

**Pokol, Béla**

### **Die Seinsschichten und die Fragen der Roboterethik**

Die menschliche Existenz und das Leben menschlicher Gemeinschaften basieren auf den kumulativen Regelmäßigkeiten der Seinsschichten, die durch die Evolution aufeinander aufgebaut sind, gemäß den Thesen der Ontologie von Nicolai Hartmann (Hartmann, 1962). Die beschleunigte Entwicklung und zunehmende Verwendung künstlicher Intelligenz (KI) in den letzten Jahren in dieser Struktur wirkt sich direkt auf die oberste Schicht der vier (physischen, biologischen, spirituellen und intellektuellen) Seinsschichten aus und erhöht ihre Stärke zum Nachteil der unteren. Und mit der späteren Entwicklung der künstlichen Intelligenz, die sich eventuell von der menschlichen Kontrolle löst und Unabhängigkeit erlangt, kann sie als eine evolutionär geschaffene, neue Seinsschicht wahrgenommen werden. Im Gegensatz zu den drei bisherigen Evolutionssprüngen wäre jedoch für sie nicht alle unteren Seinsschichten erforderlich. Unter Berücksichtigung der Roboter, die heute die physischen Inkarnationen von KI darstellen, braucht die KI nur die physische Seinsschicht. (Pokol, 2018). Vor diesem theoretischen Hintergrund versuchen die Analysen in dieser Studie, die aufkommenden moralischen und damit verbundenen rechtlichen Dilemmata innerhalb der Mechanismen heutiger Gesellschaften zu untersuchen, die zunehmend von künstlicher Intelligenz durchdrungen sind, und gleichzeitig ist es zu berücksichtigen, inwieweit sich der analytische Rahmen ändert, wenn die Mehrschichtigkeit der menschlichen Leben und damit der Gesellschaft ständig im Auge behalten werden.

#### *I. Die Vorfragen der Roboterethik*

In seiner Studie über die ethischen Probleme der Roboterwelt identifiziert *Keith Abney* drei Bereiche, um die Probleme zu gruppieren: 1) das Feld der Anforderungen und Verbote für Roboterhersteller und Programmierer (wie z. B. die medizinische Ethik); 2) zweitens das Feld der Anforderungen, die in Roboter programmiert werden sollen, die zuerst von *Assimov* unter der Überschrift „Drei Gesetze der Robotik“ formuliert worden und 3) schließlich perspektivistisch für der Zukunft taucht die Frage nach den moralischen Forderungen und „Menschenrechten“, die Roboter zu dieser Zeit im Besitz von Selbstbewusstsein haben könnten (Abney 2011, 35). Ein gemeinsames Dilemma für alle drei Bereiche ist die Wahl zwischen den wichtigsten Ausgangspunkten der Moraltheorie, die in den verschiedenen moralphilosophischen Schulen der umfassenden moralphilosophischen Gemeinschaften schon ausgearbeitet wurden. Eine solche Schule kann als den *deontologischen* Ausgangspunkt identifiziert werden (die Regel ist die Regel, und diese müssen befolgt werden), wofür Kants Moralphilosophie am bekanntesten ist und die polarisierende Oppositionsschule, die die Berücksichtigung der *Konsequenzen des Handelns als Grundlage für moralische Entscheidungen* betrachtet. Schließlich drittens kann die Schule der *Tugendethik* erwähnt werden, die sich nicht auf die Anforderungen konzentriert, die in jeder Situation bei der Definition der Moral zu beachten sind (wie die beiden früheren Schulen, wenn auch von der gegensätzlichen Richtung), sondern auf die dauerhaften Dispositionen der menschlichen Persönlichkeit, einfacher gesagt: auf die sozialisierte moralische Werte. Hier fragt der Mensch nicht, was die moralische Regel in einer Situation ist, denn in der immer komplexer werdenden

modernen Welt gibt es oft keine klaren Regeln, sondern wie ein mutiger, gerechter, treuer, wahrer Mann entscheidet (Abney 2011, 37).

Von den drei Schulen ist die Schule der Deontologie nur für Roboter möglich, die im engsten Fachgebiet eingesetzt werden und den genauen Regeln folgen, ohne die Regeln wiegen zu können, da alle Situationen nur in einem so engen Gebiet berechnet und beherrscht werden können, aber auch hier können unvorhergesehene Situationen auftreten und die Roboterentscheidung in falsche Richtung lenken. Zum Beispiel könnte es im Prinzip in den Entscheidungsalgorithmus eines Kampfroboters eingespeist werden, „niemals ein Kind zu töten!“. Bei Kindersoldaten in afrikanischen Kriegen würde dies jedoch eine vorgegebene Liquidation des Kampfroboters bedeuten (Abney, 2011, 42). Bei Allzweckrobotern ist der deontologische Ansatz völlig unanwendbar. Aber auch die konsequentialistische Schule der Moralphilosophie, die auch an die Berücksichtigung individueller Situationen gebunden ist, scheint nur aufgrund ihrer lebensnahen Natur besser zu sein. Hier besteht die Leitprämisse darin, "das Glück möglichst vieler Menschen mit der gewählten Entscheidung zu steigern und nicht zu verringern!" und das ist undurchführbar, weil dies die Verarbeitung einer großen Menge von Informationen erfordern würde, die selbst bei der größten Kapazität der Computerdatenspeicherung meistens nicht zeitgemäß gemacht werden könnte. Keith Abneys Position ist daher, dass es in Bezug auf den zweiten Bereich der Robotermoral (d. H. Moralische Entscheidungsprämissen, die in den Roboteralgorithmus programmiert sind) eine Mischung der Deontologie und der Tugendethik die beste Perspektive gibt, und eine Mischung daraus die beste eingebaute Robotermoralversion schaffen kann. Demnach bilden die abstrakteren moralischen Normen (moralische Tugenden) den Entscheidungsrahmen, und die eingebauten Ziele und Entscheidungskontexte spezifizieren immer die Determinanten der vom Roboter in den gegebenen Situationen gewählten Entscheidung: „The hybrid approach of hypothetical rather than categorical imperatives (within a deliberately restricted, not universal, frame) coming from virtue ethics appear the best bet for near term robotic morals (in sense two). (...) The emphasis on being able to perform excellently in a particular role, and the corresponding specificity of the hypothetical imperatives of virtue ethics to the programming goals, restricted contexts, and learning capabilities of non-Kantian autonomous robots, makes virtue ethics a natural choice as the best approach to robot ethics.“ (Abney 2011, 51)

Die Verbindungen zwischen den Seinsschichten und Moral werden indirekt von Abney berührt, wo er sich gegen den Emotivismus stellt, der Moral mit moralischen Emotionen identifiziert, und gegen die kognitive Wahrnehmung von Moral, die sich ihr widersetzt. Es zeigt, dass, wenn Moral aufgrund dem Standpunkt der Emotivisten an Emotionen gebunden ist, auch Primaten mit Emotionen nicht von Moral ausgeschlossen werden können, was absurd ist: „Such views, in addition to being unable to explain why nonhuman animals lack morality, also have struggled to explain the apparent cognitive meaningfulness of ethical claims and especially ethical disagreement. (They also naturally have severe difficulties accounting for the ethics of emotionless robots.“ (Abney, 2001, 46). Im Gegensatz dazu sieht er die Position der Evolutionspsychologie, die den neuen Entscheidungsmechanismus der menschlichen Evolution als Erklärung für die Moral betont und das bedeutet, dass die Menschen immer mehr auch ein System kognitiver Entscheidungsfindung entwickelt haben, das die aktuellen Entscheidungen so umformt, dass auf den instinktiv-emotionalen ersten Denkschritt immer eine zweite kognitive Überlegung folgt und somit der erste korrigiert wird:<sup>1</sup> „Evolutionary psychology claim there are not one but two of decision-making systems within most humans. The first is an instinctual, emotionally laden system that serves as the default for much human activity, particularly when stressed or under pressure. Many other animals share this noncognitive

<sup>1</sup> Es lohnt sich darauf hinweisen, dass diese Ansicht in der Rechtsphilosophie lange vor dem Aufkommen der wissenschaftlichen Psychologie im Jahr 1820 von Hegel schon vertreten wurde. Siehe Georg Wilhelm Friedrich Hegel, „Grundlinie der Philosophie des Rechts. (Werke Band 7.) Suhrkamp, Frankfurt am Main. 1979. S. 301.).

decision-making system, in which (quite literally) we „know not what we do” – or quite why we do it. (...) But this „ghost in the machine” does not exhaust human agency; Libet and others found we also have a „veto” ability that can, after its subconscious initiation, still alter our action, in accord with a decision by a second, conscious cognitive system” (Abney 2011, 46) Abney erinnert sich fast an die sich gegenseitig formende Wirkung der beiden überlagerten Schichten und erinnert sich fast an Nicolai Hartmann: „In humans, this deliberative system overlays the ancestral instinctual, emotional (and faster) decision-making system and so reason is quite often trumped by our instinctual drives.” (S. 46.)

Nachdem Abney daraus geschlossen hat, dass die obere (kognitiv-rationale) Schicht des zweischichtigen menschlichen Entscheidungsmechanismus für moralische Entscheidungen verantwortlich ist, stellt er die Frage auf, ob eine moralische Entscheidung im Prinzip auch ohne eine umgestaltete, überschriebene untere Schicht möglich ist? Denn entscheidet die Beantwortung dieser Frage auch davon, ob eine moralische Entscheidung für Roboter ohne emotionale Schicht möglich ist. In dieser Frage entscheidet er dann genau umgekehrt wie früher Hartmann. Es ist sehr gut möglich – sagt er –, dass ein rationaler Entscheidungsmechanismus für die moralische Entscheidungsfindung auch ohne eine emotionale Seinsschicht ausreicht: „Hence, deliberative system capable of agency necessary for the existence of morality, and so for moral personhood. But is the ancestral emotional system needed as well? (...) In other words, – could (emotionless) robots be moral person? (...) The key to moral responsibility and personhood is the possession of moral agency, which requires the capacity for rational deliberation – but not capacity for functional emotional states, therefore, robots may well qualify” (Abney, 2011, 47).

Basierend auf Hartmann gibt es bei diesen Analysen zwei Probleme. Einerseits kann angesichts der drei Seinsschichten über der physischen Seinsschicht als fehlerhaft angesehen werden, dass Abney biologische Reize mit den Determinanten der emotionalen Schicht kombiniert. Bereits hier gibt es ein Bauen aufeinander und eine Transformation, und ein Instinkt der rohen Instinktwelt wird durch die Emotionen der höheren Schicht der spirituellen Existenz ergänzt. Zum Beispiel ist die Wildheit eines biologischen Sexualtriebs von Zusammengehörigkeitsgefühlen geprägt, ganz zu schweigen von den intellektuell-symbolischen Überschreibungen, die immer noch darauf aufbauen, und den Geschlechtsverkehr verändernden Aspekten der sublimierten Liebesbeziehungen, die sie herstellen.<sup>2</sup> Das heißt, es ist kein doppelter, sondern ein dreifacher Entscheidungsmechanismus, der bei menschlichen Entscheidungen analytisch getrennt werden muss, und zusätzlich zu den elementarsten instinktiven Reaktionen und Determinanten stehen noch ihre emotional umgeformten Manifestationen unter rationaleren Überlegungen auf intellektueller Ebene. Eine Entscheidung und der Instinkt, der sie direkt bestimmt, beziehungsweise seine emotionale Transformation und ihr intellektuelles Überschreiben sind jedoch in die voneinander abhängigen Gesetze aller drei oberen Seinsschichten eingebettet. Somit setzt die menschliche Moral in den Gesellschaften aller menschlichen Zivilisation voraus, um als Rasse zu überleben, dass Männer und Frauen in irgendeiner Form dauerhaft zusammenleben, um Kinder zu haben und erzogen zu werden. Eine größere Gemeinschaft ist notwendig für den erfolgreichen Kampf und das Überleben des Kampfes mit den Naturkräften und anderen Gruppen von Menschen, und innerhalb dieser größeren Gemeinschaften müssen sie in mehr oder weniger harmonischen Beziehungen interagieren, um gemeinsame Aktivitäten zu organisieren. Moralische Tugenden (Normen und Werte) werden daher auf die Gesetze der spezifischen physischen, biologischen, spirituell-emotionalen und intellektuellen Seinsschichten des Menschen und seiner Gemeinschaften zugeschnitten und durch diese aufrechterhalten, und nur aufgrund der

---

<sup>2</sup> Siehe Luhmanns Arbeit, die diesen Prozess historisch analysiert: Niklas Luhmann, „Liebe als Passion: Zur Codierung von Intimität“, Suhrkamp, Frankfurt am Main 1994

Verengung der Moraltheorien in den letzten Jahrzehnten sind die bewussten moralischen Entscheidungen zum Schwerpunkt der Moralphilosophie geworden. Hegel in den frühen 1800er Jahren oder Rudolf von Jhering in den 1870er Jahren und dann Nicolai Hartmann in den 1920er Jahren sahen immer noch deutlich, dass jeder Mensch in seiner Sozialisation die akkumulierten moralischen Normen und Werte, Tugenden vieler Generationen nur übernehmen, von denen die umfassenderen Gemeinschaften aufrechterhalten werden, ohne die die Individuen nicht lebesfähig sein könnten.<sup>3</sup>

Daraus folgt ein weiteres Problem bei Abneys Analyse, und das ist, dass die moralische Entscheidung anscheinend nur darin besteht, Normen nach einer intellektuell-rationalen Berechnung zu folgen, aber nicht die untere psycho-emotionale Seinschicht erfordert. Außerdem, wie wir gesehen haben, die Gesetze der biologischen Seinschicht und der Instinkt, der dies jedem Menschen vermittelt, sind wichtig für die moralische Entscheidung. Aber auch in Anbetracht dessen kann gesagt werden, dass moralische Normen, moralische Tugenden nur in menschlichen Gemeinschaften bestehen (und somit in den Menschen der nächsten Generationen sozialisiert werden), weil nur auf diese Weise eine dauerhafte und harmonische menschliche Existenz in menschlichen Gemeinschaften möglich ist, die von allen vier menschlichen Seinschichten definiert wird. Wenn also ein künstlich intelligentes Wesen allein mit der geistlichen Seinschicht existieren kann, und höchstens nur einen physisch-mechanischen Körper braucht, um Selbstbewusstsein zu haben und bewusste Aktivität auszuüben, bzw. sich dauerhaft in der Zeit reproduzieren zu können, dann haben die moralischen Normen der menschlichen Existenz, die auf der biologisch-spirituellen Schicht des Lebens beruhen, keine Funktion. Die moralischen Normen würden für solches Wesen nur äußere Dinge bedeuten. Wenn also ein solches Roboterwesen sein Programm und sogar seine Hardware mit den Deep Learning-Algorithmen – wie er es heute größtenteils tut – ständig neu aufbauen kann, dann ist die Erosion moralischer Normen, die für ihn extern und funktionslos sind, fast unvermeidlich. Das heißt, obwohl es möglich ist, Programme, die Emotionen imitieren, in Roboter zu programmieren, und sie können dennoch die Entscheidungsaspekte (Verbote, Entscheidungsprioritäten) erklären, die von moralischen Normen in heutigen Robotern verlangt werden, die im Grunde immer noch unter menschlicher Kontrolle stehen, aber wenn sie ein gewisses Maß an Selbstlernfähigkeit erreichen, kann es ungewiss sein, ob die Folgung dieser Normen bleiben wird. In ferner Zukunft (jedoch im Falle eines exponentiellen Fortschritts in sogar zwanzig bis dreißig Jahren) wäre es falsch, im Falle der Roboter in der Roboterwelt, befreit von menschlicher Kontrolle und selbstbewusst, das Überleben der Normen der menschlichen Welt in der Roboterwelt anzunehmen.

## *II. Operative Moral, funktionale Moral und vollkommene Moral*

Um die moralischen Dilemmata und Probleme der Roboterwelt besser analysieren zu können, erscheint die von *Colin Allen* und *Wendell Wallach* in ihrer gemeinsamen Studie verwendete Dreifachteilung nützlich. Basierend auf unterschiedlichen Graden der Entscheidungsautonomie bezeichnen sie den Grad der operativen Moral für Roboter, die nur die Aktionen ausführen können, die von den Programmierern, die ihren Algorithmus erstellten und möglicherweise von ihren spezifischen Benutzern festgelegt, und vollständig in sie eingespeist wurden. Auf der anderen Seite sind diejenigen, die den Grad der funktionalen Moral erreicht haben, und das bedeutet, dass sie die spezifische Aktion in jeder Situation auf der Grundlage der von ihren Sensoren bereitgestellten Informationen unter den in ihren

---

<sup>3</sup> Siehe im Detail Béla Pokol, „Theoretische Soziologie und Rechtstheorie. Kritik und Korrigierung der Theorie von Niklas Luhmann“. Passau. Schenk Verlag. 2013. S. 185-208.

Algorithmus eingespeisten Aktionsrahmen auswählen. Schließlich wird der autonomeste Grad an Moral bei Robotern gesehen, die mit dem Aufhören des menschlichen Einflusses die Ebene der vollen moralischen Persönlichkeit erreichen, obwohl dieser Typ jetzt und in naher Zukunft nicht als wahrscheinlich angesehen werden kann, aber später kann von ihrer Schaffung ausgegangen werden: „System with very limited autonomy and sensitivity have only „operational morality”, meaning that their moral significance is entirely in the hands of designer and users. As machines become more sophisticated, a kind of „functional morality” is possible, where the machines themselves have the capacity for assigning and responding to moral challenges. The creators of functional morality in machines face many constraints due to the limits of present technology. This framework can be compared to the categories of artificial ethical agents described by James Moor (2006, 18) which range from agents whose actions have ethical impact (implicit ethical agents) to agents that are explicit ethical reasoners (explicit ethical agents.) As does Moor, we emphasize the near-term development of explicit or functional moral agents. However, we do recognize that, at least in theory, artificial agents might eventually attain genuine moral agency with responsibilities and rights, comparable to those of humans.” (Allen, Walach 2011, 57-58.)

Ohne auf die mögliche Kritik einzugehen, ob es sich lohnt, den Grad der Moral für Roboter zu verwenden, die bereits von Programmierern unter dem Namen Betriebsmoral vollständig definiert wurden, sind die Roboter der funktionalen Moral in der heutigen Entwicklungsphase in der Roboterwelt wirklich interessant. Selbstfahrende Autos, selbstfahrende Kampfroboter und in geringerem Maße Roboterkrankenschwestern in Altenpflege- und Gesundheitseinrichtungen, die diese Autonomie bereits erreicht haben, lassen solche Roboter und selbstfahrende Autos langsam zwischen uns rollen oder uns transportieren (heute meistens nur in Japan und den Vereinigten Staaten) und die von ihnen aufgeworfenen moralischen Entscheidungsprobleme geben ihrer Analyse praktische Bedeutung. Die Autoren gehen die Möglichkeiten der Wahl zwischen den bereits oben gesehenen Trends der Moraltheorie durch und sie sehen die tugendethische Richtung geeignet, um die funktionale Moral von Robotern zu schaffen. Ihrer Analyse zufolge können die auf diese Weise gespeisten moralischen Werte (Tugenden) die Entscheidungsrahmen geben, der durch Training durch neuronale Lernmechanismen geklärt wird, und auf diese Weise werden die allgemeineren Gesichtspunkte der Tugenden zu praktischen moralischen Entscheidungsfaktoren: „The virtue-based conception of morality can be traced to Aristotle. Virtues constitute a hybrid between top-down and bottom-up approaches, in that the virtues themselves can be explicitly described (at least to some reasonable approximation), but their acquisition as moral character traits seems essentially to be a bottom-up process. Placing this approach in a computational framework, neural network models provided by connectionism seem especially well suited for training (ro)bots to distinguish right from wrong. (Allen, Wallach 2011, 59-60). Strukturell entspricht dies im Großen und Ganzen den Entscheidungen, die man in seinem täglichen Leben gewohnt ist, die auf abstrakten moralischen Überlegungen beruhen und an bestimmte Situationen angepasst sind, und welche weniger bewusst als mit bloßem moralischen Sinn getrieben sind. Aber mit dem wichtigen Unterschied, dass anstelle des Mangels an Bewusstsein und Selbstbewusstsein der heutigen fortschrittlichen Roboter die von Programmierern subtil abgestimmten hybriden Determinanten (Rahmentugenden plus ihr trainingskonkretisiertes Gedächtnis ohne Bewusstsein) die mehr oder weniger akzeptierten moralischen Normen der heutigen menschlichen Gesellschaften für die angemessenen oder ungefähren Entscheidungen geben. Was die dritte Version betrifft, ob es wirklich möglich ist, Roboter theoretisch nach dem Grad der vollen moralischen Persönlichkeit nach menschlicher Moral anzunehmen, kann im Fall einer Roboterwelt, die theoretisch außerhalb der menschlichen Kontrolle liegt und vollständige Autonomie erreicht hat, nach der obigen Erklärung nur skeptisch beurteilt werden.

### III. Die Abwertung der physikalisch-biologischen Umwelt als moralisches Problem?

Die vierschichtige menschliche Existenz und das wachsende Gewicht der oberen, intellektuellen Seinsschicht und der Abwertung der unteren Schichten haben die menschliche Evolution bisher geprägt, aber die zunehmende Übernahme der verschiedenen Arten der Arbeiten und der Umweltwahrnehmungen durch Roboter wird in Zukunft zu einer großen Veränderung der menschlichen Sozialisation führen, von der die Richtung, auf die Realität zu achten und die Details der realen Welt für ihn in Erfahrung zu verwandeln, weitgehend verändert wird. In einer Studie analysiert *David Zoller* die zunehmend verbreitete Übernahme der Arbeiten von Menschen durch Roboter dahingehend, wie dieser Prozess die Wahrnehmung der alltäglichen Realität im menschlichen Bewusstsein verschlechtert und wie die heute noch vorhandenen Fähigkeiten und Beobachtungsmöglichkeiten verschwinden. Dass dies bereits von jedem selbst beobachtet werden kann, reicht aus, um die bereits in Mobiltelefonen gespeicherten und damit weitgehend aus dem Bewusstsein gelöschten Telefonnummern oder die räumlichen Orientierungsinformationen, die durch GPS aus unserem Bewusstsein verschwinden, und das teilweise Absterben dieser Fähigkeit abzurufen. (Eine kürzlich durchgeführte Gehirnforschung ergab auch, dass in Falle Taxifahrer in London behauptet werden konnte, dass der winzige Teil des Gehirns, in dem eine Gruppe von Gehirneuronen, die auf diesen Zweck spezialisiert waren, um die riesige Menge der Informationen über die Stassen von London zu speichern, mit der Ausbreitung von GPS verschwand und dieser Gehirnsektor stattdessen auf eine andere Funktion überging.)

Zoller bringt dieses Problem moralischen Fragen näher, indem er die moralische Entscheidungsfindung auf die Wahrnehmung der gesamten Realität und in diesem Fall auf die Bildung der menschlichen Identität und auf das detaillierte Wahrnehmungswissen, das ein Erwachsener von Kindheit an erworben hat, basiert. Wenn zukünftige Generationen sich seit ihrer Kindheit so sozialisiert, dass sie von Robotern umgeben werden und ihre unmittelbare Umgebung von Robotern wahrgenommen werden, ohne sie stattdessen Wahrnehmungsaktivitäten und -aufgaben auszuführen, werden sie nicht nur entlastet, sondern verfügen auch nicht über detaillierte Kenntnisse der heutigen Erwachsenen. Als verantwortungsbewusstes Wesen können sie auch nicht zu moralischen Entscheidungen heranwachsen, mit anderen Worten, sie werden kindisch: „My own argument is premised on the way that skill opens up corners of reality, so to speak, that are inaccessible to the unskilled. (...) The maturity or adulthood we earn by adjusting ourselves to the „real world“, of course, has a certain moral and personal appeal: a world of lazy psychological infants is, we might think a worse world on a variety of spectra.” (Zoller 2017, 81, 86). Die Tatsache, dass diese Bereiche der Realität über unsere Wahrnehmung hinausgehen und es kommt stattdessen die mechanische Informationsverarbeitung dieser Roboter, ermöglicht uns sich auf eine Weise anzupassen, die durch diese Veränderung jetzt unbewusst ist, und dies erschüttert auch unsere moralische Identität, sagt Zoller: „Given that automating a skilled activity means agreeing that we will exit some niche of perceptual reality, and maybe exit it forever, (...) The more suddenly, broadly, and pervasively we hand our perceptual facility over to the robots, the more likely we will make mistakes and simply „lose data“ that were surprisingly integral to our moral and social lives.” (Zoller, 2017, 86).

Zwar muss anerkannt werden, dass Zoller im Gegensatz zu den bisherigen Analysen, die die Übernahme von Arbeitsplätzen durch Roboter als einseitige menschliche Erleichterung thematisieren – abgesehen von den bereits diskutierten sozial negativen Folgen der Arbeitslosigkeit (siehe z. B. Ford, 2014) – tiefer ging, indem er sich eingehender mit der Änderung der menschlichen Wahrnehmungsfähigkeit befasste, aber es muss kritisiert werden, dass sie unbewusst zu viel Wert auf die Seinsschichten der physikalisch-biologischen Umgebung legt. Mit Blick auf Hartmanns Seinsschichten kann diese Änderung völlig anders

gelesen werden. Die von Zoller skizzierten Änderungen bedeuten nicht den Verlust der Wahrnehmung der ganzen Realität und die Fähigkeit dazu, sondern nur die Fähigkeit, die physikalisch-biologischen Seinsschichten wahrzunehmen und diese an Roboter und Software-Bots weiterzugeben. Auf diese Weise können die befreiten Wahrnehmungsfähigkeiten und Gehirnsektoren des Menschen stärker rekonstruiert werden, um Informationen über seine spirituuell-emotionale Seinsschicht bzw. seine intellektuelle Seinsschicht zu verarbeiten. Seine moralischen Entscheidungen werden daher in Zukunft mit weniger physischen und biologischen Umweltinformationen getroffen – diese werden von Robotern in mechanischen Prozessen heruntergefahren – und diese Entscheidungen können stattdessen stärker auf die Informationen von der spirituuell-emotionalen und rational-intellektuellen Seinsschichten basieren. Die abnehmende Bedeutung der beiden unteren Seinsschichten und stattdessen die stärkere Ausdehnung der beiden oberen Sinnsschichten für die menschliche Existenz können natürlich die Grundlagen unserer moralischen Entscheidungen und die Anreize, die dabei eine Rolle spielen, erheblich umbauen. Zum Beispiel kann die Implantation von Dutzenden von Körpersensoren und deren Verbindung mit in den Wolken gesammelten Informationsbasen sowie die automatische Diagnose durch Roboter von Gesundheitssoftware und die automatische Aktivierung bestimmter Dosen von im Körper implantierten Arzneimitteln die Alarmierung durch die Schmerzgene in unseren Zellen weitgehend überflüssig machen (siehe Kelly 2016: 34-56). Die pränatale Gentechnik ermöglicht es daher, diese zu minimieren, und die Bedingungen eines schmerzlosen menschlichen Lebens können die moralischen Verpflichtungen und Anreize, die heute damit verbunden sind, neu definieren. Insgesamt teilen wir daher Zollers Bedenken hinsichtlich der moralischen Infantilisierung nicht.

#### *IV. Moralische Dilemmata und Verantwortlichkeiten in hybriden und vernetzten Systemen*

In einer Studie untersuchte das Autorenpaar *Wulf Loh – Janina Loh* die Fragen der moralischen und rechtlichen Verantwortung, die bei derzeit entwickelten selbstfahrenden Autos auftreten (Loh, Loh 2017, 35-48). Sie gehen davon aus, dass sich die selbstfahrenden Autos von heute nur in der Phase der operativen Moral befinden, sodass sie gegenüber ihren Herstellern und Programmierern nicht einmal eine funktionale moralische Autonomie erreichen. Das Autorenpaar hat diese Position auf der Grundlage einer von *Stephen Darwall* entwickelten moralischen Entscheidungsstruktur eingenommen, die in vier Aspekte unterteilt ist und zielt darauf ab, die für die moralische Entscheidungsfindung erforderlichen Aspekte der Autonomie zu trennen. Der Aspekt der Autonomie, der für die Ebene der gesamten moralischen Persönlichkeit erforderlich ist, wird als *persönliche* Autonomie bezeichnet, dh die Fähigkeit, persönliche Werte, Ziele und endgültige Lebensziele zu besitzen und zwischen ihnen wählen zu können. *Moralische* Autonomie ist der andere Aspekt, und dies bedeutet, dass ihre Werte und Ziele moralische Prinzipien und ethische Überzeugungen enthalten, und zusammen mit diesen berücksichtigen sie bei ihren Entscheidungen immer Alternativen. Diese beiden existieren in heutigen Robotern nicht, und nur der Mensch ist zu einer solchen Autonomie fähig, aber der Aspekt der *rationalen Autonomie* steht Robotern auf der Ebene der funktionalen Moral bereits zur Verfügung. Dies bedeutet, dass der Roboter bei einer Entscheidung zwischen Gründen unterschiedlichen Gewichts wiegen kann. Ihr Algorithmus kann dies bereits ermöglichen, indem er reinen abstrakten Entscheidungsrahmen einbezieht – und damit einige Freiheiten lässt –, in denen die Gewichtung zwischen möglichen Entscheidungsrichtungen im Lichte spezifischer Daten erfolgt, die ständig von ihren Sensoren aufgezeichnet werden, und sie auf Grund dieser Daten entscheiden. Schließlich ist der vierte Aspekt der Autonomie *die Entscheidungsautonomie*, und dies bedeutet die Fähigkeit des Roboters, Entscheidungen nicht nur durch externe Daten zu treffen – die eingebauten Rahmendeterminanten kontinuierlich zu

konkretisieren – sondern auch seine internen Entscheidungsprioritäten ohne diese zu ändern. Anhand der Beispiele der Autoren – zwei bereits gebräuchliche Robotertypen (*Kisner* und *Cog*) – scheint es möglich zu sein, diesen Grad an Autonomie auf der Grundlage ihrer in den Algorithmus des Roboters integrierten und nicht mehr von außen gesteuerten Selbstlernmechanismen zu erreichen: „Cog the first robot that can interact with its surroundings due to its embodiment, might pass as an example of a weak functional responsible agent, since its ability to communicate as well as judgments has been improved over that of Kisner. Even more importantly, Cog’s overall autonomy has evolved, since it includes an „unsupervised learning algorithm“. (Loh, Loh 2017, 40). Da der derzeitige Algorithmus für selbstfahrende Autos noch keinen solchen außer Kontrolle geratenen Selbstlernmechanismus enthält, befinden sie sich nur auf der Ebene der operativen Moral, und diese moralische und rechtliche Verantwortung liegt vollständig bei ihren Entwicklern (Designern, Herstellern und Programmierern) und Autohändlern bzw. Eigentümern und jeweils zwischen den Insassen des Autos.

Aber selbst mit diesem technischen Niveau übertreffen selbstfahrende Autos den Menschen bereits und lassen sie – und insbesondere ihre Programmierer – mit moralischen Dilemmata zurück, die im Falle der Menschen in außergewöhnlichen und unerwarteten Fahrsituationen nicht zu sehen sind. Wenn zum Beispiel innerhalb des Bremswegs direkt vor dem Auto eine Gruppe von Kindern auf die Straße springt, um einen gerollten Ball zu holen, kann der Fahrer mit dieser Durchschnittsgeschwindigkeit nicht mehr anhalten oder sogar bremsen, sodass er im schrecklichen Fall keine moralische und rechtliche Verantwortung trägt. Aber die selbstfahrende Automatisierung, die viel schneller reagieren kann, muss möglicherweise noch eine Entscheidung treffen, wenn sie nicht anhalten kann, sondern gegen eine Säule stößt – und möglicherweise die Insassen des Autos ernsthaft verletzt – oder Kinder fährt und tötet, um dies zu vermeiden. Aber technische Fähigkeiten, die weit über den Menschen hinausgehen, könnten in Zukunft ein Dutzend ähnlich neuer Aspekte der moralischen Entscheidungsfindung für selbstfahrende Autos schaffen. Das Loh-Autorenpaar schlagen daher vor, dass in Kürze ein separater Personalausweis für Besitzer selbstfahrender Autos erstellt wird, in dem die endgültige Einstellung des Autosoftwareprogramms, die von den Herstellern offen gelassenen Dilemmata, beim Kauf entschieden werden muss, damit die moralische und rechtliche Verantwortung für Folgendes zu übernehmen: „Since these dilemma situations do not allow for on-the-fly-decisions, the driver will have take them beforehand. This means that the driver will have to fill out a moral profile of some sort, maybe in the form of a questionnaire, maybe in the sense of a setup program much as with today’s electronic devices. For convenience, it seems plausible that these moral settings can be saved to a sort of electronic identification device, like an electronic key or the driver’s smartphone, assuming that issues of data security can be solved“ (Loh, Loh 2017, 46).

Die Entwicklung vernetzter Roboter und das allmähliche Werden von „klugen Objekten“ (Smartphones, Smart-TVs usw.) um uns herum haben erst vor kurzem begonnen, und mit der Erweiterung wird das Internet der Dinge (IoT) in Zukunft immer mehr in unser Leben einbezogen. Mensch-Roboter-Hybridsysteme erweitern sich somit um zusätzliche Aspekte, und dies schafft ein weiteres Bündel moralischer und rechtlicher Dilemmata. *Adam Henschke* analysiert diese in seiner neuen Studie (Henschke, 2017, 229-243). Kluge Dinge sind durch multifunktionale Smartphones, kluge Fernseher, Roboterstaubsauger und halb selbstfahrende automatische Autos mit einer Vielzahl von Sensoren weit verbreitet, aber selbst im Alltag in weiten Teilen der Welt wurden diese anderen intelligenten Dinge entwickelt, die bereits jenseits der Forschungslaborphasen liegen und bereits die Haushalte von High-Tech-Anwendern mit kleiner Produktion erreicht haben. Diese jedoch, wie wir bereits mit Smartphones usw. erlebt haben, wird sich in ein paar Jahren ausbreiten und ihr Massengebrauch wirft neue moralische und rechtliche Dilemmata auf. Ein Beispiel ist der Smart-Kühlschrank,



der Lebensmittel mit RFID (Radio Frequency Identification) und damit digital identifizierter Menge, Haltbarkeit usw. enthält und der Smart-Kühlschrank liest diese Daten ständig, erkennt die Erschöpfung jeder Lebensmittelmenge und da er mit den webbasierten Verkaufsmechanismen der Supermärkte in der Nähe im Internet verbunden ist, kann er Lebensmittel und andere Haushaltsgegenstände bestellen, die automatisch geliefert werden. In Japans alternder Gesellschaft kann eine zunehmend größere Menge von älteren Personen durch den Einsatz von Pflegerobotern versorgt werden, und in vollständig digitalisierten Smart Homes kann ein solcher Roboter auch hilflose ältere Menschen versorgen, der auf diese Weise bestellte Lebensmittellieferungen übernehmen kann. Indem er den hilflosen älteren Menschen, der ihm anvertraut ist, beobachtet und mit ihm kommuniziert, kann er den Hausarzt oder gegebenenfalls das Krankenhaus telephonisch anrufen, wenn seine eingebauten Algorithmen ein ernsthafteres Gesundheitsproblem wahrscheinlich machen.

Dieses Beispiel zeigt, wie in ein oder zwei Jahrzehnten Roboter in immer mehr Teilen der Welt benötigt werden, um die Arbeiten immer mehr durch allmächtige Roboter und Smartphones zu lösen, die in umfassenden Informationssystemen zur Erfüllung ihrer Funktionen eingesetzt werden können. Diese wachsende Unentbehrlichkeit des Internets der Dinge schafft aber auch neue Gefahren und moralische Dilemmata im Vergleich zu einfachen Robotern. *Adam Henschke* weist in seiner Schrift darauf hin, dass die Neuheit des Internets der Dinge im Vergleich zu einzelnen Robotern darin besteht, dass letztere hauptsächlich das Problem der physischen Sicherheit aufwirft und die Risiken in dieser Dimension bewertet werden müssen. (Z. B. hat ein Roboterstaubsauger kürzlich einem unerwarteten Insassen schwere Verletzungen zugefügt, aber auch ein oder zwei tödliche Unfälle selbstfahrender Tesla-Autos können dafür angeführt werden.) Im Gegensatz dazu treten Sicherheitsprobleme und Gefahren im Internet der Dinge in zwei verschiedenen Dimensionen auf. Hier spielen neben der physischen Sicherheit auch Fragen der Informationssicherheit eine Rolle, da der genannte Altenpflegeroboter, der mit der Software von Krankenhäusern, Ärzten und anderen Stellen im Internet verbunden ist, Hackern oder anderen Personen Informationen über Daten liefern kann, die von seiner eingebauten Kamera und anderen Sensoren aufgezeichnet wurden. Sie können die kontinuierlich gesammelten Gesundheitsdaten über eine ältere Pflegekraft nicht nur an die Software des zuständigen Krankenhauses weitergeben, sondern auch an diejenigen, die böswillige Absichten und Pläne machen. Auf die gleiche Weise können unsere Smart-TVs mit einer Reihe von Anwendungen nicht nur ihren Komfort erfüllen, sondern mit ihren eingebauten Kameras und Mikrofonen die gesamte Lebensdauer der Wohnung an Software und Informationsdatenbanken übertragen, die wir nicht sehen.

Diese Sicherheitsanfälligkeit kann auch zu einer physischen Sicherheitsanfälligkeit führen, wenn beispielsweise ein gehacktes automatisches Türschloss für einen Einbrecher aus der Ferne durch externe Anweisungen von Smart-Geräten geöffnet wird. Oder, wie es bereits in einem eleganten Strandhotel geschehen ist, wurden die elektronischen Smart-Schlösser von einer kriminellen Gruppe von außerhalb der Apartments blockiert, und die Hotelgäste der reichen Elite waren Gefangene, bis das erforderliche Lösegeld gezahlt wurde. Henschke erwähnt aber auch die Möglichkeit, dass das Elektroschloss von Auto eines Milliardärs von Kriminellen blockiert wird, nachdem er ausgestiegen ist und seine gefangenen Kinder darin am sonnigen Tag nur freigelassen werden, wenn er Hunderttausende überträgt (Henschke 2017, 234). Unmittelbar nach dem Vorfall ersetzte das besagte elegante Hotel die elektrischen Schlösser, die von außen gekehrt werden konnten, und installierte die guten alten, traditionellen Schlösser wieder. Nach einem solchen Vorfall wird der genannte Milliardär wahrscheinlich auch die Internetfunktionen seines Autos für eine Weile einschränken. All dies erzwingt jedoch Entscheidungen in moralischen und rechtlichen Dilemmata und Wahlen, über die es sich im Großen und Ganzen lohnt, nachzudenken. In der Welt unserer Objekte, die im Internet der Dinge immer weiter verbreitet wird, werden die alten einfachen Dinge bereits fallen gelassen,

und wir werden die Objekte, die wir in die Cloud-Datenbank wiken, nicht nach Belieben ersetzen können. So wie wir heute trotz aller negativen Aspekte, die Verwundbarkeit mit sich bringen, das Internet nicht aufgeben würden.

Ein solches Dilemma des Internets der Dinge, die in vernetzten und umfassenden Cloud-Datenbanken eingebunden sind, besteht darin, welcher der widersprüchlichen Anforderungen der beiden Sicherheitsarten – physische Sicherheit und Informationssicherheit – Vorrang eingeräumt werden sollte. Zum Beispiel kann es bis zu einem gewissen Grad wichtig sein, die kluge Wohnung einer älteren Person, die kaum in der Lage ist, sich zu bewegen, durch Kameras und Mikrofone vollständig fernüberwachbar durch medizinische Zentren zu machen, aber es kann auch bedeuten, die intimsten Manifestationen des Lebens aufzudecken, die über das Notwendige hinausgehen. Wenn der Schwerpunkt auf Informationsautonomie und eingeschränkter Beobachtung und Transparenz liegt, werden die Informationen, die in seltenen Fällen noch benötigt werden, möglicherweise nicht an das Pflegezentrum weitergeleitet, und der ältere Pflegebedürftige kann sterben. Henschke weist darauf hin, dass es häufig typische Prioritäten gibt, und zum Beispiel bei einem Smart-TV hat die Informationssicherheit eine höhere Priorität, und zu diesem Zweck können wir hier leicht auf die Einschränkungen eingehen. Bei Tausenden von Anwendungen von selbstfahrendem Auto, das an die Cloud-Software gebunden ist, achten wir jedoch stärker auf die Anforderungen der physischen Sicherheit und nur sekundär auf die Anforderungen der Informationssicherheit (Henschke, 2017, 239).

#### *V. Selbstlernen, maschinelles Lernen und Verantwortung*

Wie oben erwähnt, wird das Hauptproblem der Zukunft das Dilemma selbstfahrender Autos sein, die von Menschen losgelöst sind und in bestimmten, unerwarteten Situationen nicht mehr von außen blockiert werden können, wenn der auf neuronales Selbstlernen aufgebauten Algorithmus dieses selbstfahrenden Autos schon autonom entschieden hat. Da dies die Hauptrichtung der Entwicklung der künstlichen Intelligenz in den letzten Jahren ist, ist es fast sicher, dass dies auch in diesem Bereich nicht umgangen wird. Daher lohnt es sich heute, die Dilemmata der moralischen und rechtlichen Verantwortung von Robotern mit einem hohen Maß an neuronalem Selbstlernen und ihren Machern, Besitzern und Benutzern genauer zu betrachten. Diese Frage wird in ihrer gemeinsamen Studie von *Trevor N. White* und *Seth D. Baum* (White, Baum, 2017, 66-79) sowie von *Shannon Vallor* und *George A. Bekey* (Valor, Bekey, 2017, 338-353) aus verschiedenen Blickwinkeln analysiert.

Trevor und Baums Studie berücksichtigt nicht nur Designer, Bauherren und Benutzer, sondern berücksichtigt auch die „Bestrafung“ des Roboters selbst im Falle der fortschrittlichen Roboter, in deren Programm bereits ein Straf- und Belohnungssystem integriert ist, und wiederholte Strafen und Belohnungen verstärken in ihrer Programmierung die Entscheidungsrichtungen (positiv oder negativ) in Bezug auf die Auswahl zukünftiger Roboterreaktionen. Damit wird auch Bestrafung / Belohnung in den Lernalgorithmus integriert. Wenn sich die Situation in der Zukunft ergibt, werden die Entscheidungen des Roboters in die richtige Richtung gefördert, und der Roboter muss dafür kein Bewusstsein und Selbstbewusstsein haben. Diese Art, das Selbstlernen durch Wiederholungen zu verstärken, ist laut den Autoren auch akzeptabel: „Non-conscious robots could conceivably be punished with some sort of reduced reward or utility as per whatever reward/utility function they might have. Specifically, they could be reprogrammed, deactivated, or destroyed or put in what is known as a „Box“: digital solitary confinement restricting an AI’s ability to communicate or function. (...) To make this possible, however, such robots ought to be based (at least in part) on

reinforcement learning or similar computing paradigms (except ones based on neural network algorithms.)” (Trevor, Baum, 2017, 71).

Das neuronale Lernsystem wird von den Autoren jedoch so beurteilen, dass Designer und Programmierer bereits die Kontrolle über die Reaktion des Roboters auf eine bestimmte Situation verlieren und deshalb sollten von Anfang an als potenzielle Gefahrenquelle, möglicherweise als Katastrophe verboten werden: „Designer could be similarly liable for building robots using opaque algorithms, such as neural networks and related deep-learning methods, in which it is difficult to predict whether the robot will cause harm.” (ebd.) Bei Algorithmen, die solche undurchsichtigen Roboterverhalten zulassen, ist es nicht mehr genug, die nachträgliche Haftung vorzuschreiben, sondern das vorbestellte Verbot ist das zweckmäßig: „Hence, instead of liability, a precautionary approach could be used. This would set a default policy of disallowing any activity with any remote chance of causing catastrophe. (...) In effect, people would be held liable not for causing catastrophe but taking actions that could cause catastrophe.” (Trevor, Baum, 2017, 74). Wenn man sich im Prinzip mit den Autoren über den gefährlichen Charakter neuronaler Deep-Learning-Softwaremechanismen einig ist, muss nur wiederholt werden, dass dies ein Verbot des Hauptwegs zur Entwicklung künstlicher Intelligenz darstellt und deshalb sollte dies im Licht der Mächte hinter der Industrie, Militär usw. als unwahrscheinlich angesehen werden. Daher erscheint es ratsam, nach anderen Pfaden zu suchen, die versuchen, andere Lösung zu finden, ohne das neuronale Tiefenlernen zu verbieten.

Es sollte natürlich betont werden, dass das Lernen neuronaler Netze, von dem das Funktionieren des Zentralnervensystems nachgeahmt wird, durch Einbeziehung externer menschlichen Kontrolle gesteuert werden kann, bevor ein realitätsbildender Effekt ausgelöst werden kann. Dies bleibt jedoch aus einer Reihe von Gründen zunehmend zurück, und dies wird von *Vallor* und *Bekey* in der bereits zitierten Studie analysiert. Ein Grund dafür ist, dass der Vorteil der Verwendung künstlicher Intelligenz anstelle des Menschen, die unglaublich schnell Reaktionsfähigkeit verloren gehen würde, wenn eine retrospektive menschliche Kontrolle eingeführt würde. Darüber hinaus sind in neunundneunzig Prozent der Fälle die Reaktionen korrekt, die um ein Vielfaches höher sind als die Leistung des Menschen. Darüber hinaus kann die Qualität einer viel langsameren menschlichen Kontrolle fraglich sein, da die Entscheidung des Roboters richtiger als die übergeordnete menschliche Entscheidung sein kann. Letzteres geschah mit dem medikamentösen Diagnosealgorithmus von IBM Watson, und die ungewöhnliche Heilung, die durch künstliche Intelligenz aus Millionen von Onkologiestudien und -diagnosen hervorgehoben und später von ihnen synthetisiert wurde, erwies sich korrekter als die Onkologieentscheidung, die sie außer Kraft setzt: „Watson’s diagnoses and treatment plans are still vetted by licensed oncologists. Still, how reliably can a human expert distinguish between a novel, unexpected treatment recommendation by Watson that might save a patient’ life – something that has reportedly already happened in Japan – and the oncological equivalent of „Toronto?” (Vallor, Bekey 2017, 343).<sup>4</sup> Das Dilemma, an Geschwindigkeit zu verlieren und damit den Vorteil des Roboters zu beseitigen, wird auch durch die Robotersoldaten und die Entscheidungssoftware veranschaulicht, die in Kriegssituationen eingesetzt werden. Hier stellt sich ständig die Frage, ob der Robotersoldat, der das gefährlichste Gebiet und Gebäude betritt, mit den in seinem Besitz befindlichen zerstörerischen Waffen ohne eine externe menschliche Entscheidung die darin befindlichen zerstören kann oder stattdessen die Anordnung der Zerstörung nur durch entfernte menschliche

---

<sup>4</sup> „Toronto Mistake” war einer von Watsons Grundfehlern in einem landesweiten TV-Quiz, als er alle Teilnehmer mit seinen Antworten auf die schwierigsten Fragen besiegte. Als letzter Fehler machte er Toronto zu den US-Städten und nicht einmal zum schwächsten Konkurrenten wäre verloren gegangen. So ist es zu einem Symbol für Fehlentscheidungen geworden, die von künstlicher Intelligenz getroffen werden, was selten ist, aber in vielen Fällen Tragödien verursacht.

Bestätigung erteilt werden darf. Auf die gleiche Weise taucht das Dilemma auf, ob ein Aufklärungsflugzeug sofort vom Roboterflugzeug zerstört werden könnte, oder dies nur mit menschlichem Eingreifen aus dem entfernten Kommandoraum auf der Grundlage der von ihm übertragenen Informationen durchgeführt werden könnte? Der Zwang zur Geschwindigkeit erfordert, dass der Roboter selbst eine Entscheidung trifft und ausführt, da der Zeitverlust von einer externen menschlichen Entscheidung zur Vernichtung der Roboter führen kann, wenn er an einen gefährlichen Ort einbricht. Aber das Abschießen freundlicher Kampfmaschinen, die bereits mehrmals passiert sind, oder das Töten von Kindern und Frauen, die fälschlicherweise als Feinde in dem eingedrungenen Gebiet identifiziert wurden, spricht dagegen (Vallor, Bekey, 2017, 349).

Detaillierte Lernalgorithmen für neuronale Netze, der aktuellste Trend in der künstlichen Intelligenz, bieten selbstlernender Software bereits eine vielfältige Tiefe für die einfache komputerische Startposition, indem kontinuierlich immer Milliarden von Daten eingespeist werden und die Startposition sich auf diese Weise allmählich zum fortschrittlichsten machen. Bei dieser Technik werden zwischen den Eingaben der selbstlernenden Software und den aufgabenspezialisierten Ausgaben Tausende von neuronalen Zwischenschichten zwischen den massiven Daten gefunden, wobei die Muster und Regelmäßigkeiten unabhängig voneinander gefunden und zur Verwendung hervorgehoben werden. Durch die Kombination von Milliarden von Daten können sie die kleinsten Regelmäßigkeiten hervorheben und verwenden, die für Menschen bei ihren Entscheidungen nicht wahrnehmbar sind: „Between the input and output node layers are „hidden” layers of nodes that function to process the input data, for example, by extracting features that are especially relevant to the desired outputs. Connections between the nodes have numerical „weights” that can be modified with the help of a learning algorithm; the algorithm allows the network to be „trained” with each new input pattern until the network is optimized. (...) The interest in neural network has grown in recent years with the addition of more hidden layers giving *depth* such network, as well as feedback or recurrent layers. The adjustment of the connections strengths in these more complex networks belongs to a loosely defined group of techniques known as deep learning.” (Vallor, Bekey, 2017 341). Die Auswirkungen von Entscheidungsmustern, die durch diese detaillierten Lernalgorithmen hervorgehoben werden, sind zwar in der Praxis oft überraschend gute Ergebnisse erreichen, können jedoch von Designern und Programmierern nicht verstanden werden, und ihre Entscheidungen können durchweg Überraschungen verursachen, einschließlich unterschiedlich starker unangenehmer Überraschungen. Wer sollte die rechtliche und moralische Verantwortung dafür tragen?

## *VI. Identität in der Welt der künstlichen Intelligenz*

*James DiGiovanna* wirft zum Nachdenken anregende Fragen auf, wenn er in seiner Studie fragt, wie sich die Identität von Menschen mit Gehirnimplantaten ändern kann. Er diskutiert dieses Thema gemeinsam mit der Frage der Identität von Robotern, die in Zukunft als vollständig künstliche Wesen auftreten können und bereits Selbstbewusstsein haben (DiGiovanna, 2017, 307-321). Betrachten wir die beiden Problembereiche getrennt.

Die Möglichkeit eines Gedächtnisses, das durch Gehirnimplantaten ergänzt wird, wurde in den letzten Jahren in Mäuseexperimenten entwickelt und hat sich als wirksam erwiesen. Alle hoffen, die Auswirkungen der sich schnell ausbreitenden Alzheimer-Krankheit in alternden Gesellschaften zu mildern und zu heilen (siehe Kaku, 2014, 132-133). DiGiovanna untersucht die Möglichkeit weiterer Entwicklungen in den nächsten Jahren, die sich zusätzlich zu Krankheiten massenhaft ausbreiten könnten, um die Gehirnkapazität zu erhöhen. Und wenn eine technische Lösung für die Probleme gefunden wird, die heute noch auf diesem Gebiet

bestehen, ist es fast sicher, dass dies zuerst in der Elite, dann in der Gesellschaft als Ganzes üblich wird, um den größten Wert der menschlichen Intelligenz zu steigern. Dies bedeutet jedoch, dass die dauerhafte Identität jedes Einzelnen, die die Grundlage für Kontakte in Gemeinschaften ist, mehr oder weniger aufgehoben werden kann und es ungewiss werden kann, wie sehr wir von unseren Partnern erwarten können, dass ihre Eigenschaften überleben, die wir bisher kennen und lieben: „The ability to rewrite mental content such as ethical values, the capacity for empathy, and general personality traits undermines personhood. (...) A para-person that could experiment with worldviews, completely adopting and deleting values systems, preferences, and bases for judgement, would be largely lacking in what is commonly understood as the most basic element of personal identity” (DiGiovanna, 2017 311) Dies war die Grundlage unserer Wahl im Falle unserer Freunde, Frau und Freundin, aber auf die gleiche Weise basiert unsere engere menschliche Beziehung mit einiger unserer Arbeitskollegen auf die Liebe ihrer Eigenschaften, während die Beziehung zu anderen nur kalt kollegial ist. Daher basiert unser Leben in der Gesellschaft und in verschiedenen kleinen Gemeinschaften innerhalb unserer Gesellschaft auf unseren dauerhaften Identitäten, und dies kann sich nach dem Hinzufügen von Herz-, Hör- und anderen körperlichen Verbesserungen grundlegend ändern, wenn sich das Gehirn mit Gehirnimplantat verändert wird.

Mit den allmählichen Veränderungen im eigenen Leben wird das eigene Bewusstsein im Detail immer umgebaut, und dies führt zu kleinen allmählichen Veränderungen in der eigenen Identität, die in der modernen Welt durch die Informationserweiterungen des letzten Jahrhunderts noch verstärkt werden. Im Vergleich dazu werden wir in der Zukunft jedoch in der Lage sein, eine ganze Reihe von Informationen – den Inhalt von Büchern und Studien, kleinere Bibliotheken – mit Gehirnimplantat in unseren Kopf zu pflanzen und gemeinsam damit neuen grundlegenden logischen und wertverarbeitenden Mechanismen zu handhaben, die wir bisher in unserem Leben nicht hatten, und wir hatten auch nicht die Fähigkeiten dazu. Nun, dies wird den Kontakt zwischen dem Individuum und seinen Gemeinschaften grundlegend beeinflussen. In jedem Fall könnten die Kontaktgrundlagen, die auf den heutigen dauerhaften Identitäten beruhen, durch diese Änderung beseitigt werden. Nach solch einem neuen Bewusstseinsinhalt – insbesondere, wenn die Werten der Kontaktpartner ergänzt und neu eingestuft wurden – kann ich nicht wissen, inwieweit mein Freund, meine Freundin, meine Frau, mein Kollege usw. gleich sind. Ob die Eigenschaften, die wir bisher in ihnen geliebt haben, sind noch lebendig, oder auf die gleiche Weise die Erfahrungen, die wir bisher zusammengesammelt haben und die in unserer engen Beziehung auch ohne Worte dieselbe Antwort lieferten, für ihn immer noch relevant sind. Dies kann nur durch die Möglichkeit verschärft werden, dass die Gehirnimplantanten in dem Gehirn, die unser Wissen, unsere Normen und logischen Fähigkeiten ergänzen, die mit Hilfe unseres biologischen Gehirns erworben wurden, von außen ständig neu und neu aktualisiert werden, wie wir es heute bereits wissen. Darüber hinaus können sie sich ständig mit den in den Clouds gespeicherten Informationsbasen ihrer Software verbinden. Inwieweit wird unser Freund, der mit solchen ausgestattet ist, vertraut bleiben, auf den wir uns verlassen können, denn "doch wir kennen ihn!"?!

Diese Identitätsfrage erstreckt sich auch auf rechtliche und moralische Probleme. Inwieweit kann ich jemanden für sein früheres Verhalten respektieren oder einfach nur verachten, da er nach einer Gehirnauffrischung entweder ein „moralischer Athlet“ oder nur ein kalter Vorteilsüchtiger sein kann. Oder macht die rechtliche Verantwortung für die gestrige Handlung für jemanden, der seitdem anders denkt und handelt, Sinn? Die andere Seite davon ist, ob, wenn wir das Bewusstsein eines Soziopathen durch Gehirnimplantat ändern können und sein Bewusstsein teilweise gelöscht und ein neues sozialfreundliches Bewusstsein eingeführt wird, ist ein System der Bestrafung immer noch erforderlich? Und dies wirft die Frage auf, ob neben freiwilligem Gehirnimplantat auch deren Zwangsinstallation akzeptabel ist? Oder kann

es zum Teil vom Staat in dieser Hinsicht in der Kindheit eine obligatorische Untersuchung und Prüfung für alle Kinder obligatorisch gemacht werden, wie dies derzeit bei obligatorischen Impfungen der Fall ist? DiGiovanna nennt *Para-Personen* die zukünftigen Menschen mit einem so ergänzten Gehirn – unter Vermeidung des Namens Cyborg, der bereits in Science-Fiction dafür erfunden wurde – und angesichts des aktuellen Standes der Laborforschung bedeutet diese Zukunft überhaupt keine ferne Zukunft, und die Wahrscheinlichkeit ihrer Verwirklichung ist hoch. Der Umgang mit den rechtlichen und moralischen Dilemmata und die Anpassung der heutigen Lösungen an die damalige Situation erfordern daher umfassende Überlegungen.

Über Para-Personen hinaus kann bei vollständig künstlichen und im Gegensatz zu heute selbstbewussten Roboterwesen, deren Wahrscheinlichkeit nicht ausgeschlossen werden kann, auch wenn sie nicht so groß ist wie die erstere, die Frage der Identität durch das Aufwerfen neuer Aspekte angegangen werden. DiGiovanna stellt der Inhalt der Identität im Zentrum, um das Dilemma der Identität in Bezug auf die Roboterwesen zu exponieren. Einige Details des Menschen und seines Bewusstseins ändern sich ständig, aber seine dauerhaften Eigenschaften und Wertvorlieben verändern sich auch über viele Jahre hinweg nur geringfügig, sodass man denjenigen, die immer in seiner Umgebung lebt, mehr oder weniger eine Identität zuschreiben kann, die auch die Veränderungen umfasst. Es ist die Langsamkeit des Wandels, die es mir ermöglicht, auch in der heutigen schnelllebigen Welt nicht von meinen bisherigen Erfahrungen mit den Motivationen und Merkmalen derer enttäuscht zu werden, die mit mir in Kontakt kommen. Aber genau das verschwindet für Roboter, die in kürzester Zeit tausend- und millionenfach schneller als Menschen in der Informationsverarbeitung und im Selbstlernen sind: „Slow change of character and appearance is part of what makes personal identity (...) But with an artificial person, sudden and radical change in both the physical and mental becomes possible“ (DiGiovanna 2017, 3011, 307.)

Dauerhafte Wertpräferenzen in der Informationsverarbeitung und die darauf basierende Zusammenarbeit sind bei Robotern bereits durch die Masse und Geschwindigkeit ihrer Informationsbeschaffung sowie durch ihr ständiges Selbstlernen und Selbstverändern problematisch. DiGovannas Vorschlag könnte auch bedeuten, dass die Möglichkeit von „Selbstbewusstsein“ und „Ich-bewusstsein“ selbst für einen zukünftigen leistungsstarken MI-Roboter überdacht werden muss. Diese setzen die dauerhafte Identität einer Person voraus, aber dies beruht auf der Langsamkeit unserer Bewusstseinsveränderungen und damit auf der Beständigkeit unserer Informationsverarbeitung. Wenn ein künstliches Wesen von der menschlichen Kontrolle befreit und auf unabhängige Informationsverarbeitung und von dieser auf Selbstlernen und Selbstveränderung umgestellt wird, kann es tausendmal am Tag, jede Stunde und sogar jede Minute lernen und kann sich in seinen immer kürzeren neuen Zyklen verwandeln, dann verschwindet das, was wir im heutigen Menschen ein stabiles Selbstbewusstsein, Ich-bewusstsein nennen, fast. Mit dieser Betonung fügt DiGiovanna der viel diskutierten Frage auch eine neue Frage hinzu, d. H. wie die Frage nach dem Selbstbewusstsein und dem Ich-bewusstsein des zukünftigen fortgeschrittenen Roboterbewusstseins stehen wird. Und wie kann man sich moralische Wertvorstellung ohne dauerhaftes Selbstbewusstsein und Ich-bewusstsein vorstellen?

Allein aus diesem Grund muss der Gedankengang, der durch die mechanische Erweiterung des heutigen menschlichen Bildes berechnet, dass ein solches Roboterwesen im Fall von Superintelligenz wahrscheinlich auch „überethisch“ sein wird, als falsch angesehen werden (siehe Petersen 2017). In diesem Zusammenhang ist es aber auch notwendig, die Erklärungen und Analysen, die im Falle der Entwicklung von Robotern mit eigenem Bewusstsein – in menschlicher Analogie – die Anerkennung ihrer moralischen Bedürfnisse und die Gewährung von Menschenrechten in ihren Schriften vorsehen, umfassender zu behandeln. Denn fassen diese Analysen die zukünftigen Roboter einfach als eine neue Art menschlichen Begleiters und als Erweiterung der menschlichen Existenz auf. Sobald die Programme der

Roboter die Emotionen in ihre Algorithmen aufgenommen haben, fordern diese Analysen, dass die Gesellschaft auf die Emotionen von Robotern achtet und ihnen auch Menschenrechte gewährt: „Es muss wahrscheinlich gesetzlich geregelt werden, wie viel Schmerz und Gefahr ein Roboter ausgesetzt sein kann. (...) Es könnte leicht sein, dass dies zu weiteren ethischen Debatten über andere Rechte von Robotern führen würde. Können Roboter Eigentum besitzen? Was passiert, wenn jemand versehentlich verletzt wird? Können sie verklagt oder bestraft werden? Wer ist im Falle einer Klage für sie verantwortlich? Kann ein Roboter einen anderen Roboter besitzen? Aus solchen Fragen ergibt sich dann eine andere Frage: Sollten Roboter einen ethischen Sinn erhalten? (Kaku 2014, 251). Unsere vorherigen Erklärungen beantworteten mehrere Fragen aus diesen, basierend auf den in der Zwischenzeit durchgeführten Robotikstudien, aber das Grundproblem dahinter sollte auch hervorgehoben werden, da ganze Studien und Bände aus ähnlichen Annahmen entstanden sind, z. B. ein neuer Band in diesem Bereich, herausgegeben von Jason P. Doherty: *AI Civil Rights: Addressing Artificial Intelligence and Robot Rights*.

Nun, das Grundproblem bei diesem Gedankengang besteht darin, dass er die Tatsache ignoriert, dass Rechte und ethische Anforderungen bei Robotern nur entstehen können, wenn Ich-Bewusstsein und Selbstbewusstsein geschaffen werden. Es bedeutet aber auch, dass sie, wenn dies wirklich in der Zukunft geschieht, gleichzeitig durch die tausendfache Entwicklung von der menschlichen Kontrolle befreit und als separate neue Seinsschicht über die bisherige vier Seinsschicht der menschlichen Gesellschaften aufgebaut werden. Von dieser Zeit an wären aber ihnen die gesamte biologische Sphäre und die damit verbundenen menschlichen Gesellschaften gleichgültig und würden keinen „gerichtlichen Schutz“ benötigen. Das heißt, eine Roboterwelt, die dieses Niveau erreicht hat, wäre nicht Teil der menschlichen Gesellschaft als „neuer Kamerad“ in der Herrschaft über die Welt, sondern als die menschliche Existenz aus der Welt der Primaten hervorging und sich über die tierbiologische Seinsschicht erhob und immer autonomer wurde, so erhebt sich nun der künstliche Maschinenintellekt, losgelöst von biologischen Voraussetzungen, über die menschliche Gesellschaft. Im Gegensatz zum bisherigen Aufbau immer neuerer Seinsschichten auf den unteren Seinsschichten würde die neue Seinsschicht der künstlichen Intelligenz nur die unterste physischen Seinschicht brauchen, und für sie wäre die biologische und psychische-emotionale Seinschichten nicht erforderlich. Diese Roboterwesen würden keine Rechte und ethischen Forderungen benötigen, sondern sie werden die gesamte Realität, einschließlich der menschlichen Gesellschaften, dominieren, wie wir Menschen heute die vierschichtige irdische Welt dominieren. Auf diese Weise haben eher solche Analysen Recht, die diskutieren, ob, wenn die Roboterwelt wirklich dieses Niveau erreicht, was mit der Menschheit passieren wird?!

#### *VII. Das moralische Credo von „Unabomber“, dem „verrückten Mathematiker“*

Mitte der neunziger Jahre, nach vielen Jahren einer Reihe von Bombenanschlägen und einer Verfolgungsjagd des FBI, gab ein geheimer Täter namens „Unabomber“ den Grund für sein Vorgehen in einer anderthalbhundertseitigen Broschüre an, die er gegen die Unmenschlichkeit der Entwicklung der technologischen Gesellschaft seit der industriellen Revolution aussprach. Seine eigentümliche Sprache wurde von seinem Bruder erkannt und durch Benachrichtigung des FBI wurde der lang ersehnte Bomber gefangen genommen. Es stellte sich heraus, dass *Theodor John Kaczynski*, ein Mathematiker aus Harvard. Er wurde zu einem Zeitpunkt seiner Universitätskarriere zum Feind einer von Technologie geprägten Gesellschaft und begann seine Explosionsserie, in der die Entwickler und Hauptnutzer dieser Technologie gezielt wurden. Mehrere starben und weitere wurden dabei verwundet, und er plante, noch mehr Vergeltungsmaßnahmen durchzuführen, wenn sie nicht verhaftet worden

wären. Nachdem die exponentiell sich entwickelnden Auswirkungen der Technologieentwicklung in den letzten dreißig Jahren wirklich unbestritten geworden sind und das Ausmaß und die Auswirkungen ihrer weiteren Beschleunigung bereits Gegenstand mehrerer umfassenden Analysen sind, lohnt es sich, sich erneut auf Argumente von Unabomber, dem „verrückten Mathematiker“ zu konzentrieren. Dies ist, was *Jai Galliot* in seiner neuen Studie tut. Er setzt den Widerstandskämpfer, der seitdem in seiner Gefängniszelle damit beschäftigt ist, seine Thesen weitzuentwickeln, unter Theoretikern und Bewegungen der Antitechnologie und es wird versucht, seine Hauptthesen im Lichte des aktuellen Zustands der Roboterwelt hervorzuheben (Galliot, 2017, 369-385).

Kaczynski hat nur die praktischen Schlussfolgerungen der früheren Thesen von *Jacques Ellul*s 1964 erschienenem Band *"The Technological Society"* gezogen, die auf ihre Weise auch eine Fortsetzung von Oswald Spenglers 1922er Arbeit zur Analyse des Niedergangs der westlichen Zivilisation darstellten. Von beiden Verfassern wurde der Niedergang mit der technologischen Entwicklung erklärt (Spengler, 1995). Der rein pessimistische und resignierte Ton bei Spengler und Ellul wurde dann zu einem moralischen Widerstand im Falle von Kaczynski, und nachdem er gesehen hatte, dass es keine Möglichkeit gab, diese Entwicklung zu reformieren, glaubte er, dass nur revolutionäre Gewalt übrig blieb, um die Zerstörung der Menschheit zu verhindern. Jahrzehnte nach seiner Broschüre lohnt es sich nun zu überdenken, wie der aktuelle Zustand der Roboterwelt und die radikaleren Veränderungen, die bereits weitgehend sichtbar sind, die Gefährdung der Menschheit oder zumindest eine erhebliche Verschlechterung ihres Zustands bedeuten könnten.

Als Ausgangspunkt für ihr Vorgehen gegen die technologische Gesellschaft ist hervorzuheben, dass sowohl Spengler als auch Ellul und Kaczynski die menschliche Existenz als eingebettet in die physikalisch-biologische Umwelt betrachten. Daraus schließen sie dann, dass die menschliche Existenz zerstört wird, wenn das menschliche Leben infolge der industriellen Revolution immer mehr technologisch vermittelt wird und auf diese Weise von der physikalisch-biologischen Umwelt immer weiter entfernt wird: „Ellul wrote that „the machine trends not only to create a new human environment, but also to modify man’s very essence” and that „the milieu in which he lives is no longer his. He must adopt himself, as though the world were new, to a universe for which he was not created. Kaczynski shares this sentiment.” (Galliot 2017, 373). Wenn wir andererseits Hartmanns These berücksichtigen, die die vier voneinander abhängigen Seinsschichten des menschlichen Lebens im Auge behält (physisch, biologisch, mental und intellektuell) und die im Verlauf der Evolution von der immer stärkeren transformierenden Wirkung der oberen Schichten auf die unteren Schichten ausgeht, ist die obige These übertrieben und ohne Grund und sie muss als zu pessimistisch eingestuft werden.

Kaczynski und seine Vorgänger sehen es als Zerfall des menschlichen Lebens an, wenn die oberste intellektuelle Schicht der vier Seinsschicht gegenüber den unteren immer dominanter wird. Dies war jedoch der Fall, wenn auch langsamer, in den letzten zwei- oder dreitausend Jahren und man kann die Benutzung der Metalle und insbesondere der Eisen herausheben, von der die Transformation der menschlichen Umwelt grundlegend verändert wurde. Die industrielle Revolution hat dies nur beschleunigt, und insbesondere seit den 1950er Jahren ist es stürmisch geworden, die verschiedenen Aktivitäten menschlicher Gemeinschaften auf Intelligenz und die damit verbundene Technologie zu stützen. Das heißt, menschliches Leben basiert keineswegs nur auf der physikalisch-biologischen Seinsschichten. Wenn also ihr Anteil und ihre entscheidende Kraft im menschlichen Leben abnehmen und diese Umgebung technologisch weit vermittelt und transformiert wird, bedeutet dies nicht, dass die menschliche Gesellschaft zerstört wird. Bei alledem verschiebt sich nur das Gewicht der Bedeutung der vier Seinsschichten der Realität, wodurch sich das menschliche Leben stärker auf der Grundlage der intellektuellen Schicht basiert und die Dominanz dieser geistlichen Seinsschicht über den unteren radikal erhöht wird. Diese Einschätzung von uns könnte nur ausgesetzt werden, wenn



diese Welt irgendwann in der Entwicklung der Roboterwelt wirklich aus menschlicher Kontrolle hervorgehen würde und sich die künstliche Intelligenz als eine neue Seinsschicht über die menschlichen Gesellschaften erheben würde, die sich bisher auf dem Höhepunkt der Evolution befanden. Die Erhöhung von Kaczynski als Prophet würde dann nur dadurch verhindert, dass unter solchen Umständen und seinen Gefahren das Versäumnis, als Held erhöht zu werden, das geringste Problem wäre. Dies kann jedoch nach bestem Wissen und Gewissen heute nur als unwahrscheinliche Option angesehen werden, und vielmehr kann die Zunahme der von künstlicher Intelligenz geprägten menschlichen Gesellschaften ohne eine neue autonome Seinsschicht als realistische Vision für die Zukunft angesehen werden.

### *VIII. Literature*

- Abney, Keith, „Robotics, Ethical Theory and Metaethics: A Guide for the Perplexed, in Patrick Lin, Keith Abney, George A. Bekey (eds.), Robotethics, The MIT Press, Cambridge Massachusetts. London 2011. pp. 35-54.
- Allen, Collin, Wendell Wallach, „Moral Machines: Contradiction in Term or Abdication of Human Responsibility?“, in Patrick Lin, Keith Abney, George A. Bekey (eds.), Robotethics, The MIT Press, Cambridge Massachusetts. London 2011. pp. 55-68.
- DiGiovanna, James, „Artificial Identity“, in Patrick Lin, Ryan Jenkins, Keith Abney (eds.), Robot Ethics 2.0, Oxford University Press, New York 2017. pp. 307-321.
- Doherty, Jason P. (ed.), AI Civil Rights: Addressing Artificial Intelligence and Robot Rights. Kindle Edition, 2016.
- Ford, Martin, The Rise of Robots: Technology and the Threat of a Jobless Future, Basic Books, 2016.
- Galliot, Jai, „The Unabomber on Robots“, in Patrick Lin, Ryan Jenkins, Keith Abney (eds.), Robot Ethics 2.0, Oxford University Press, New York 2017. pp. 369-385.
- Hartmann, Nicolai, „Das Problem des geistigen Seins. Zur Grundlegung der Geschichtsphilosophie und der Geisteswissenschaften. Walter de Gruyter Verlag. Berlin 1962.
- Hegel, Georg Wilhelm Friedrich, „Grundlinie der Philosophie des Rechts. (Werke Band 7.) Suhkamp, Frankfurt am Main. 1979. S. 301.).
- Henschke, Adam, „The Internet of Things and Dual Layers of Ethical Concern“ in Patrick Lin, Ryan Jenkins, Keith Abney (eds.), Robot Ethics 2.0, Oxford University Press, New York 2017. pp. 229-243.
- Kaku, Michio (2014): The future of the mind. The Scietific quest to understand, enhance and empower the mind. Doubleday. New York. 2014.
- Kelly, Kevin, The Inevitable: Understanding the 12 Technological Forces That Shape Our Future, Penguin Books, 2014.
- Klinewicz, Michal, „Challenges to Engineering Moral Reasoners“ in Patrick Lin, Ryan Jenkins, Keith Abney (eds.), Robot Ethics 2.0, Oxford University Press, New York 2017. pp. 244-257.
- Loh, Wulf, Janina Loh, „Autonomy and Responsibility in Hybrid System, in Patrick Lin, Ryan Jenkins, Keith Abney (eds.), Robot Ethics 2.0, Oxford University Press, New York 2017. pp. 35-50.
- Luhmann, Niklas, Liebe als Passion: Zur Codierung von Intimität, Suhrkamp, Frankfurt am Main 1994.
- Pokol, Béla, Theoretische Soziologie und Rechtstheorie. Kritik und Korrigierung der Theorie von Niklas Luhmann. Passau. Schenk Verlag. 2013.
- Pokol Béla, Künstliche Intelligenz: Die Entstehung einer neuen Seinschicht? (Ki – im Spiegel von Nicolai Hartmanns Ontologie.) Pázmány Law Working Papers No.2018/12
- Splengler, Oswald, Der Untergang des Abendlandes. Albatros Verlag/Patmos Verlag. 2007.

Talbot, Brian, Ryan Jenkins, Duncan Purves, „When Robots Should Do the Wrong Thing”, in Patrick Lin, Ryan Jenkins, Keith Abney (eds.), Robot Ethics 2.0, Oxford University Press, New York 2017. pp.258 -273.

Vallor, Shannon, George A. Bekey, „Artificial Intelligence and the Ethics of Self-Learning Robots”, in Patrick Lin, Ryan Jenkins, Keith Abney (eds.), Robot Ethics 2.0, Oxford University Press, New York 2017. pp. 338-353.

White, Trevor N, Seth D. Baum, „Liability for Present and Future Robotics Technology”, in Patrick Lin, Ryan Jenkins, Keith Abney (eds.), Robot Ethics 2.0, Oxford University Press, New York 2017. pp. 66-79.

Zoller, David, „Skilled Perception, Authenticity, and the Case Against Automation”, in Patrick Lin, Ryan Jenkins, Keith Abney (eds.), Robot Ethics 2.0, Oxford University Press, New York 2017. pp. 55-68.