Würde und interpersonalen Anerkennung durch andere Personen.³ Im Zentrum des Personenbegriffes steht die Fähigkeit von Menschen, sich autonom zu entwickeln, Zwecke selbst zu bestimmen und dadurch den eigenen Charakter selbstkontrolliert auszubilden. Andere, ebenfalls von Mohr gelistete Eigenschaften von Personen wie Autonomie in der Bestimmung des eigenen Handelns, Freiheit in der Wahl von Zwecken und Normen, Intentionalität als die Fähigkeit, zu absichtlichen und zweckvollen Handelns, ein Vermögen zu praktischen Urteilen gehen in die komplexe Fähigkeit zur selbstkontrollierten Entwicklung des Charakters durch die Ausbildung von Zwecken als Abstraktionen konstitutiv ein.

Für das, was Personen zu sich individuell entwickelnden Personen macht, ist ihre Fähigkeit entscheidend, Zwecke und Ziele nicht nur zielorientiert zu verfolgen. Denn das können auch Computer, Tiere und Thermostaten. Wesentlich wichtiger ist vielmehr, dass Personen fähig sind, Zwecke unter dem Einfluß von Erfahrungen und Bewertungen auszubilden,

3

In: „Was ist eine Person? Begriffsgeschichte und aktuelle philosophische Diskussionen“, in: *Was ist eine Person?*, hg. von G. Mohr, Schriftenreihe des Zentrums Philosophische Grundlagen der Wissenschaften, Band 25, S. 9 - 29, Bremen 2002. Mohr listet die folgenden Eigenschaften: 1. Individualität, 2. Leiblichkeit, 3. Sinnlichkeit und Bewusstsein, 4. Perspektivität, 5. Selbstbewußtsein und synchrones Identitätsbewußtsein, 6. diachrones Identitätsbewußtsein, 7. Zukunftsbewußtsein, Selbstsorge, 8. Intentionalität, 9. Theoretisch-instrumentelle Rationalität, 10. Selbstverhältnis, 11. Sympathie für andere Lebewesen, 12. narrative Identität, 13. Autonomie, Willensfreiheit, 14. Selbstbestimmung (Entscheidung über das eigene Handeln) 15. praktisch-legitimatorische Rationalität, 16. Vernunft, 17. interpersonale Anerkennung 18. Dignität (moralische Würde aufgrund von Rechten und Pflichten).

zu verändern und so weiter zu entwickeln. Erst dadurch kann auch ein die Person formendes Selbstverständnis entstehen. Aus der Differenzthese folgt: Eine Person ist zur Bildung und Entwicklung von Zwecken in der Lage, weil sie relativ zu ihrer Erfahrungs- und menschlichen Interaktionsumgebung durch Reflexion und Kritik ihre Zwecke bewertet, verwirft oder entwickelt. Dies ist z.B. dadurch möglich, dass sie durch Akte der Selbstkritik und Selbstkontrolle ihr Handeln, Denken und Empfinden durch ihre eigenen Entscheidungen entweder über die Kontrolle von Bedingungen mittelbar beeinflusst oder durch ihre Entschlüsse unmittelbar festlegen kann. Nach diesem Bild werden Menschen insbesondere dadurch zu Personen, dass sie nicht nur einzelne Zwecke ausbilden und entwickeln. Was den Charakter eines Menschen zu seiner Persönlichkeit werden lässt, ist die Fähigkeit, lebenslang bis zu irgendeinem Grade eine teilweise konsistente Ordnung der Präferenz und Komplementarität zwischen den verfolgten Zwecken und den ihnen entsprechenden Handlungen, unwillkürlichen Reaktionen, herzustellen. Personen in diesem starken Sinne sind wir nicht immer und überall und in jeder Hinsicht. Aber wir sind es in einem erstaunlichen Maße, und wir sind es dann nur, insoweit wir fähig sind, Ziele, Zwecke, Impulse und Handlungen einen Zusammenhang zu geben, der sie übergreift und relational vereinheitlicht.

Das Resultat dieses Bildungsprozesses, die Persönlichkeit eines Menschen, wird manchmal irreführend die Identität einer Person genannt - als handelnde es sich um etwas irgendwann völlig Feststehendes. Besser scheint mir, vom Charakter einer Person zu sprechen. Alle einzelnen Handlungen von Cäsar

beispielsweise, kann man extern - physikalisch und physiologisch - so beschreiben wie die Handlungen jedes anderen Menschen: Schlafen, essen, lachen, sprechen, gehen, kämpfen usw. Aber Cäsar hat diese Handlungen auf eine Weise ausgeführt und miteinander verknüpft, dass sie eine Verbindung erhalten haben, die der keines anderen Menschen gleicht. In dieser Ordnung von Cäsars Handlungen zeigt sich Cäsars Charakter, der Ausdruck der von ihm verfolgten Präferenzordnung seiner Ziele und Zwecke ist. Um den Zusammenhang zwischen Charakter und Erfahrungsbewältigung in einer Biographie zu klären, benötigen wir eine genauere, 2. Hypothese über die Rolle von Abstraktionen:

Es gibt lebensformative Abstraktionen: Die begriffslogische Grundlage für die Bildung, Entwicklung und mitmenschliche Vermittlung von Zielen, Inhalten, Zwecken und Werten ist die kognitive Fähigkeit zur erfahrungsbezogenen Abstraktion von Inhalten aus der Wahrnehmung, die selbstkritisch und selbstkontrolliert aufgenommen und verarbeitet werden.

Nach unserer Hypothese spielen Selbstkontrolle und Abstraktion in dem Bildungsprozeß, der die Entwicklung eines Menschen ausmacht, also eine zentrale Rolle, die wir wie folgt beschreiben können: Menschen können nur zu Personen in dem starken Sinne werden, der einen Charakter ausprägt, wenn sie in der Lage sind, in ihrem Verhalten und Denken einen nicht begrenzbaren (aber deshalb nicht unbegrenzten) Grad der Selbstkontrolle zu erreichen. Wir wollen zwischen Selbstkritik erster Stufe und den offenen Stufen reflektierter logischer Selbstkritik, die dann zur Selbstkontrolle moralischer Art werden kann, unterscheiden. Natürlich können auch Computerprogramme so geschrieben werden, dass eine Maschine das

Ergebnis der Ausführung eines ihrer Programms anhand von den Kontrollparametern bewertet, die den Erfolg der Anwendung des Programms kontrollieren und dann extern etwa durch die Abwägung mit anderen Zielen bewerten. Ein noch einfacherer Fall der Selbstkontrolle erster Stufe ist z.B. ein Thermostat als Element eines Systems, das durch Rückkoppelung kontrolliert, ob ein bestimmter Soll-Zustand erreicht wurde. Natürlich sind heutige Computer in der Lage, mehrere aufeinander bezogene Programme zu verarbeiten, die so aufeinander bezogen sind, dass das eine Programm durch ein zweites oder drittes so überprüft wird, dass das zweite oder dritte Programm auch die Kontrollparameter verändert, die zur Überprüfung des Erfolgs des ersten Programms dienen. Aber natürlich ist auch diese Veränderung der Kontrollparameter im Programm z.B. Rückkoppelungsschleifen und „wenn – dann“ Verknüpfungen in seinem Parameterbereich festgelegt.

Anders dagegen die menschliche Intelligenz. Für das Ausbilden von höheren Stufen reflektierter Selbstkontrolle, die an Zielen, Idealen und Zwecken orientiert sein kann, ist ein selektives Bilden von Abstraktionen erforderlich, die dann den Kontrollparameter nicht nur verändern, sondern den gesamten Bereich und die Relevanzordnung verändern. Die Art der hierfür erforderlichen Abstraktionen, die neue Werte, Zwecke und Ziele schaffen, werde ich „hypostatische Abstraktion“ nennen. Dieser Begriff wurde von dem amerikanischen Philosophen C. S. Peirce eingeführt. Um etwas zu einem Zweck für eine Person zu machen, ist es nötig, dass ein Erfahrungsgehalt z.B. aus einer vollzogenen Wahrnehmung, die uns mit

diesem qualitativen oder relationalen Gehalt vertraut machte, herausgelöst und zum Gegenstand des Bewertens, Planens und des ablehnenden oder suchenden Verhaltens gemacht wird. Erst dieser Status als Denkgegenstand ermöglicht es, dass etwas auch zu einem Zweck und schließlich in Objekte des Wollens, Begehrens und Bezweckens umgesetzt werden kann. Nehmen wir ein Beispiel: Im abstrahierenden Übergang von der Beobachtung

„Die Rose ist rot“

zu der Suche nach „etwas Rotem“ wird der Begriff „Röte“ zum eigenständigen, substantivierten Objekt kognitiver Haltungen und Entscheidungen gemacht: In einem reflexiven Akt wird ein fiktiv-abstraktes Objekt dadurch gebildet, dass das Denken einen bereits erfaßten Inhalt des Ahnens, Wahrnehmens oder Denkens noch einmal, nun aber als separates Objekt und an Subjektstelle von Aussagen über es, aufmerksam darstellt. Diese gedanklich reflexive und gleichzeitig vergegenständlichende Wirkung des Akts versetzt unser Denken in die Lage, sich von beliebigen Anlässen zu emanzipieren - eben z.B. der Wahrnehmung jener einzelnen roten Rose. Der abstrahierte Begriff der „Röte“, der dieselbe Eigenschaft als Allgemeinbegriff fasst, ist nun als Denkobjekt ein Gegenstand, dessen Beziehungen zu wirklichen Gegenständen weiter bedacht, geprüft, interpretiert und eben auch als Zweck angestrebt werden kann.

Großraumbegriffe wie „Tugend, Gerechtigkeit, Wahrheit, Schönheit“ sind übrigens sekundäre, durch mehrfache Abstraktion vermittelte Begriffe und keine primär abstrahierten Denkobjekte. Primär wird dagegen z.B. im Übergang von „Dieser Honig ist süß“ zu „Diesem Honig kommt Süße zu“

eine solche abstraktive Bildung eines Denkobjekts vollzogen.

Was ist damit für unser Verständnis der Fähigkeit von Menschen gewonnen, Zwecke auszubilden und so die höherstufige Eigenschaft des persönlichen Charakters zu entwickeln? Ein Mensch kann sich nur dann für das Ziel entscheiden, stets rote oder süße Gegenstände zu bevorzugen, sie z.B. zu kaufen oder herzustellen, weil er oder sie fähig war, den Begriff der Röte und des Süßen unabhängig von einzelnen derartigen Wahrnehmungen als mögliche Zielzustände zu abstraktiv denken und mit den konkreten Umständen möglichen Handelns erfolgreich zu verknüpfen. Es ist eben diese Reflexivität der Substantivierung zum Denkobjekt, die eine autonome und interne Gegenstandsbildung innerhalb des Feldes praktischer und kognitiver Einstellungen erlaubt.⁴

Man kann diesen Befund auch als Beschreibung der inneren Denksemantik einer Person beschreiben. Dann lautet er: Erst der abstrahierte Begriff konstituiert ein fiktiv-abstraktes Objekt, das ein vorgängiges Urteil erneut auf allgemeine Weise faßt. Ein Zweck kann eine Person erst dann ausbilden, wenn sie einen Teil eines vorausgegangenen Gedankens, die gedachte Eigenschaft, als eigenständiges Objekt des Denkens ausgewählt hat. Für das Problem der Bildung einer Person ist dabei wichtig: Diese Objektbildung beruht auf einer echten Selbstreferentialität des Denkens.

4

In einer Reihe von Aufsätzen, insbesondere aber in „Das Denken und die Struktur der Welt“, hat der amerikanische Philosoph H.N. Castañeda eine Ontologie der Konstitution aller Gegenstände durch die Schichtung minimaler Denkobjekte (thin objects) vorgeschlagen. Dieser Ansatz und diese Arbeit wird dargestellt und von 16 deutschen Philosophen kommentiert in dem Band „Das Denken und die Struktur der Welt - H.N. Castañedas epistemische Ontologie in Darstellung und Kritik“, Berlin 1990.

Folglich erlaubt erst die hypostatische Abstraktion die Emanzipation des Denkens von seinen Umständen und die Aufprägung individueller selektiver Haltungen - später auch Entscheidungen - auf den weiteren Umgang mit Erfahrung. Ausgehend von Wahrnehmungs- Erinnerungs oder auch Fantasiegehalten wird so der mögliche Entwurf von späteren Objekten des Wollens vorgezeichnet. Trotzdem verweist das z.B. aus einer Wahrnehmung abstrahierte Denkobjekt auf die Gelegenheit seiner Einführung zurück: Sie kann deshalb an neuen Wahrnehmungen überprüft und mit ihnen verknüpft werden. Die obige Differenzthese verweist auf eben diese grundlegende abstrahierende Operation, durch die Personen die Entwicklung ihrer Zwecke selbst bestimmen und kontrollieren können.

Doch wie hängt die Fähigkeit zur Zweckbildung mit einer selbstkontrollierten Persönlichkeits- oder Charakterentwicklung zusammen? Und was trägt dies zum Unterschied zwischen künstlicher und natürlicher Intelligenz bei? Wir wollen diesen Fragen nachgehen, in dem wir die Fähigkeit zur Selbstkritik und Selbstkontrolle am Beispiel einfacher logischer Maschinen diskutieren. Die Einfachheit des Aufbaus dieser Maschine hat den Vorteil, dass sie Operationen durchführen, die durchschaubar sind, weil sie einfache aussagenlogische Operationen (genauer: einige aristotelische Syllogismen) vollziehen. Doch entwertet die Einfachheit nicht jedes Ergebnis zu dem wir gelangen können? Besteht die Komplexität und Leistungsfähigkeit heutiger Computer nicht gerade darin, dass sie jene zusätzliche Grade der Selbstkontrolle bereitstellt, die Computer mit dem menschlichen Geister vergleichbar macht? Nun, ich glaube das nicht. Denn

wie soeben dargestellt, ist für den Geist menschlicher Personen die flexible, d.h. im Grad nicht begrenzbar und auf kontrolliert ausgewählten Abstraktionen aufbauende Selbstkontrolle, konstitutiv und charakteristisch. Es mag sein, dass auch solche Abläufe nachträglich durch ein Computerprogramm nachvollziehbar sind. Aber ein solches Nachvollziehen verfehlt den Punkt, um den es geht: Die Abfolge der Stufen der Selbstkontrolle und der Wahl der Abstraktionen ist offen und wird flexibel organisiert, so dass es auf das Verhältnis der Abläufe zum jeweils nächsten Stufe der Selbstkontrolle ankommt.

Aber selbst wenn das Modell der einfachen logischen Maschinen unangemessen sein sollte, so ist damit niemanden geschadet: Es müßte doch für denjenigen, der die Fähigkeit von modernen Computern zur Selbstkontrolle beweisen will, dann um so leichter, an diesem einfachen Modell nachzuweisen, durch welche Weiterungen moderne, komplexere Maschinen in der Lage sind, selektive Abstraktionen mit kontrollierter Selbstkontrolle zu verbinden.

Logische Selbstkontrolle und die Grenzen logischer Maschinen

Ich beginne mit einem kleinen Addendum zur Geschichte der technischen Realisierung maschineller Intelligenz. Die Geschichte beginnt an der Johns Hopkins University gegen Ende des 19. Jahrhunderts. In den Jahren 1883 und 84 hatte der Philosoph und Logikdozent C.S. Peirce dort einen begabten Studenten. Er hieß Allan Marquand und baute 1886 eine logische Maschine. Diese Maschine bestand aus Rädern und Walzen, die vier

Buchstaben mechanisch gegeneinander verschieben konnte. Peirce schreibt über diese Maschine: „Tatsächlich ist alles, was letztlich dargestellt wird, eine Liste der möglichen Produkte von vier Buchstaben. ... Es gibt praktisch keine Tasten außer den acht für die Buchstaben und ihre Negationen, denn die zwei Tasten, die für den Prozeß des Löschens verwendet werden, sollten hier nicht zählen“, bemerkt Peirce. (W 6, S. 66) Mit dieser Maschine lassen sich die Folgerungen einiger klassischer Syllogismen (jedoch nicht alle), dargestellt durch vier Buchstaben und ihre Negationen, mechanisch ziehen. Was Peirce detailliert an formalen und materiellen Mängeln anmerkt, aber auch seine mathematisch-logische Beschreibung der Operationen dieser logischen Maschinen, kann ich hier nicht diskutieren.

Ich wende mich vielmehr zunächst einem Brief an Marquand zu. In ihm schlägt Peirce eine Verbesserung dieser logischen Maschine vor. Er empfiehlt ihm, statt mechanischer Mittel doch besser Elektrizität zu verwenden. Er schreibt: „Es seien A, B, C drei Tasten oder andere Punkte, an denen der Stromkreis offen oder geschlossen sein kann. In Abbildung 1 wird ein Stromkreis nur dann hergestellt, wenn alle geschlossen sind; in Abbildung 2 wird ein Stromkreis nur dann hergestellt, wenn einer geschlossen ist. Dies entspricht der Multiplikation [Konjunktion „und“] und Addition [Disjunktion, „oder“] in der Logik.“ (NEM, III, S. 632) Dieser Brief ist ein für die Geschichte der Technologie der künstlichen Intelligenz wichtiges Dokument. Denn er liefert die erste Beschreibung des Ansatzes zu einer Schaltalgebra, die für elektrische Ströme konzipiert ist. Eine Schaltalgebra setzt logische Funktionen in Schaltungen für elektrische Ströme um. Konjunktion und

Disjunktion reichen ja bereits aus, um zusammen mit der Negation alle aussagenlogischen Beziehungen durch Schaltungen auszudrücken.

Es ist dieses Modell einer sehr einfachen logischen Maschine, an dem Peirce 1887 in dem Aufsatz „Logical Machines“ (in: American Journal of Psychology, repr. in: NEM III,⁵ S. 625-632 und in W 6, S. 65-72) die Möglichkeiten und Grenzen der KI diskutiert. In einem ersten Schritt beschreibt Peirce das logische Prinzip, das menschliches und maschinelles Denken teilen. Dies ist das Prinzip einer konstanten semantischen Beziehung: Jede Umformung und jede Folge von Operationen von Symbolen muß den Wahrheitswert unverändert lassen. Peirce formuliert dies so: „Es besteht darin, dass unabhängig davon, welche Beziehung zwischen den Objekten im Fokus Angelpunkt einer Schlußfolgerung stehen soll, es stets möglich sein muß, dieselbe allgemeine Relation zwischen den Teilen der Maschine herzustellen.“ (NEM III, S. 628) Aber er hat bereits die Verallgemeinerung der mechanischen Umsetzung von Schaltungen vor Augen. Die Maschine von Marquand vergleicht Peirce mit einem Handwebstuhl, und er blickt auf die nächsten Entwicklungsschritte, wenn er schreibt: Die Logik würde sehr viel lernen können, wenn sie untersuchen würde, wie der Übergang zu dem Analogon zu einem automatischen, durch Lochkarten gesteuerten Modell möglich wäre.

Peirce geht in „Logical Machines“ davon aus, dass dieselben allgemeinen Gesetze einer Logik 1. Stufe in jeder funktionierenden logischen Maschine und ebenso durch jedes logisch schließende menschliche Denken

⁵ NEM“ ist das Sigel für „The New Elements of Mathematics by Charles S. Peirce“, 4 Bde. in fünf Teilbänden: I, II, III/1, III/2, IV, hg. v. Carolyn Eisele, Den Haag/ Paris/Amsterdam 1976.

realisierbar sind. Doch logische Maschinen sind im Unterschied zur natürlichen Intelligenz zwei Einschränkungen unterworfen:

„Erstens entbehren sie aller Originalität oder aller Initiative. Sie können ihre eigenen Probleme nicht ausfindig machen; sie können sich nicht selbst füttern. Sie können nicht selbst entscheiden, welche von verschiedenen möglichen Verfahrensweisen zu wählen ist. ... Sie kann nicht die Initiative ergreifen. Dies jedoch kein Fehler der Maschine; wir wünschen nicht, dass sie ihre eigenen Zwecke verfolgt, sondern die unseren. ... Zweitens hat die Fähigkeit der Maschine absolute Begrenzungen; sie wurde entworfen, um bestimmte Dinge zu tun, und sie kann nichts anderes tun. ... Der Geist ohne weitere Hilfsmittel ist ebenfalls in dieser wie in anderen Hinsichten begrenzt; aber der Geist, der mit einem Bleistift und viel Papier arbeitet, kennt keine derartige Begrenzung. Er strebt immer weiter, und welche Grenze seiner Fähigkeit auch immer heute gesetzt werden mag, kann morgen überschritten werden. Dies ist es, was die Algebra zu dem besten Instrument des Denken macht für das nichts zu kompliziert ist.“

(Ebenda)

Peirces erster Kritikpunkt, die mangelnde Initiative logischer Maschine habe ich teilweise ganz zu Anfang bereits dort diskutiert, wo ich über die mangelnde Fähigkeit von Computer zur Selbsterhaltung und Selbstsorge gesprochen habe. Im folgenden sahen wir, daß Originalität und Initiative das sind, was durch die selbstkritischen Entscheidungen und kontrollierten Abstraktionen eines Denkenden vor allen im engeren Sinne logischen

Folgerungen zu leisten ist. Die Rolle von Selbstkontrolle und Selbstkritik ist intern: Sie verbindet Episoden logischen Schließens und gibt ihnen einen normativen Kontext. Bei allen Unterschieden im Einzelnen besteht Selbstkritik stets darin, dass wir vergangene Handlungen oder Gedanken nach einem Maßstab, anhand eines verfolgten Zwecks oder Ziels, in ihrem Erfolg bemessen. Das Ergebnis dieses Urteils wird dann in komplexen Überlegungen beurteilt. Dabei geht es um die Konsequenzen, die aus der in diesen Überlegungen vorgenommenen Bewertung des Erfolgs zu ziehen sind. Diese zweite Phase der Ergebnisabwägung wird irgendwann zu einem Entschluss führen, der Verhalten in ähnlichen Fällen, den Zweck selbst oder den Zusammenhang mit anderen Zwecken betrifft. Nur wenn es zu einem Entschluss kommt, kann der Zweck, der Zusammenhang mit anderen Zwecken und das künftige Handeln oder Denken in ähnlich gelagerten Fällen durch die kritische Ergebnisabwägung beeinflusst werden. Abwägungen konkurrierender Zwecke, Entscheidungen und Bewertungen münden also in der Ergebnisbewertung und Entschlußfassung. Die Ausführung eines einmal gefaßten Entschlusses ist dagegen ein berechenbarer, mechanischer Prozeß. Dies ist sozusagen die algorithmische Phase des Denkens.

Die oben beschriebene Bild von der abstrahierenden Bildung von Zwecken machte diese Form menschlichen Denkens davon abhängig, dass eine natürliche Intelligenz kritisieren und kontrollieren kann, welche abstraktiv erfassten Erfahrungsgehalte und Qualitäten sie zu Zwecken macht, die aufgrund zu strukturierenden Faktoren im biographischen Leben einer Person werden und welche nicht. Aus diesem Bild folgt, dass dieser

Zweckausbildungsprozeß solange eine Grenze für die KI bildet, wie für diese keine unbegrenzt selbstkontrollierte Entwicklung von Zwecken möglich oder wünschbar ist. Dieser verallgemeinerte Einwand besagt dann, dass ein selbstkritisches Lernen in dem Verfolgen eines Zusammenhangs zwischen den Zwecken, Gedanken und Handlungen im Leben einer Person durch eine offene, aber kontrollierte Entwicklung von Zwecken nur für eine natürliche Intelligenz möglich ist.

Eine solche allgemeine Einschätzung der Möglichkeiten künstlicher Intelligenz setzt sie zu der Intelligenz von Personen in Beziehung und bemißt sie daran, ob und inwiefern maschinelles Denken die Fähigkeit zur selbstkritischen Entwicklung erreicht. Der Ansatzpunkt ist dabei, dass der charakterbildende Gebrauch menschlicher Intelligenz nur möglich ist durch die kontrolliert-kritische Bewertung nicht nur der eigenen logischen Überlegungen, sondern auch von Entscheidungen, Handlungen und Zwecksetzungen. Peirce machte einen ganz ähnlich gelagerten Vorschlag. Er hat dabei jedoch nicht an die Entwicklung von Personen im Blick. Es geht ihm allein um die Autonomie des menschlichen Denken, die nur durch eine prinzipiell unbegrenzbare Selbstkontrolle gewährleistet werden kann: „Beim echten Schlußfolgern sind wir nicht an unsere Methode gebunden. Wir billigen sie mit Bedacht, sind aber auch stets bereit und darauf eingestellt, sie aufs neue zu prüfen und zu verbessern und unsere Kritik an ihr zu kritisieren - ohne Unterlaß.“ (NZ, S. 332)]

Die Auswahl sowohl der Denkmethode, Zwecke und Handlungsmaximen wie die Art ihrer Prüfung sollte nicht nur selbst geprüft

werden, sondern die Auswahl und Abfolge von Denk- und Prüfoperationen sollte sich nach dem angestrebten Zweck richten. Der Zweck und das Ziel einer Überlegung oder Handlung sollte somit selbst der Kritik zugänglich sein. Damit haben wir ein Kriterium, nach dem wir eine intelligente, nämlich kritische Konstruktion einer Überlegung oder Handlung von einer unkritischen, weniger intelligenteren unterscheiden können. Kurzum: Eine Maschine, die nicht in der Lage ist, ihre Vorgehensweise selbstkritisch am Ergebnis und am Ziel ihrer Aufgabenstellung zu überprüfen, ist nicht im vollen Sinne intelligent zu nennen. Aus seinen Überlegungen zieht Peirce deshalb die folgende Konsequenz:

„Wenn eine Maschine nach einem festen Prinzip arbeitet, das in ihrem Plan enthalten ist, kann sie eine nützliche Hilfe beim Schlußfolgern sein. Falls sie jedoch nicht so eingerichtet wäre, dass sie sich in dem Fall, dass ein Fehler in ihr steckt, selbst verbessert, so würde die Maschine selbst keine Garantie dafür bieten, dass ihre Konklusionen korrekt wären, obwohl sie jede mögliche Konklusion aus den Prämissen korrekt berechnen könnte. Eine derartige Garantie könnte sich lediglich aus unserer kritischen Prüfung von ihr ergeben. Folglich wäre sie strenggenommen keine Folgerungsmaschine.“ (NZ, S. 332)

Wir wissen, was Peirce noch nicht wußte: Wir können prinzipiell einen Computer - auf jeden Fall eine universale Maschine - so programmieren, dass er jedes Programm, das er ausführt, durch ein Kontrollprogramm überprüft wird, das seinerseits wieder ein Algorithmus ist, also auf effektiv berechenbare Funktionen basiert. Doch ist die Unbegrenzbarkeit der

Selbstkontrolle nur im Prinzip - also für die universelle Maschine - richtig, und sie berücksichtigt nicht die Art von Kontrolle, die von den kontingenten Bedingungen des Gebrauchs und der Anwendung der Computers abhängt. Unter den Anwendungsbedingungen kann nicht jedes Kontrollprogramm wieder kontrolliert werden, und wir können auch die Flexibilität der Wahl von Kontrollstufen und das kontrollierte Einbeziehen von Abstraktionen, die sich erst aus der Anwendung ergeben, nicht im Programm festlegen.

Doch ist die natürliche Intelligenz des Menschen nicht auch in Fragen der Kontrollfähigkeit ähnlichen Einschränkungen ausgesetzt wie die KI? In der Tat: Jeder Mensch ist solange rational wie er oder sie in der Lage ist, seine Handlungen und Überlegungen zu kontrollieren und die Verantwortung dafür zu übernehmen, wie gut sie durchgeführt sind und ob sie erfolgreich waren. Alles, was verhindert, dass wir unser Denken und Handeln selbstkontrolliert ausführen, verhindert auch, dass wir rational und damit intelligent agieren und für das Ergebnis unseres Denkens und Handelns verantwortlich sind. Die Einschränkungen des Ausübens von natürlicher Intelligenz und damit der menschlichen Rationalität können interner und externer Natur sein: Eine seelisch bedingte Unfähigkeit wie das traumatisch bedingte Ignorieren bestimmter Tatsachen oder aufgrund einer emotionalen Wertung schränkt ebenso eine selbstkontrollierte Rationalität ein wie eine äußere, gewalttätige Bedrohung, die uns daran hinderte bestimmte Informationen zur Kenntnis zu nehmen oder Schlußfolgerungen zu ziehen. Doch im Unterschied zu Computern können menschliche Personen sich auf die durch ihren Charakter gesteuerte Flexibilität durch die Entwicklung ihrer Selbstkontrolle und

Abstraktionen zurückgreifen.

Selbstkontrollierte Zwecke und die Freiheit der Person

Trotzdem gilt: Es gibt immer irgendeine endlich nahe Grenze für die Zahl der Akte der Selbstkontrolle und der selektiven Abstraktion, die auch für die natürliche Intelligenz möglich ist. Jede Folge von kontrollierenden Überlegungen wird stets nach einer endlichen Zahl von Schritten abbrechen. Folglich kann eine vollständige Selbstkontrolle geistiger Prozesse nicht einmal annähernd verwirklicht werden. Ich hatte deshalb weiter oben auch nur behauptet, dass eine menschliche Person in der Entwicklung ihrer Zwecke gegenüber der selbstkritischen Korrektur unbegrenzt offen, also stets zu weiterer Selbstkritik bereit sein sollte.¹⁶

Doch reicht die Bereitschaft zur Selbstkritik aus? Wir können diese Disposition dadurch stärken, dass wir ihr einen den Status einer epistemisch-ethischen Maxime unseres kognitiven Vorgehens geben. Etwa so: „Glaube niemals, dass irgendein Ergebnis Deiner Überlegungen oder eine Handlungsweise unwandelbar gültig und richtig ist und nicht durch weitere Kritik verbessert werden könnte“. Wie können wir das Verhältnis von natürlicher und künstlicher Intelligenz hinsichtlich der möglichen Grade der Selbstkritik genauer bestimmen? Eine KI ist in stärkerem Maße als eine

61 Peirce hat die Unmöglichkeit einer vollständigen Selbstkritik zwar eingeräumt, aber betont, dass die bloße charakterliche Eigenschaft der Bereitschaft zur Selbstkritik ihre effektive Unvollständigkeit ausgleicht: „Selbstkritik kann niemals vollkommen perfekt sein. Denn die beste Kritik ist ihrerseits stets für Kritik offen. Aber solange wir zu Selbstkritik sowie zu weiterem Lernen bereit sind, haben wir in dieser Bereitschaft eine Garantie dafür, dass wir in dem Fall, dass man in irgendeiner Frage überhaupt jemals zur Wahrheit vorstoßen kann, am Ende auch zu ihr vorstoßen werden.“ (NZ, S. 332)

natürliche I. einer häufigeren, effektiven Selbstkontrolle der 1. Stufe fähig. Da nur eine durch Programme vorgegebene Reihe von Zielen von ihr erreicht werden soll, ist ihre Entwicklungsmöglichkeit durch ihre notwendig eingeschränkten Möglichkeit, Zwecke zu entwickeln begrenzt. Doch wäre auch eine KI, einen idealen Programmierer und entsprechende ideale technische Möglichkeiten vorausgesetzt, in der Lage, immer neue Stufen der Selbstkontrolle zu erreichen. Andererseits wird menschlich-natürliche Intelligenz stets unter stark beschränkenden Bedingungen operieren.²⁷ Doch ist das wirklich eine Einschränkung von der Freiheit der Person? Die der direkten logischen Kontrolle entzogenen Arten geistiger Prozesse wie die instinktiven Reaktionen, Körperempfindungen, Emotionen, Gewohnheiten und Interessen kommt eine wichtige kognitive Funktion zu. Sie begrenzen nicht nur rationale Selbstkontrolle, sondern sie machen sie, indem sie eine eingrenzende Vorauswahl vornehmen, sie allererst möglich. Sie bestimmen den Rahmen und die vorgelagerten Ziele zu Selbstkritik fähiger Denkaktivität.

Die Klagen von Philosophen und Wissenschaftlern über die in öffentlichen Debatten geäußerte und wirksame „irrationale“ Emotionalität und das Hängen der veröffentlichten Meinung an überkommenen, also gewohnheitsmäßigen Überzeugungen, sind uns allen vertraut. Doch gerade Philosophen unterschätzen die konstruktive Funktion dieser Komponenten. Denn die rein logische Selbstkontrolle wird erst wirksam, weil sie durch Emotionalität und Gewohnheit eingeschränkt worden ist: Ein sich vollständig

²⁷ Denn richtig ist, was auch Peirce bemerkt, dass es nichts gibt, „nicht einmal unsere Gedanken, worüber wir vollständige Kontrolle hätten.“ (NZ, 332)

rein logisch kontrollierender Mensch, könnte keinen individuell “eigentümlich” personalen Charakter entwickeln, der einen umfassenden Zusammenhang unterschiedlicher Aspekte und Teile seiner Person in relativer Eigenständigkeit zueinander in Beziehung setzt. Doch nur durch die eigentümliche, nämlich sowohl individuelle wie kreative Gewichtung von in manchen Bereichen einander ausschließender Aspekte, Zwecke und Erfahrungen ist individuell gelebten Lebens möglich. Der vollständige logische Kontrolleur dagegen könnte keine eigentümliche Ordnung von Zwecken, Impulsen, Gewohnheiten und Emotionen zulassen. Er könnte die “lichtbringenden Zufälligkeiten” der Erfahrung nicht wahrnehmen oder gar ernstnehmen - eben weil er sie nicht einzubeziehen weiß. Wenn wir abschließend die Möglichkeiten KI mit den restriktiven Bedingungen der Intelligenz von Personen vergleichen, zu welchem Ergebnis kommen wir? Zum einen sehen wir: Die kontingenten Bedingungen natürlicher Intelligenz durch andere geistige Prozesse, haben gerade nicht die Wirkung, die natürliche Intelligenz zu behindern, sondern sie ermöglichen es erst, dass rein formale Prozesse zum inhaltlich reichen und Denken von Menschen werden können, die durch selbstkontrollierte Entwicklung einen vielschichtigen Charakter entwickeln können. Natürlich sind auch alle Realisierungen der KI nur unter endlichen Bedingungen der jeweiligen Hardware und Software möglich. Aber diese Bedingungen sind nicht so, dass wir Grund hätten, ihnen eine selbstkontrollierte autonome Entwicklung von Zwecken zuzubilligen. Peirce schreibt: „Demgemäß müssen wir ... damit rechnen, unter unseren natürlichen Handlungen jeden Grad von Kontrolliertheit zu finden, von der

fast völligen Blindheit bis zur völligen Vernünftigkeit.“ (NZ, S. 335) Ich fasse zusammen: Es gibt einen wichtigen Unterschied zwischen KI und natürlicher Intelligenz, der von der Aufgabe abhängt, die wir der KI zubilligen. Denn dieser begrenzende Unterschied hat mit dem Werkzeugcharakter von Computern zu tun, der die Konstruktion wirklich autonomer Stufen der Selbstkontrolle widersprüchlich machen würde: Ein autonomes Werkzeug, das sich selbst Zwecke setzt und weiterentwickelt, ist kein Werkzeug mehr. Andererseits besteht die natürliche Intelligenz von Menschen darin, auch für die von ihnen verfolgten Zwecke neue Grade der Selbstkritik und damit der Selbstkontrolle einführen zu können. Dabei können sie auf dem bereits hohen Selektionsvermögen aufbauen, das durch die Antriebs- und Gefühlsausstattung vorgegeben ist. Ihre selbstkritischen Akte einer rational gewichteten Wahl werden so erst möglich. Darin zeigt sich die Autonomie natürlicher Intelligenz, dass sie sich in der Fähigkeit zum Erschließen von Wahlmöglichkeiten äußert. Sie ist es, die es Personen ermöglicht, ihren Charakter durch die Entwicklung von Zwecken auszubilden.

Diese Freiheit der Entscheidung über das eigene Handeln und der Wahl von Zwecken stellt die Verbindung zu der zu Anfang formulierten Differenzthese her. Für Personen, im Unterschied zu Computern, ist die Fähigkeit konstitutiv, durch kontrollierte Abstraktion aus dem Kreis ihrer Wahrnehmungen, ihres Spürens, ihrer Träume und Fantasie hervorzutreten, indem sie sie durch die Wahl von Zwecken so entwickeln, dass sich ein individueller Charakter ausbilden kann. Vielleicht gibt es so etwas wie Freiheit im metaphysischen Sinne nicht. Aber es reicht für ein

menschenwürdiges Leben und ist Zeichen menschlicher Kultur, dass wir die Wahl haben, Personen zu werden.