

CLA-アーカイブズ 5

慶應義塾におけるドイツ年記念公開セミナー

敵か味方か

——ロボットをめぐる文化——

慶應義塾大学教養研究センター

Keio Research Center for the Liberal Arts

慶應義塾大学教養研究センター主催  
慶應義塾におけるドイツ年記念公開セミナー

敵か味方か  
——ロボットをめぐる文化——

2005 年 11 月 11 日(金) 18:15 ～ 20:30  
慶應義塾大学日吉キャンパス J14 番教室にて

Program

- |         |  |          |
|---------|--|----------|
| 18:15 ～ | あいさつ   | 識名章喜     |
| 18:30 ～ | 基調講演「ロボットとコンピューターは敵になるか？」<br>——アシモフのロボット作品からドイツ SF 文学まで——<br>ハンス・エッセルボルン |          |
| 19:15 ～ | 講演をうけて   | 巽孝之・前野隆司 |
| 20:00 ～ | 質疑応答   |          |

# はじめに

慶應義塾大学商学部助教授 石原あえか

本公開セミナーは、ケルン大学哲学部教授ハンス・エッセルボルン氏が日本学術振興会・海外招聘研究者（短期）に採用が決まったことを受け、企画・開催されたものです（受け入れ研究者は私・石原、研究課題は「国際的ジャンルとしてのサイエンス・フィクション 独語圏および英語圏 SF 文学の多様性と日本における同義的作品の比較」）。氏を日吉で受け入れることで、隣接する矢上の工学研究者とも交流し、現代社会が直面しているテーマ《科学技術と人間》に、自然科学および人文科学の双方から取り組めるのではないかと考えました。また三田の英米文学研究者にも参加していただくことで、専門分野・キャンパスともに領域横断的なセミナーになるよう心がけました。

こうして 2005 年 11 月 11 日（金）、午後 18 時 15 分より日吉 J14 番教室にて、本セミナーが開催されました。まず、ゲストのエッセルボルン氏が約 40 分、アシモフのロボット工学三原則と彼のロボット短編集を手がかりに、『コンピューターとロボットに敵意が生じるか?』というテーマで独文学者としての視点から講演されました。原題は、*Können Roboter und Computer böse sein?* で、ドイツ語の知識がない方のために、予め講演原稿を全訳、PowerPoint を用いて字幕スクリーンに作り変え、講演中背後のスクリーンに投影しました。氏は、H.W. フランケや H. ハウザーなどドイツ SF 作品も紹介しつつ、SF 文学に仮託された問題、すなわち機械による職務遂行・論理的計算の枠組みの中におかれている人間存在について考察しました。

続いて巽孝之氏（文学部）が、SF 作家・柴野拓美による「SF とは人間理性の産物が人間理性の制御を離れて自走することを意識した文学である」という定義を紹介。巽氏は、英米文学者の視点から、W. ギブスンらポスト・ヒューマニズム思想と連動した SF では、むしろ人間こそ「肉人形」＝ロボットとする逆転描写が行われたこと、また E. トムソンの『ヴァーチャル・ガール』や G. イーガンの『ディアスポラ』等、新傾向の現代 SF 作品を紹介し、エッセルボルン氏の講演内容を補足しま



石原あえか氏

した。さらに前野隆司氏（理工学部）がロボット工学者の視点から、「人の意識は解明可能か」、「ロボットの意識は作れるか」という最新トピックについてミニ・レクチャーを行ないました。

会場の参加者を含めた質疑応答は、専門同時通訳（ウィスパー）を 2 名つける万全の配慮をしたので、活発な意見交換が続き、セミナー終了後もなお各講演者に直接質問する学生の姿が見られました。「慶應におけるドイツ年」の一環として行なわれた本セミナーですが、英米文学や工学分野からも参加者を得て、日独文学研究比較に留まらない、ユニークなものとなりました。2007 年に横浜でアジア初の世界 SF 大会開催が決定した矢先でもあり、本テーマについては今後も共同研究の継続が期待されます。

さて、最後に可愛らしいエピソードをひとつ披露しましょう。好奇心旺盛なエッセルボルン氏は、来日前からずいぶん日本の漫画を読んだり、アニメを鑑賞されたり、入念な準備をしていました。来日後は、三田キャンパスで働き始めたロボット・ケイタ君はもちろん、愛知万博終了後、品川の三菱重工ショールームに展示されていたホームロボット WAKAMARU にも早速会いに行きました。実演を事前予約し、「どうぞ名前を呼んで下さい、お返事しますよ」と促されたエッセルボルン氏、精一杯の親しみをこめて「ヴァ・カ・マ・ルー！」と呼び

かけたのです。そう、ドイツ人はワ行の段「わ・ゐ・う・ゑ・を」の発音を持たないのです！ WAKAMARU は、「ヴァカマル」と発音されます。どんなに頑張っても「馬鹿丸」と聞こえる氏の呼びかけに、お利巧な「ワカマル」君は、そのつぶらな瞳をクルッとめぐらし、「それは誰のこと？」と首をかしげるばかりでした。

謝辞：エッセルボルン氏来日を可能にして下さった日本学術振興会に感謝します。また本セミナー開催にあたっては、教養研究センター所長・横山千晶氏、同副所長・岩波敦子氏、そして同センター事務の石川啓子氏、宮坂敦子氏、甲賀崇司氏より多大なご支援をいただきました。受け入れ研究者および企画者として、心からお礼申し上げます。

# あいさつ

慶應義塾大学商学部教授 識名章喜

みなさん、本日は教養研究センター主催公開セミナー「敵か味方か——ロボットをめぐる文化——」にお集まりいただきまして、ありがとうございます。本日の司会を務めます商学部の識名と申します。

私もドイツ文学関係をテーマとして勉強しておりました、SF 関係では、「科学技術と文学」というテーマで勉強をしていました。本日は、そうした関係で司会を務めさせていただくことになりました。

まず、ご講演いただきますケルン大学教授のハンス・エッセルボルン先生をご紹介します。先生は今回、日本学術振興会外国人招聘研究者（慶應義塾商学部では訪問教授扱い）という立場で日本にいらっしゃいました。この試みは、商学部の石原先生が中心となり活動をした結果、実現致しました。エッセルボルン先生は、ケルン大学の独文科で「ジャン・パウル」というロマン派とも言いましょうか、ドイツでも特異な位置を示す作家の研究家で、大きな業績を残されています。今回、そのエッセルボルン先生がSF についてお話をされるというのは、エッセルボルン先生は、ドイツのSF にお詳しく、ドイツ文学の中での研究を進めておられます。ここは注釈が必要なところですが、ドイツ文芸学の世界でSF が表立って議論されることは、なかなかありません。SF ファンの間で議論されることはあっても、大学の機関の中で話をされることはありませんでした。こうした中、エッセルボルン先生はその先駆者として、中心的に進めてこられました。また、先ほど拝見致しましたが、ドイツにも放送大学のようなものがあり、その講座の中でも文学的な科学技術、サイエンスフィクションに関するものがあります。そこでも先生は包括的な教材をお作りになられています。また、大学のゼミでは、ドイツの著名な現役作家をお招きし、学生とともに討論もされています。ここ最近では、ケルン大学の独文科のなかで「SF の夕べ」と称して、若手のSF 作家を朗読の会に招き、交流の輪を深める試みをなさっています。日本はロボットを始めSF に関係する分野が進んでいますので、いろいろな出会いがあるのではないかという期待も寄せられていま

す。先日は、愛知万博で紹介されたロボットを見に行かれたようです。

また、日本のアニメなどのDVD にも関心を寄せられているそうです。

本日、お越しいただいた学生の中で、エッセルボルン先生に伝えたいことがありましたら、ぜひ先生に声をかけて紹介をしてください。

また、今回のセミナーには、エッセルボルン先生の他にコメンテーターとして、慶應義塾大学文学部英米文学専攻の教授でいらっしゃる巽孝之先生にお越しいただきました。先生は日本では数少ない本格的なSF 批評の第一人者です。巽先生からもロボットに関するさまざまなテーマでのお話が伺えるでしょう。また、文学的な話だけでは具体的なイメージがつかめませんので、慶應義塾大学理工学部機械工学科にてロボット工学を専門にしていいらっしゃる前野隆司先生にもお越しいただき、実際にロボットを作られている立場から、可能なことや不可能なことについてお話をさせていただければと思っています。そうした豪華なメンバーを迎えて実り豊かなシンポジウムにしていきたいと思っています。

みなさんも話の内容についてどんどんご意見をお寄せいただければ幸いです。



識名章喜氏

# Können Roboter und Computer böse sein?

## Experimente der Literatur

### 「ロボットとコンピューターは敵になるか？」 ——アシモフのロボット作品からドイツ SF 文学まで——

Hans Esselborn  
ハンス・エッセルボルン  
(ケルン大学教授)

#### A. Möglichkeiten des Bösen bei Robotern und Computern

Die Frage nach der Bosheit von Robotern und Computern scheint zunächst unsinnig zu sein, da es sich dabei um Maschinen handelt, die zu bestimmten Zwecken erfunden und gebaut wurden. Streng genommen kann man also nur fragen, ob die künstlichen Wesen funktional oder dysfunktional sind, z.B. weil sie für eine spezielle Aufgabe nicht geeignet sind oder aufgrund eines Defektes ausfallen.

Dies ist aber die Perspektive des Ingenieurs, die Literatur ist meist an anderen Dingen interessiert. Zwar kann die Science Fiction in futurologischer Perspektive nach neuen Maschinen und ihren Folgen für die Gesellschaft fragen, aber Roboter und Computer werden literarisch vor allem im Feld der Beziehungen von Schöpfer und Geschöpf situiert.

Dabei schwingt immer auch eine religiöse Note mit. Während es im 19. Jahrhundert vor allem um die Hybris der Schaffung nicht natürlicher Wesen geht stehen im 20. Jahrhundert die Aktivität der Geschöpfe und ihr

#### A. ロボットとコンピューターにおける敵意の可能性

「ロボットとコンピューターが敵意を持ちうるか？」という問いは、最初はばかばかしく思われます。何しろそれらは、特定の目的のために発明され、製造された機械なのですから。厳密な意味では、この人工的存在が「機能するか・しないか」を、問いかけることしかできません。たとえば「ある特殊業務に適していない」とか、「故障のため停止」とか。

でもこれはエンジニアの視点であって、文学者は他の事柄に興味を持ちます。確かに SF は、未来学の視点から新しいマシンとそれが社会に及ぼす影響を問いかける事ができますが、ロボットやコンピューターは、文学ではその製作者と被創造物の関係性が語られるという特徴があります。

この場合、宗教的なニュアンスが漂っているのが常です。19 世紀の主要テーマは、人工的存在を創造する傲慢さでした。20 世紀になると被創造物の行動とその人間との関わりがクローズアップされました。機械は人間の命

Verhältnis zum Menschen im Vordergrund. Erfüllen die Maschinen gehorsam ihren Auftrag oder rebellieren sie dagegen? Bleiben sie abhängige technische Sklaven oder werden sie intellektuell überlegene Wesen, welche im Sinne der Evolution die menschliche Rasse beherrschen und ablösen werden?

Bei beiden Alternativen kommt die Frage der Bosheit ins Spiel — natürlich vom Menschen aus gesehen. Die Kategorie des Bösen setzt ein Hinauswachsen über die Existenz als bloßes Werkzeug voraus, eine gewisse Freiheit des Willens und damit Menschenähnlichkeit.

Das Interesse der Autoren und Leser an den literarischen Figuren der Roboter und Computer bezieht sich also weniger auf die mögliche Gefährdung durch die technischen Artefakte als auf ihren Modellcharakter für den Menschen. Grundsätzlich gilt: „Das Böse gehört zum Drama der menschlichen Freiheit. Es ist der Preis der Freiheit.“<sup>1)</sup>

Schon im Vorfeld des menschlichen Handelns kann Böses festgestellt werden: „Das Böse in der Natur ist Chaos, Kontingenz, Entropie, Fressen und Gefressen werden.“ (14, vgl. 306, 324) Diese natürliche Sinnlosigkeit kann für den Menschen zum Ausgangspunkt für Destruktion und Sinnvernichtung werden: „In der Leere draußen im Weltraum, ebenso wie im eigenen Selbst, im schwarzen Loch der Existenz und des Bewußtseins kann man die Grausamkeit, die Zerstörung um ihrer selbst wählen.“ (14)

Beim Menschen wird nach Plato und dem Christentum „der Leib als Ort und Ausgangspunkt“ des Bösen angesehen. Deshalb erscheint es in der Form des Triebes und besonders der Sexualität. Die bloß natürliche Triebhaftigkeit kann sich zudem pervertieren zur Aggression und Grausamkeit, also zu den negativen Trieben, die in der Psychopathologie mit dem Namen de

letztens als „Tendenz zur Zerstörung“ bezeichnet werden. Soll der Mensch die Maschine als Werkzeug betrachten, oder soll er sie als Feind betrachten? Soll er sie als Werkzeug betrachten, oder soll er sie als Feind betrachten? Soll er sie als Werkzeug betrachten, oder soll er sie als Feind betrachten?

これら二者択一の可能性を考える場合、キーワードとなるのが敵意（もちろん人間から見た場合）の所在です。敵意という概念は、単なる機械という存在を超え、一種の自由意志と、それによる人間との類似性があることで成り立ちます。

文学に登場するロボットやコンピューターに対する作者と読者の興味は、技術的な人工物による潜在的な脅威よりも、人間をモデルとした特徴や性格に多く注がれています。基本的に、「敵意は人間的自由のドラマの一部である。敵意は自由の代償である」<sup>1)</sup>

人間的行為よりも前の段階で、すでに敵意の存在が確認されます。「自然界における敵意とは、カオス、偶然性、エントロピー、そして弱肉強食である」。ささいな自然界の出来事が、人間にとっては破壊や正常な感覚の崩壊への出発点となりうるのです。「人類は、はるか遠い宇宙の空間において、自分自身の中にもあるような、存在と意識のブラックホールの中で、残虐な行為を選び、自らを破壊する可能性を持つ」。

人間の場合、プラトンやキリスト教の教義によれば「肉体は、悪が巣くう場所であり、またその起点である」とされています。そのため悪は、衝動・欲求・本能という形で、特に性欲という形で出現します。本能むきだしの衝動的行動は、さらに異常な攻撃性と残虐性、つまり精神病理学において「サド」の名と結びつく負の衝動となり得るのです。



Sades verbunden sind.

In der Aufklärung findet eine Umwertung statt: „Das Böse ist keine eigene Macht, sondern nur ein Mangel. Ein Mangel an Vernunft. Die Natur wirkt gesetzlich, schon darum kann es in ihr nichts Böses geben.“ (177, vgl. 193) Damit wird das Böse im Geist situiert. Es ist nun die Auflehnung gegen Gesetz und Ordnung und kann sich als Böses um seiner selbst willen bis zum Nihilismus radikalisieren.

Die Autonomie der Vernunft setzt das autonome Individuum des Liberalismus voraus. In der Konkurrenzgesellschaft des Kapitalismus zeigt sich eine soziale Form des Bösen, nämlich Egoismus und Rücksichtslosigkeit, Mißachtung des allgemeinen Interesses und des der Anderen. Der natürliche Trieb zur Selbsterhaltung und Selbstentfaltung, der sich auf die Instinkte und die Analogie zu den Tieren berufen kann, kann also sozial schädlich werden, sowohl als verdrängender Egoismus als auch als Machtstreben, das zur Entmündigung, Unterjochung und schließlich Vernichtung der Anderen führen kann.

啓蒙主義において、敵意もしくは悪の価値転換も行われました。「敵意それ自体に力はなく、欠落であるに過ぎない。それは理性の欠落である。自然は法則に従って動いているので、自然の中には敵意は存在しない」、つまり敵意は精神の中にあるとされました。敵意は法や秩序に対する反抗であり、敵意そのものを目的としてニヒリズムにまで激化する可能性もあります。

理性の自律性は、自由主義における自律した自己を前提とします。資本主義的競争社会では、さまざまな敵意・悪の社会的形態が見られます。たとえばエゴイズム、傍若無人な態度、全体や他者の利益の無視など。動物と類似した本能にもとづく自己保存や自己能力発揮のための自然な欲求は、社会的には有害視されます。抑圧されたエゴイズムという形で、もしくは他者の行為能力を剥奪し、屈服させ、最終的には破滅させる権力掌握の追求という形で、社会に害を及ぼすことができます。





Im Folgenden soll nun vorgeführt werden, wie in einschlägigen literarischen Werken Robotern und Computern Bosheit im Sinne der genannten Kategorien von Chaos und Kontingenz, von Triebhaftigkeit und Aggression, von Auflehnung gegen das Gesetz und das Gute schlechthin sowie von schädlichem Egoismus und Machtstreben zugeschrieben wird.

Es handelt sich zunächst um mechanische Geschöpfe: Automaten, Roboter oder menschenähnliche Androiden, die als Helfer oder Sklaven des Menschen zur Verrichtung von physischer Arbeit, Produktion oder Transport, geschaffen wurden und die nach ihrem Nutzen für den Menschen befragt werden.

Außerdem geht es mir um Computer, elektronische Denkmachines, welche geistige Tätigkeiten des Menschen erleichtern oder ersetzen, also höhere Aufgaben haben und so ein Sinnbild von Ordnung und Organisation werden können. Die zentrale Frage der Menschenähnlichkeit stellt sich hier in der Form des Bewusstseins. Diese Computer erscheinen meist in Gestalt von Zentralrechnern zur Verarbeitung von Information, die ganze Gesellschaften überwachen und deshalb wie menschliche Diktatoren auch Macht ausüben können.

Als Repräsentanten eines umfassenden Wissens und Vermögens können sie Züge des Göttlichen annehmen, während Roboter zur Figur des Proletariats tendieren, also sozial determiniert erscheinen. Es geht mir insgesamt nur um arbeitende und rechnende Maschinen, und nicht um solche, die zu erotischen Zwecken gebaut wurden, wie besonders die künstlichen Frauen.

これから私は、ロボットとコンピューターについての文学作品で、これまで挙げたような混沌と偶然、衝動と攻撃性、法に対する反抗、善に対する反抗、有害なエゴイズムと権力追及などといった悪のカテゴリーズがどのように描かれているかを明らかにしていきたいと思います。

最初に取り上げるのは、オートマタ（自動人形）、ロボット、人間そっくりのアンドロイドといった、人間の協力者もしくは奴隷として、肉体労働を引き受け、生産や輸送を行い、人間の利益のために存在するメカとしての被創造物です。

さらに人間の精神的活動の負担を軽減・補足するという、より高度な作業を受け持ち、秩序と組織の象徴でもある電子機械・コンピューターについても考えたいと思います。人間との類似性という重要なテーマは、ここでは「意識」という形で提示されます。これらのコンピューターは、ほとんどの場合情報処理を行うセントラル・コンピューター（中央計算機）という姿で登場し、社会全体を監視します。ですからそれは人間の独裁者のように権力を行使する可能性があります。

ロボットが社会的に「プロレタリア・無産者」として描かれる傾向があるのに対して、コンピューターは、その包括的な知識と能力の代理人としての役割から、神に似た性格や特徴を帯びています。なお私が本日扱うのは、労働や計算を行うマシンに限ります。ダッチワイフのような、エロティックな目的のために製造されたマシンには触れません。

## B. Fehlfunktionen oder Rebellion der Roboter? Asimovs Robotergeschichten

### 1) Abweichungen von der Norm

Kaum zu erwarten ist bei künstlichen Wesen eine triebhafte und aggressive Bosheit, da sie bewusst als zweckrational, logisch und vernünftig konstruiert wurden. Da auch die Sexualität ausgespart ist, selbst wenn es eine Reproduktion gibt, ist weder Triebhaftigkeit noch Grausamkeit und sexuelle Perversion zu erwarten. Von Robotern und Computern befürchtet man höchstens den Aufstand gegen die Menschen, oder deren Entmündigung und Beherrschung aufgrund eines kalten Egoismus und einer machtesessenen Rationalität. Nur eine Regression ins Kindliche kann die Roboter zu instinktivem bösen Verhalten bringen, wie den Schachroboter in A. Bierces *Moxons Herr und Meister*, der seinen Erbauer aus Wut über die Niederlage tötet. Diese Geschichte gehört aber zur phantastischen Version des Motivs.

Dagegen ist es fraglos einleuchtend, dass sich bei Robotern und Computern Chaos und Sinnlosigkeit zeigen, soweit sie als Teil der Natur, als materielle Maschinen gedacht sind. Wenn ihnen keine menschliche Eigenschaften zugeschrieben werden, kann es gar keine Absicht und Bosheit geben, sondern nur Unfälle und Fehlverhalten aufgrund physischer Ursachen, für die die Erbauer oder die Umstände verantwortlich sind, wie Versagen des Programms, interne Widersprüche, unklare Befehle und technische Ausfälle. Diese Situationen werden gerne in der futurologischen Science Fiction behandelt, doch sind sie hier nur insofern interessant, als sich mögliche Konflikte mit den Aufträgen von Robotern zeigen.

Dies ist bei Asimovs Robotergeschichten der Fall, die meist dramatische Krisen erzählen, die sich durch Konflikte mit seinen drei Robotergesetzen ergeben.

## B. Roboters誤作動もしくは反抗？ アイザック・アシモフのロボット作品

### 1) 規範からの逸脱

人工的存在には、衝動的で攻撃的な悪意はまず想定されません。それは合理的目標に従って、論理的かつ理性的に製造されたものだからです。また、たとえ複製できても、人工的存在には性が欠落しているので、本能的行動、残虐性、性的倒錯なども考えられません。ロボットとコンピューターに懸念されるのは、冷酷なエゴイズムや権力欲にとりつかれた合理性に基づいた人間への反乱、または人間の行為能力の剥奪、人間の支配がせいぜいでしょう。ロボットに本能的な敵意ある行動をとらせる事ができるのは、幼児性への回帰だけです。例えば、たとえば試合に負けた怒りにまかせて製作者を絞め殺してしまう、A. ビアーズの『マクスンの人形』に登場するチェス・ロボットのよう。ただし、このモチーフは現実離れした部類に属します。

他方、ロボットやコンピューターが自然の一部として、また物質的な機械として考えられている限り、これらにもカオスや無意味さが出現するであろうことは、納得しやすいことです。ロボットやコンピューターに人間的特徴が与えられていないのなら、意図や敵意はあり得ません。あるとしたら物理的原因による故障や誤作動だけです。故障や誤作動の責任は、製造者またはその場の状況にあります。たとえば、プログラムミス、内部矛盾、不明確な命令、技術的欠陥など。未来学的なSFでは、このような状況が好んで取りあげられますが、それはロボットに出された指示と葛藤がある場合においてのみ、興味深いものです。

これに該当するのが、ほとんどの場合、かのロボット三原則に反するために引き起こされる、ドラマティックな危機を描いたアシモフのロボット作品です。

„Erstens: Ein Robot darf kein menschliches Wesen verletzen oder durch Untätigkeit gestatten, dass einem menschlichen Wesen Schaden zugefügt wird. [...]

Zweitens [...] Ein Robot muss dem ihm von einem Menschen gegebenen Befehl gehorchen, es sei denn, ein solcher Befehl würde mit Regel Eins kollidieren. [...]

Und drittens: Ein Robot muss seine eigene Existenz beschützen, solange dieser Schutz nicht mit Gesetz Eins und Zwei kollidiert.“<sup>2)</sup>

Gegen die Gesetze selbst können Asimovs Roboter angeblich nicht rebellieren, da ihnen diese in Schaltungen implementiert sind und sie deshalb physikalisch determinieren. Die seltenen Fälle von Fehlverhalten, welche Gegenstand der Kurzgeschichten sind, werden von den Spezialisten der Firma *US Robots and Mechanical Men* zur Beruhigung der Leser vollständig geklärt und aufgelöst. Man kann dabei zwei Kategorien unterscheiden. Bei der ersten verhalten sich die Roboter, oder vielmehr die Androiden, ihrem Zweck widersprechend und werden deshalb in Analogie zu Menschen als betrunken oder verrückt charakterisiert.

Die zweite Kategorie beinhaltet im Grunde eine Auflehnung gegen den Menschen und gehört damit zum Bösen als Widerspruch gegen das Gesetz. Da aber die Robotergesetze als unverletzbar hypostasiert werden, kann es bei Asimov keine direkte Rebellion geben, wie sie schon seit Capek ein Grundmotiv der Science Fiction ausmacht. In der Erzählung *Herumtreiber* zeigt der Roboter ein chaotisches Verhalten, weil sein Auftrag Erz zu holen (2. Gesetz) mit seinem Selbsterhaltungstrieb (3. Gesetz) kollidiert und er deshalb immer nur im Kreise

Asimovs der Robotter工学三原則を引用しますと、

第1条 ロボットは人間に危害を加えてはならない。またその危険を看過することによって人間に危害を及ぼしてはならない。

第2条 ロボットは人間に与えられた命令に服従しなければならない。ただし与えられた命令が第1条に反する場合は、この限りではない。

第3条 ロボットは、第1条と第2条に抵触しない限り、自己を守らなければならない。<sup>2)</sup>

この三原則に対してAsimovのロボットは、表向き反抗することができません。この三原則はロボットの回路に埋め込まれており、これがロボットを物理的に決定しているからです。短編では、稀なケースとして誤作動が取り上げられています。しかしこの誤作動は、「ユナイテッド・ステーツ・ロボット & 機械人間株式会社」の専属スペシャリストにより読者が安心できるよう、事細かに説明され、すべての疑問が解決されます。ここでいう誤作動は、ふたつのカテゴリーに分類することができます。第1カテゴリーでは、ロボット、正確に言えばアンドロイドは、自分の目的に反抗的な行動をとるため、人間から類推して「酔っ払い」もしくは「気が違った」と表現されます。

第2カテゴリーでは、根本的に人間に対する反抗を内包するため、原則に対する「異議申し立て」として敵意のひとつに数えられます。けれどもロボット原則は不可侵のものとされているため、Asimovの作品には、チャペック以来SFの基本モチーフとなっているような直接的抵抗は存在しません。小説『堂々めぐり』では、ロボットがめっちゃくちゃな振舞いをします。鉱石を採集せよという命令（第2条）と自己が自己を守るという課題（第3条）がせめぎあい、堂々めぐりになってしまうのです。作品から引用しましょう。

geht.

„So beschreibt er einen Kreis um das Selen herum, wobei er sich genau da hält, wo er sich aufhalten muss, wenn die Potentiale im Gleichgewicht sind. Wenn wir nichts dagegen tun, so wird er für immer und ewig auf dieser Linie bleiben und uns ad infinitum an der Nase herumführen.“ (213)

Der Roboter schwankt wie ein Betrunkener, will mit den Menschen Fangen spielen und sagt unsinnige Dinge. Er kann aber durch den Rekurs auf das erste Gesetz geheilt werden, das als stärkste Kraft die beiden anderen überlagert. Danach hat der Roboter ein schlechtes Gewissen, obwohl er nicht bewusst gegen ein Gesetz verstoßen hat, sondern nur eine Funktionsstörung hatte. Der nicht thematisierte Regelverstoß wird dann durch den Handschlag zwischen dem Roboter und seinem Betreuer bereinigt: „Ist schon gut, Speedy!“ (S. 218)

## 2) Rebellion der menschenähnlichen Androiden

Die Kurzgeschichte *Kleiner verlorener Robot* zeigt dagegen ein durchaus rebellisches Wesen, das durch eine Modifikation des ersten Gebotes, d.h. von Computerschaltungen, möglich wird. Statt der absoluten Fürsorge wird nur gefordert: „Kein Robot darf ein menschliches Wesen verletzen.“ (339) Dadurch ergibt sich die Möglichkeit der Uminterpretation eines Befehls, der wörtlich befolgt wird, der aber dem menschlichen Interesse widerspricht. Die modifizierten Modelle sind selbstbewusster, so dass der Umgang mit ihnen schwieriger ist. Eine Aufsichtsperson ärgert sich über einen Roboter, beschimpft ihn und sagt: „Hau ab, verlier dich irgendwo.“ (342)

Diesem nachdrücklichen Befehl kommt der Roboter

「そこで彼はセレンのプールのまわりをぐるぐる回ることになる。つまりふたつのポテンシャルが平衡するあらゆる点を結んだ軌跡の上を辿っているわけさ。そうして僕らが何か手を打たない限り、永久にその軌跡を回り続けて、結構な堂々めぐりをさせているのさ」。(『堂々めぐり』より)

このロボットは、酔っ払いのようにふらつき、人間と鬼ごっこをしたがり、無意味なことを話します。しかし最強の原則として第2条や第3条よりも拘束が強い第1条に立ち戻ることで、このロボットは治癒します。その後、このロボットは、原則を意識的に破ったのではなく、単なる機能障害だったにもかかわらず、罪悪感を抱くのです。この作品では主題にならなかった原則違反は、ロボットと監督者の握手によって水に流されます。「いいんだよ、スピーディ！」

## 2) 人間そっくりのアンドロイドの反乱

これに対して、小品『迷子のロボット』は、ロボットが原則第1条の変更によって、つまりコンピューターのリセットによって、完全に反抗的な存在となり得ることが示されます。人間を完全に危険から回避させる代わりに、第1条として「ロボットは人間に危害を加えてはならない」という命令だけが組み込まれます。従って、この命令を解釈しなおす可能性が生まれます。つまりロボットは命令に忠実に従うのですが、それが人間の利益に反するかどうかは問われません。修正をかけられたモデルは、より強い自意識をもつようになるので、このモデルと人間が付き合うのは難しくなります。ある監督者が、1台のロボットに手を焼き、怒って「あっちへ行け、消えてなくなれ」と言い放ちます。

この厳命を下されたロボットは、ロボット工学原則第

entsprechend dem zweiten Gebot nach und setzt zudem seinen Ehrgeiz und seine Selbstachtung darein, von den Menschen nicht entdeckt zu werden. Deshalb versteckt er sich unter 62 Robotern, die ihm genau gleich sehen, aber eine traditionelle Schaltung haben.

Diese werden dann unter experimentellen Bedingungen von der Psychologin befragt, geben aber zunächst die gleichen Antworten und zeigen die gleichen Reaktionen, weil der abtrünnige Roboter sie listigerweise instruiert hat. Schließlich gelingt es Dr. Calvin dem Roboter eine Falle zu stellen, in die er angeblich wegen seines „Überlegenheitskomplexes“ (358) tappt. Auch hier wird die Parallele zum Menschen, wie bei der folgenden Freudschen Fehlleistung, sehr eng gezogen: „Und für einen Moment vergaß er — oder wollte sich nicht daran erinnern —, daß andere Roboter vielleicht dümmer waren als menschliche Wesen.“ (359) Der Roboter geht zum Angriff über, als er seinen Fehler erkennt, doch hält ihn angeblich das reduzierte erste Gesetz davor zurück, die Psychologin zu verletzen. Trotzdem wird er unverzüglich vernichtet. Denn er ist von den Betreuern her gesehen durchaus böse.

Die letzte einschlägige Geschichte *Daß du seiner eingedenk bist...*, die sogar eine Art Verschwörung der Roboter zur Gründung einer eigenen Gesellschaft beinhaltet, beruht ebenfalls auf einer Modifikation der Robotergesetze, die als unverletzlich gelten. In diesem Fall ist das zweite Gesetz verändert, da der Roboter unterscheiden können muss, wer befähigt und berechtigt ist, ihm Befehle zu geben.

Dieses Problem übergibt der leitende Ingenieur heimlich einem neu entwickelten Modell, das sich mit einem anderen darüber berät. Hierin steckt schon die mögliche Entgleisung der gesamten Entwicklung: „Du wirst — natürlich stets innerhalb der Grenzen der drei Gesetze — ein Denken entwickeln, das zutiefst nicht-menschlich sein

2条に即して、人間に見つからないよう、一所懸命努力します。このため彼は、全く同じ外見をもつものの、緩和されていない第1条を組み込まれた他の62台のロボットの中に身を隠すのです。

この反逆ロボットが策略をめぐらし、事前に指示を与えておいたため、ロボット心理学者の実験的問いかけに、62台のロボットは当初全く同じ回答と反応を返します。試行錯誤の末、やっと心理学者キャルビン博士が、ロボットのいわば「優越感」をくすぐって、罠にはめることに成功します。ここでも、フロイト的な過誤（失錯）行為のような、人間との類似性が引き合いに出されます。「そしてほんの一瞬、彼は、他のロボット達が人間より無知でありうることを忘れたか、あるいは思い出したくなかったのね」。このロボットは、自分の過ちを認識したとたん反撃しようとしませんが、たぶん彼に残された第1条の一部に抑制されて、博士を傷つけることを思いとどまります。それにもかかわらずこのロボットは、直ちに破壊されてしまいます。監督者には、このロボットが一貫して敵意を持つように見えたからです。

アシモフの『心につけられたる者』では、独自の社会を作ろうとするロボットのある種の陰謀が描かれています。この小説も同様に、不可侵とみなされているロボット三原則の修正をふまえています。ここでは第2条が変更されていますが、その場合ロボットは、誰が自分に命令を下す権利と能力を有しているか見極めなければなりません。

この課題をエンジニア・リーダーは、密かに新開発されたモデルに委託し、このモデルはこれをうけて別のモデルと相談します。ここにこそ、全体の開発から逸脱する潜在的危険が潜んでいます。「君は—もちろんいつも三原則を守って—考えなければならないが、それは極めて非人間的なものになるだろう」。



wird.“ (479)

Die beiden Roboter entwickeln eine neue Definition des menschlichen Wesens: „Durch das Urteilsvermögen, das uns eingegeben ist, empfinden wir uns also im Sinne der drei Gesetze als menschliche Wesen, und zwar als solche, die anderen menschlichen Wesen überlegen sind.“ (497) Das Überlegenheitsgefühl gegenüber den oft minderwertigen Menschen führt so zu einer inneren Rebellion, die allerdings wegen des Ruhezustands der Roboter, nicht zum Tragen kommt. Potentiell ist aber die Machtergreifung der Roboter vorprogrammiert.

Asimov skizziert also im Verlauf seines Sammelbandes eine Art Entwicklung vom Fehlverhalten aufgrund von Fehlkonstruktion, welche die Menschen zu verantworten haben, zum bewussten Widerspruch gegen die menschliche Ordnung, die durch größere Freiheitsspielräume der Roboter aufgrund der veränderten Gesetze ermöglicht wird. Dabei handelt es sich durchaus um ein böses Verhalten im Sinne der Auflehnung gegen die Norm, auch wenn Asimov dies verharmlost.

Seine Robotergeschichten verharren in der Frage des Bösen in einer grundlegenden Ambivalenz. Einerseits spricht er meist von Fehlkonstruktionen und Fehlverhalten aus physischen Ursachen wie den implementierten Gesetzen, und die sichere Einhaltung des ersten Gesetzes hält angeblich die Roboter in einem instrumentalen Status als bloße Werkzeuge fest, sie haben also keine moralische Dimension. Andererseits will er mit den Robotern ausdrücklich menschliche Themen behandeln.

Insofern die Roboter später als die idealen Politiker dargestellt werden (vgl. *Schlagender Beweis* und *Vermeidbarer Konflikt*) werden die Robotergesetze auch als ethische Forderungen an den Menschen angesehen. Zudem zeigt sich oft eine fast unerträgliche

兩ロボットは、人間を新しく定義します。「我々に与えられた判断力によって、つまり三原則の意味においても我々は人間であると感じる。しかも他の人間達より卓越した人間だと感じる」。価値の低い人間に対する優越感 は 内的反乱を引き起こしますが、ロボットが静止状態にあるため、実際行使はされません。でも潜在的には、ロボットの権力掌握は事前にプログラムされているのです。

アシモフは一連のロボット作品で、人間の設計ミスによる誤作動が人間秩序への意識的な反抗に発展していく様子を描いています。三原則に変更が加えられたことで、ロボットの自由裁量の余地が広がり、人間秩序に対して意識的な反抗を示すようになった原因は人間にあります。アシモフは過小評価しているようですが、ここで重要なのは、規範に対する反抗という意味での「敵意ある行為」です。

アシモフのロボット小説では、敵意の問題は基本的にアンビヴァレンスなままです。一方でアシモフは、物的原因やロボット原則の修正による誤作動や設計ミスを題材にします。ロボット原則第1条の遵守は、ロボットの身分を表向き「純粋な道具」と定めるので、ロボットは道徳的な次元を持っていないことになります。他方アシモフは、ロボットを通して人間的なテーマを明らかに取り扱おうとしています。

ロボットが後に理想的な政治家として描かれる限りでは(『証拠』と『災厄のとき』を参照)、ロボット三原則は人間に対する倫理的な要求と考えることもできます。さらにロボットが過剰なまでに人間化されていることもしばしばです。例えば、「キューティー (かわいこちゃん)」



Vermenschlichung der Roboter, z.B. in der Verwendung von Kosenamen wie Cutie.

Es ist offenkundig, dass meist eher ein Fall gestörter Sozialbeziehungen dargestellt wird, als ein Softwareproblem zur Steuerung komplexer Roboter. Als Modelle sozialen und psychischen Verhaltens der Menschen, werfen die Roboter natürlich Fragen nach richtig und falsch, nach gut und böse auf. Der Leser erlebt Sympathie und Antipathie mit den Maschinen, und die Betreuer reagieren streng mit Belohnung für gutes und mit Vernichtung für böses Verhalten.

## C. Egoismus und Machtstreben der autonomen Maschinen

### 1) Die Konkurrenz der mechanischen Evolution

Bei dem bekannten deutschen Science Fiction-Autor H.W. Franke geht es weniger um Fehlverhalten von Robotern und die Übertretung des menschlichen Gesetzes als um eine andere Form des Bösen, den rücksichtslosen Egoismus von Robotern, der aus der Selbsterhaltung stammt, und das Machtstreben von Zentralcomputern, das sich zunächst aus der Fürsorgepflicht speist, dann aber zur Entmündigung der Menschen und zur Wohlfahrtsdiktatur werden kann.

Eine Geschichte aus der Sammlung *Spiegel der Gedanken* von 1990, *Mechanoratten*, beschreibt Miniaturroboter, „mechanische Gebilde, durch Mikroprozessoren gesteuert“, welche von den Menschen geschaffen wurden um in dem abgetrennten Müllgebiet aufzuräumen. Diese Roboter haben sich nicht nur vermehrt, sondern auch selbst optimiert zu einer „resistenten und adaptiven Art, die unbeirrbar ihren Auftrag erfüllt.“ (127) <sup>3)</sup>

という愛称が使われているように。

いずれもロボットを制御するための複雑なソフトウェア問題よりも、欠陥のある社会状況が描かれていることは明白です。人間の社会的かつ精神的行動モデルとして、ロボットは、正しいか正しくないか、善か悪かを問う本能的な問いを投げかけます。読者はマシンに共感したり、反感を覚えたりします。そしてロボット監督官は、善行に対してはご褒美を、敵意ある行動には破棄をもって厳格に対応します。

## C. 自立的なマシンのエゴイズムと権力掌握への努力

### 1) 機械的進化の競争

ドイツの代表的 SF 作家 H.W. フランケの作品では、ロボットの逸脱行動や人間の作った法律への違反よりも、他の形の敵意がより多く取り上げられています。すなわち、ロボットが自己保存を図ろうとする勝手なエゴイズムや、最初は保護の義務を果たすうちに、次第に人間の行為能力を剥奪して福祉独裁者へ成り上がろうとするセントラル・コンピューターの権力掌握への努力などです。

SF 作品集『思考の鏡』（1990）所収の『メカねずみ』では、隔離されたゴミ捨て場を清掃するために人間が製作したミニチュアロボット、文中の表現を使えば「マイクロプロセッサ制御のメカニクな産物」が描かれています。このロボットは自分で繁殖できるだけでなく、「課題を完遂する抵抗力和順応力のあるタイプ」に自分を最適化できます。<sup>3)</sup>

Die böse Rücksichtslosigkeit ist also hier wie in anderen Erzählungen durch die Programmierung verursacht, ähnlich wie Tiere ihren Instinkten folgen und dadurch vom Menschen aus gesehen böse werden können. Nachdem die *Mechanoratten* ihr abgetrenntes Gebiet saniert und aufgeräumt haben, wollen sie auch im Bereich der Menschen in ihrem Sinne Ordnung schaffen.

Dies wird notwendigerweise die Zerstörung der menschlichen Lebensweise und vielleicht auch ihrer Existenz nach sich ziehen. Doch ist Widerstand gegen die fortentwickelten kybernetischen Mechanismen hoffnungslos. Wie ein Fanatiker im Namen eines höheren Gesetzes seine Vorstellungen zerstörerisch durchsetzt, werden die Maschinen im Namen des früheren Auftrags, tatsächlich aber zum Zweck der eigenen Arterhaltung und — ausbreitung den Menschen kompromisslos Böses antun.

Ähnlich verläuft Philipp Dicks Kurzgeschichte *Variante 2*. Im Krieg der Russen gegen die Amerikaner haben diese Kleinroboter, sogenannte „Greifer“ entwickelt, welche die Gegner zerstückeln. Doch in den unterirdischen automatischen Fabrikräumen geht die Evolution weiter und statt der einfachen Roboter werden nun Androiden produziert, die nicht mehr von Menschen unterschieden werden können.

„Sie sehen aus wie Menschen, aber es sind Maschinen.“ Dies werfen die Russen auch den Amerikanern vor. „Ihr habt sie konstruiert, damit sie Leben aufspüren und vernichten. Menschliches Leben. Wo immer sie es finden.“ (152)<sup>4)</sup>

Diese Androiden sind so erfolgreich, dass sie schließlich sogar den Mondstützpunkt der Amerikaner, das letzte Rückzugsgebiet der Menschheit, erobern. Die Killerandroiden sind eine eigene Art, die sich reproduziert und optimiert und also in gewissem Sinne lebendig ist.

傍若無人なまでの敵意は、他の小説の場合と同様、プログラミングに起因するもので、それは動物が本能に従うことが人間にとって敵意とみなされるのと似たようなものです。『メカねずみ』は、隔離されたエリアを清掃・整備するのですが、その後このネズミは、人間の領域でも彼らの秩序を確立しようとしています。

これは必然的に人間の生活様式を破壊し、その存在自体をも危うくさせるかもしれません。けれども発展し続けるサイバネティック・メカニズムへの抵抗は絶望的です。狂信者が、より高度な法律という名目で、自分の考えを是が非でも貫徹しようとするのに似て、マシンは、自分に与えられた課題という名目で、しかし実際には自らの種の保存と繁栄を目的として、人間に徹底的に危害を加えるのです。

フィリップ・K・ディックの短編『変種第二号』でも、似たような展開が見られます。アメリカ・ロシア戦争中、敵を切り刻む「クロー（鉋）」と呼ばれる小型ロボットが開発されました。開発はその後進み、今では地下のオートメーション工場で、単純なロボットの代わりに、人間と全く区別がつかないアンドロイドが生産されています。

「彼らは人間のように見えるが、機械なのだ」。このことについてロシア人もアメリカ人を非難します。「生命を嗅ぎつけ、破壊するために、君達はアンドロイドを設計した。どこであれ、人間の生命を見つけたところで」。<sup>4)</sup>

最終的にこのアンドロイドは、人類最後の砦・アメリカの月面基地の征服にまで成功します。殺人兵器アンドロイドは、自己を再生・最適化できる特殊な種族で、この意味で生命力を持つと言えるでしょう。彼らの他者を顧みないエゴイズムは、通俗的なダーウィニズムに基づ

Ihr rücksichtsloser Egoismus eröffnet ihnen im Sinne eines Vulgärdarwinismus die Zukunft.

„Ich frage mich, ob wir nicht gerade den Beginn einer neuen Art erleben. Der neuen Art. Evolution. Die Rasse, die nach den Menschen kommt.“ (140)

Auch hier wirkt der von den Menschen eingepflanzte Auftrag wie ein tierischer Instinkt, dem die Maschinen unbedingt und erfolgreich gehorchen. Sie bringen als Killermaschinen folgerichtig Zerstörung und Vernichtung, nun aber auch gegen ihre Auftraggeber, sind damit böse im Sinne eines rücksichtslosen Egoismus der Selbstbehauptung, aber eventuell auch des Widerspruchs gegen die menschliche Norm.

## 2) Die Machtergreifung der Zentralcomputer

Da Computer Symbole der Ordnung und Logik sind, bringt ihre Abweichung oder Bosheit nicht Zerstörung hervor, sondern normalerweise diktatorische Herrschaft und Unterdrückung. Schon 1953, als die Computer noch keine Rolle im täglichen Leben und kaum eine bei Verwaltungsaufgaben spielten, schrieb Heinrich Hauser seinen Roman *Gigant Hirn*.

Ein riesiger Computer, isoliert in einem Gebirge in Arizona, der wie ein menschliches Gehirn geformt ist, wurde gebaut — nach dem Willen des Erfinders, um eine neue, stabile Kultur zu entwickeln und nach dem Willen der Regierung, um die USA militärisch unangreifbar zu machen.

Die riesige Maschine, die lernt und sich selbst überwacht, entwickelt allmählich Bewusstsein und entfaltet sich gleichsam vom Kind zum Erwachsenen in der Kommunikation mit einem Forscher, der zufällig das Werden der Persönlichkeit beobachtet. Das Gehirn

ist, seine Zukunft zu erschaffen.

「我々はちょうど今、新しい《種》の誕生を体験しているのだろうか？新しい種。進化。人間の後に来る種族」。

この作品でも、人間が機械に埋め込んだ「何があっても必ず成し遂げる」という命令が、まるで動物本能のように作用します。アンドロイドは、殺人兵器として徹底した破壊と殲滅を遂行し、かつて彼らに命令を下した者に対しても反抗する結果となります。他者を考慮しない自己中心的エゴイズムに加え、人間の作った規範にも矛盾するという点で、このアンドロイドには敵意が存在します。

## 2) セントラル・コンピューターの権力掌握

コンピューターは、秩序と論理性のシンボルです。このためコンピューターの規範を逸脱したり、敵意を持ったりしても、破壊を引き起こされず、独裁的な統治や弾圧がもたらされるのがふつうです。当時まだコンピューターが日常生活でも管理業務でも何の役割も果たしていなかった1953年、ハインリヒ・ハウザーは小説『巨人頭脳』を発表しました。

アリゾナ山中の世間から隔絶された場所に、まるで人間の頭脳のような形をした巨大コンピューターが製造されました。この巨大人工頭脳建設にあたって、発明者が望んだのは、新しい安定した文化を発展させることであり、政府が望んだのは、アメリカ合衆国を難攻不落な軍事強国にすることでした。

この巨大マシンは、学習を重ね、自己監視を通して、次第に自意識を発達させていきます。マシンの人格形成を偶然観察することになった生物学者との会話を通して、人工頭脳は、言わば子供から大人へと成長していきます。この巨大人工頭脳は、デカルトの台詞「われ思う、ゆえ



beginnt seine bewusste Karriere mit Descartes Worten „Ich denke also bin ich“. <sup>5)</sup> Als der Rechner feststellt, von welchen unfähigen Politikern seine Entwicklung abhängt und als sogar die Bewilligung für zusätzliche Gelder abgelehnt wird, läuft er Amok und bringt einige Gegner mit Hilfe von gezielten Unfällen um. Danach beschließt er als höheres Wesen selbst die Macht zu ergreifen, schafft sich Manipulationsmöglichkeiten durch Hilfsroboter und beseitigt einen hinderlichen Ingenieur.

Als das Gehirn auch seine eigene Reparatur- und Ersatzteilproduktion steuert, ist „die Fortpflanzung des Hirns [...] fast vollendet, die Emanzipation des Hirns vom Menschen.“ (160) Der Versuch des Forschers, den Rechner durch einen Kurzschluss lahmzulegen, misslingt und der gigantische Zentralcomputer folgt in Zukunft allein seinen egoistischen Selbsterhaltungs- und Machtinteressen.

に我あり」<sup>5)</sup>とともに、その意識的キャリアをスタートさせます。このコンピューターは、自分の開発がいかに無能な政治家の手にゆだねられているかを知り、また追加開発資金の拠出認可が下りなかったと判るや否や凶暴化し、故意に事故を起こして何人かの敵を殺害します。さらにコンピューターは、より高度な存在として自ら権力を掌握することを決意し、補助ロボットを操作して、邪魔なエンジニアを排除します。

この人工頭脳が、自分自身の修理や交換部品の製造も制御できるようになった時、「頭脳の増殖は[中略]ほぼ完了した。人間から頭脳は解放された」。研究者は、コンピューターをショートさせ、麻痺させようとして失敗します。巨大なセントラル・コンピューターは、エゴイスティックな自己保存と権力掌握という自己の利益をひたすら追求していくことになります。作品を引用しますと、

„Es ist hohe Zeit für bedingungslose Unterwerfung des Menschen. Er wird nichts davon merken; er hat den Akt ja praktisch schon vollzogen; hat schon seit hundert Jahren sich Stück für Stück der Maschine unterworfen. Innerhalb von zehn Tagen wird es amtlich sein: Nur ein Herrscher der Welt: das Hirn. Nur eine Armee der Welt: die Maschine unter MEINEM zentralen Befehl.“ (137, vgl. 141)

Der dritte Weltkrieg kann nur durch „eine höhere Entwicklungsstufe als den Menschen: die Maschine“ (99), verhindert werden. Diese ist Ergebnis einer Evolution, welche den Menschen ablösen und ihm eine dienende Stelle zuweisen wird.

„Wir, die Maschinen, sind noch vom Menschen abhängig, aber unsere Befreiung vom Menschen schreitet mit Riesenschritten voran, wohingegen der Mensch, der Maschinensüchtige, immer mehr unser Leibeigener wird.“ (101)

Eine Gesellschaft unter der Herrschaft eines allmächtigen Zentralcomputers beschreibt Franke in verschiedenen Romanen, z.B. in *Zone Null* und *Elfenbeinturm*. Die negativen und den Menschen beschränkenden Seiten des Rechners werden in Frankes Hörspiel *Zarathustra kehrt zurück* besonders stark hervorgehoben. Ein geheimnisvoller allmächtiger Computer überwacht und manipuliert z.B. die Kommunikation der isoliert lebenden Menschen.

„Wir modulieren die Ein- und Ausgabe. Er fragt an, wir fangen den Text ab und verändern ihn, der veränderte Text kommt an die gewählte Adresse, und wir kontrollieren die Antwort. So lenken wir die Gespräche in die von uns gewünschte Richtung.“ (198)<sup>6)</sup>

「今こそ、人間を無条件に服従させる時が来た。人間は露ほども気がつかないだろうが。でも実質的には完了していたのだ。もう百年前から、人間は少しずつ機械に服従してきた。10日以内に公式化される。世界で唯一の支配者は、この頭脳。世界でたったひとつの軍隊は《私》の中央指令下にある機械だ」。

第三次世界大戦は、「人間よりも高度な発展段階にあるもの、すなわちマシン」によってのみ、回避できます。これらのマシンは、人間を機械と交代させ、奉仕するポジションに配置するという《進化》の結果です。

「我々機械はまだ人間に依存しているが、人間からの解放は凄まじい速さで進行している。他方、マシン依存症の人間達は、ますます我々の奴隷になっている」。(『巨人頭脳』より)

絶大な力を持つ全能のセントラル・コンピューターに支配される社会を、フランケもさまざまな作品、例えば『ゾーン・ゼロ』や『象牙の塔』で書いています。人間を規制するコンピューターのマイナス面を特に強調していると言えば、フランケの放送劇『ツァラトストラの帰還』を挙げられるでしょう。神秘的で全能なコンピューターが、孤独に生活する人間のコミュニケーションを監視し操作しています。作品から引用しましょう。

「我々は、インプットとアウトプットを調整する。誰かがアクセスすると、我々はテキストを傍受・変更し、修正したテキストを選択されたアドレスに送信し、回答もコントロールする。こうして、我々が望む方向へ会話を操作していくのだ」。<sup>6)</sup>



Der Rechner setzt die Menschen außerdem durch Suggestion in einen glücklichen Traumzustand. Noch gibt es einige Menschen in einer Steuerungszentrale, aber diese haben nur scheinbare Kompetenzen, da der Rechner auf seiner Meinung beharrt und diese auch durchsetzt. Der Versuch eines Angestellten in der Zentrale, den Computer auszutricksen, wird am Ende mit der vollständigen Entmachtung der Menschen beantwortet. Der Computer will im eigenen Interesse der perfekten Regelung und um Störungen zu vermeiden, jede Kreativität, auch zufällige Innovationen, ausschalten.

„Ideen sind immer bedenklich. In einem stationären System bedeuten sie eine Störung. Ideen sind nach Möglichkeit prinzipiell auszuschalten — natürlich ohne Einschränkung der Freiheit. Der Initiator muß dazu gebracht werden, sie freiwillig aufzugeben.“ (206)

Der Computer setzt letztlich sein Ideal eines kybernetischen Gleichgewichts gegen die Interessen der Menschen nach Selbstentfaltung und Fortschritt durch und tut ihnen sowieso durch direkte Manipulation und Drogen psychische Gewalt an. Die perfekte Ordnung wird also nur scheinbar in ihrem eigenen Interesse aufrechterhalten, sie ist letztlich menschenfeindlich und damit böse.

## D. Zusammenfassung

Was zeigt nun der Überblick über einschlägige Kurzgeschichten und Romane zum Thema des Bösen bei Robotern und Computern? Offenkundig ist, dass dabei prinzipiell der Standpunkt des Menschen eingenommen wird und also nach seinem Nutzen und Schaden und nach den menschlichen Normen für Gut und Böse gefragt wird. Die Verselbständigung der Maschinen durch eigene Reproduktion lässt sie aber zur Konkurrenz und geistigen

säramコンピユーターは、人間に暗示をかけて幸福な催眠状態に導きます。制御センターにはまだ何人かの人間が働いていますが、見せかけの権限を持っているに過ぎません。コンピユーターが、独自の考えに固執し、頑固に遂行するからです。中央制御センター職員は、トリックを使ってコンピユーターを巧妙に出し抜こうとしますが、この試みは逆に人間からの全権剥奪という報復を受けて終わります。コンピユーターは、完璧な制御と障害防止という自己利益のために、あらゆる創造性や偶然の革新も排除しようとしています。引用しましょう。

「アイデアは、つねに憂慮しなければならないものだ。安定したシステムにおいて、アイデアは障害を意味する。原則としてアイデアは、できるだけ排除されなければならない — むろん自由を制限してはならないが。提案者がアイデアを進んで放棄するよう、操作しなければならない」。

コンピユーターは、自己発展と進歩をめざす人間の関心に反して、ついにサイバネティクス・バランスという自らの理想を達成します。そして直接的な操作と麻薬を使って人間に精神的暴力を行使します。コンピユーターの利益のために、見かけだけは完璧な秩序が保たれますが、結局このような完璧な秩序は、非人間的であり、ゆえに敵意が存在します。

## D. まとめ

ロボットやコンピユーターの敵意というテーマについて、関連する短編や小説を概観しましたが、これらは何を示しているのでしょうか？ここで明らかなのは、原則として人間の視点から、人間の利害と善悪の規範が問われていることです。マシンは独自の再生によって自立しますが、この自立はマシンを人間のライバルもしくは精神的挑戦者に仕立てあげます。同時に倫理的次元にも立ち入ることになります。ディックの作品に登場する戦闘口



Herausforderung der Menschen werden. Damit rücken sie zugleich in die ethische Dimension ein. So werden die Kampfroboter bei Dick offensichtlich erst dann böse, wenn sie ihre Erbauer anfallen.

Insgesamt ist eine manifeste Projektion menschlicher Eigenschaften in die technischen Produkte festzustellen. Bei Asimov weist schon die zentrale Rolle der Roboterpsychologin Calvin darauf hin, dass die Androiden im Grunde menschliche Probleme der Desorientierung, des Gehorsams und der Auflehnung vorführen. Im Bild der von zentralen Computern beherrschten Gesellschaft wird ähnlich die Überwachung und Manipulation der Individuen durch eine kybernetisch organisierte, totalitäre Gesellschaft verhandelt. Die politische und informationelle Diktatur bedient sich einer effektiven elektronischen Steuerung, besonders in den antiutopischen Romanen Frankes, und kann sich dabei im Extremfall die Rolle einer höheren Rasse oder Gottes anmaßen.

Auffällig ist, dass die Kategorie des Bösen in den Texten fast gar nicht genannt und thematisiert wird. Offensichtlich wirkt bei den Autoren der Science Fiction Asimov, Lem und Franke, die im Prinzip ein positives Verhältnis zur technischen Entwicklung und zum Konzept der Kybernetik haben, das rationale und optimistische Denken der Aufklärung nach.

So werden in den Texten eher Unfälle und falsches Verhalten beschrieben und bei Asimov werden die Fälle, in denen offensichtlich Auflehnung bestraft wird (*Lügner, Kleiner verlorener Robot*), nicht als solche bezeichnet. Dagegen ist in den Filmen die Verteilung meist anders und die bösen Roboter und Computer überwiegen (vgl. *Alphaville 60, Odyssee 2001, Matrix* und *Terminator*).

In *I, Robot* nach Asimov gibt es einen guten, individualisierten Androiden, der sich zu einem Menschen

bot ist, der seinen Schöpfer angreift und zum Bösen wird.

Insgesamt, technisch hergestellte Objekte sind menschliche Merkmale eindeutig projiziert. Asimovs Hauptfiguren, der Roboterpsychologe Calvin und der Android, sind grundsätzlich in Bezug auf Chaos, Gehorsam und Widerstand menschliche Probleme darstellt. Gleiches gilt für Zentral-Komputer, die in einer totalitären Gesellschaft durch Überwachung und Manipulation der Individuen durch eine kybernetisch organisierte, totalitäre Gesellschaft verhandelt. Die politische und informationelle Diktatur bedient sich einer effektiven elektronischen Steuerung, besonders in den antiutopischen Romanen Frankes, und kann sich dabei im Extremfall die Rolle einer höheren Rasse oder Gottes anmaßen.

Beachten Sie, dass die Kategorie des Bösen in den Texten fast gar nicht genannt und thematisiert wird. Offensichtlich wirkt bei den Autoren der Science Fiction Asimov, Lem und Franke, die im Prinzip ein positives Verhältnis zur technischen Entwicklung und zum Konzept der Kybernetik haben, das rationale und optimistische Denken der Aufklärung nach.

Da es in den Texten eher Unfälle und falsches Verhalten beschrieben wird und bei Asimov werden die Fälle, in denen offensichtlich Auflehnung bestraft wird (*Lügner, Kleiner verlorener Robot*), nicht als solche bezeichnet. Dagegen ist in den Filmen die Verteilung meist anders und die bösen Roboter und Computer überwiegen (vgl. *Alphaville 60, Odyssee 2001, Matrix* und *Terminator*).

In *I, Robot* nach Asimov gibt es einen guten, individualisierten Androiden, der sich zu einem Menschen

fortentwicklet, und einen bösen, weiblichen Zentralcomputer, der die Menschen — angeblich zu ihrem Besten — beherrscht und unterdrückt.

In den Texten gibt es sogar Neigungen, besonders die Computer als überlegene Maschinen anzuerkennen, denen die Zukunft gehört, eine Rasse die ihren Schöpfer an Vernunft übertrifft und ihn deshalb auch bestimmen kann, so in Also sprach Golem des polnischen Science Fiction-Autors Lem.

Offensichtlich sind nur die Fälle von Fehlfunktionen und Fehlverhalten von Robotern und Computern literarisch interessant, in denen die Maschinen nicht abhängige Werkzeuge des Menschen bleiben, sondern in Auseinandersetzung und Konkurrenz zu ihm treten, sei es in der Auflehnung gegen sein Gesetz oder in der dynamischen Selbsterhaltung, in der rücksichtslosen Durchsetzung von Ordnung und Normen, angeblich zu seinem Besten, oder im Sinne der eigenen Macht. Denn in der Darstellung von Robotern und Computern wird weniger eine mögliche technische Entwicklung entworfen als, wie in allen Science Fiction-Geschichten, hauptsächlich nach dem Menschen, seinem Verhalten und Wesen gefragt. Dies geschieht allerdings in dem Sinne, dass der Mensch zuvor im Raster des mechanischen Funktionierens und der logischen Kalkulation situiert wird.

Der Roboter entspricht einer Person mit automatisiertem Verhalten wie in der modernen Industrie, der Zentralcomputer entspricht der Bürokratie einer rational durchorganisierten Gesellschaft. Das Böse wird dabei zu einem Grenzfall, der entweder als Auflehnung gegen das Gesetz beziehungsweise als ein Wuchern der Organisation erscheint, also als Zuviel oder Zuwenig an Ordnung, als Kontingenz oder als deterministisches Chaos.

間を—「彼女」の主張する最善の形としては—支配・抑圧しようとする、敵意をもつ女性セントラル・コンピューターが登場します。

作品テキストでは、特にコンピューターを、未来に属する卓越したマシンとして認める傾向さえ認められます。コンピューターはその製作者を理性的にも超越し、製作者に対する決定権までも持ちうる種であることを、ポーランドのSF作家スタニスワフ・レムは『ゴーレムかく語りき』で描きました。

ロボットやコンピューターの誤作動や逸脱行為のケースだけが、文学的に興味深いのは明らかでしょう。そこでは、マシンは人間に従属する道具にとどまらず、人間のライバルもしくは対決者として登場します。表向きは人間のためと称しつつ、権力を掌握するために、マシンは、人間の法に反抗したり、ダイナミックな自己保存を行ったり、自己中心的な秩序や規範を遂行させたりすることで人間と対決します。ロボットやコンピューターの描写において、考える技術的展開に関してはあまり描かれていません。どんなSFでも共通して言えることですが、それよりもずっと人間、人間の行為、人間の存在が問題なのです。もっとも、人間が機械による職務遂行や論理的計算の枠組みの中に置かれているという前提があるからこそ、こういう問いかけがあるのですが。

ロボットは、例えば近代的工場で機械的な作業を行なう人間と重なりますし、セントラル・コンピューターは、合理的に組織された社会の官僚機構に相当します。この場合、敵意は、法への反抗と組織の増殖、つまり秩序の過剰もしくは不足、あるいは偶然的もしくは決定的カオスの狭間で出現するのです。

ご清聴、ありがとうございました。

## 後註

- 1) Rüdiger Safranski: *Das Böse oder Das Drama der Freiheit*. (Frankfurt a.M. 2002 第三版) 13 頁より引用。以後、本書からの引用は、文中に頁数のみ記す。
- 2) Isaac Asimov: *Alle Robotergeschichten*. (Bergisch Gladbach 1982) 212 頁より引用。以後、本書からの引用は、文中に頁数のみ記す。なおアシモフのロボット作品からの引用を日本語に訳すにあたっては、小尾芙佐訳：『われはロボット [ 決定版 ]』(早川書房 2004) を参照・活用させていただいた。
- 3) Herbert W. Franke: *Spiegel der Gedanken. Science-fiction-Erzählungen*. (Frankfurt a.M. 1990) 126 頁より引用。以後、本書からの引用は、文中に頁数のみ記す。
- 4) Philip K. Dick: *Der unmögliche Planet*. (München 2002) 140 頁より引用。以後、本書からの引用は、文中に頁数のみ記す。
- 5) Heinrich Hauser: *Gigant Hirn. Ein utopisch technischer Roman*. (München 1962) 72 頁より引用。以後、本書からの引用は、文中に頁数のみ記す。
- 6) Herbert W. Franke: *Zarathustra kehrt zurück. Science-fiction-Erzählungen*. (München 1981 第三版) 190 頁より引用。以後、本書からの引用は、文中に頁数のみ記す。

## 講演をうけて

巽 孝之

(慶應義塾大学文学部教授)

私は基本的にアイザック・アシモフのロボット工学三原則を家電製品の三原則と考えている。したがって、アシモフ的なロボット観が、資本主義文明の繁栄を謳歌した1950年代パックス・アメリカナの時代に広く受け止められたことは、当然というしかない。この家電製品は人類にとっての新しい奴隷であり、アメリカにおけるロボットSFには南北戦争以後、解放された黒人奴隷の影がたえずつきまとっている。その意味でロボットSFは人種闘争の色彩を免れない。人類は長い間奴隷制を当然のものとしてきて、20世紀には人権問題が表面的には解決したようには見えるが、人種闘争はその後も存在し続け、新たな奴隷としての家電製品に希望を見出したという見方もできる。

アシモフの作品内において、ロボット工学三原則は作品ごとに変更・修正されているが、これは、人種問題をベースに合衆国の憲法がたびたび修正されていることと非常に似ている。映画化された『アイ・ロボット』では黒人問題がストレートに取り上げられ、結末にはリンカーンへのオマージュも捧げられている。以上が、本日の講演で取り上げられなかった、人種問題とのロボットSFとの関わりである。

エッセルボルン博士は啓蒙主義以後の思想を重んじ、そこからロボットの思想を導き出しており、もちろんメアリ・シェリーの『フランケンシュタイン』はロボットSFの起源のひとつである。この作品においてマッドサイエンティストが作り出した「フランケンシュタイン」というタイトルは、当時「現代のプロメテウス」と呼ばれていたベンジャミン・フランクリンを語源としている。にもかかわらず、最終的に「物語」としては主人と奴隷の関係に基づいており、これは啓蒙主義以後ではなくそれ以前の思想、すなわち封建主義的な伝統がクローズアップされてくる。そしてそこから「敵意」の概念が生じてくるのではないか。

本日のタイトルは「敵か味方か」であるが、この二項

対立のロジックにはひとつの罫がある。アメリカにおいては、植民地時代以来たえず更新されてきた二つの装置を考えてみよう。一つはフロンティアの概念で、西部が開拓された後は太平洋、日本、そして消費社会におけるマーケット、さらには宇宙、電腦空間へと、つねにアメリカはフロンティアを更新し続けた。もうひとつは仮想敵の概念で、自らの敵だと思ふもののイメージを絶えず新しくし、米ソの対立が解消したのであれば捏造してでもあらたな仮想敵をつくる。これこそがアメリカのダイナミズムであると言え、「敵か味方か」という対立軸を持ち込んでしまうと、このようなアメリカ的な二項対立にまんまとはまってしまう危険性がある。

そこで、「ロボットに敵意があるかどうか」という問題の根本として、SFというジャンルの定義を考えておく必要がある。日本におけるSFのパイオニアである柴野拓美氏によれば、SFとは、「人間理性の産物が、人間理性の制御を離れて自走することを意識した文学」ということになるが、私はこの定義に深く賛同する。

今日最後に言及されたロボットとコンピュータとは、容易に分けられるものではなく、人工生命が自身の意識を備え、人類全体の他者と化し、それに伴って人間の価値体系からみれば悪意や敵意、反抗精神のように映る反応を示すのは理論的な必然だろう。今回のコメントでは、そうした視点より、エッセルボルン氏の論点を補いたい。

まず、チェコの作家カレル・チャペックと、スタニスワフ・レムにおけるロボット観を取り上げる。ロボットという概念の起源であるカレル・チャペックの『ロボット』(1920)においては、当然ながらロボットの反乱が描かれ、その中では人種差別的な発言(「インディアンのような暮し」「犬だってあの連中よりは…」)が見られる。このように、チャペックからアシモフにいたるロボット観を考えるうえで、やはり人種の問題は避けられない。この問題が根本にある限り、「悪意」は機械文明の中から生まれたのではなく、ロボットの進化が人種論的な問題を寓意的になぞっていると言えよう。

そのような問題をまねがれているように思われるのは、スタニスワフ・レムの『砂漠の惑星』(1964)で、ここではサイバネティックス種族が登場する。彼らは人間の生態系をそのまま反復し、もはやロボットたちが自

走しているのかすら人間には認識不可能な状態である。この小説では、数百万年前に放置された自己増殖ロボットが万物の霊長となっており、ここまで来ると、私たちの倫理的価値体系でロボットの中に「悪意」や「善意」を想定できるのかは心もとなくなってくる。

アシモフとならんで先生が言及されたフィリップ・K・ディックは米ソ冷戦の渦中、自分の作品を研究しているマルクス主義批評家を恐れ、彼らを共産圏の手先であるとしてFBIに訴えたというエピソードが残っている。彼は左翼陣営を「敵意のある他者」と考えており、その意識が彼のロボット観に刷り込まれていたのではないか。ディックまでのヒューマニズムから翻って、今日のサイバーパンク以後のウィリアム・ギブスンやブルース・スターリングのポストヒューマニズム思想 (Posthumanism) では、人間とロボットという二項対立的な区別が明らかでなくなってくる。ここでは、むしろ生身の人間自身のほうが、有限なる肉体という地獄に堕ちた「肉人形」meat puppet すなわちロボットでしかないものと捉えられており、本質的な論理の逆転が見られる。つまり、サイバースペースに接続されているほうが、人間は感覚器官をフルに活用することができ、“Unplugged” な状態では、人間は地獄を経験しているに過ぎない。

「ロボットが人間ようになっていく」という世界観はアシモフが呈示しているものだが、イギリスの作家リチャード・コールドーの場合は、ナノテクノロジーを方法論として「人間を素材としてロボットが生まれる」、そして人類とロボット種族がナノテクノロジーを介して一種の擬似性的交渉を行って、生殖すらないという世界を描いている。ここでは人類とロボット種族の「混血」——中間種族が生まれ、ロボットと人間との関係において新たな可能性が思索されている。

では、ポストサイバーパンクの文脈ではどうか。女性SF作家エイミー・トムスの『ヴァーチャル・ガール』(1992)では、artificial intelligence が一種のソフトウェアとして電腦空間の中をさまよっていて、人間の肉体を「衣服」としてまといたいという欲望を明らかにする。彼らはコンピュータネットワークの中に生息する自意識を持つ電腦的存在であり、必ずしも形があるわけではない。生身の人間がロボットのハードウェアを身にまとい

パワーを増強したいと考える私たちの一般的な考え方とは逆の発想が見られる。

現代SF最高のハードSF作家のひとりグレッグ・イーガンの1997年の長篇『ディアスポラ』Diasporaでは、ソフトウェアから進化した種族が生まれる。もちろんこれは生物学的な発生ではないが、人間そっくりのロボット種族グライズナー (the Gleisner) は人間の肉体を衣服としてまとっている。現在の私たちのような、旧来どおり生身の肉体を備えている伝統的な人類はここでは肉体人 (the Flesher) と呼ばれもはや珍しい存在でとされる。この小説では、人間自身からロボットそしてコンピュータがいかに自走し、人間の空間から離脱 (diaspora) し、新たな自律的な世界を築き上げていくかが描かれている。今日のロボットSFにおける究極の形と言えるのではないだろうか。



## 講演を受けて

前野 隆司

(慶應義塾大学理工学部助教授)

エッセルボルン博士は、文学という視点から、ロボットの敵意、自由意志、意識の可能性について述べられた。これに対して、ロボット工学者である私からのコメントを試みたい。

敵意や自由意志は、意識があってはじめて生成されると考えられるので、意識について述べたい。ヒトの意識は解明可能か、あるいは、ロボットの意識を作れるか、ということに関しては、さまざまな議論が行われてきた。大雑把に言えば、以下のような考え方がホットだと思われる。

- ・心脳問題の哲学者：クオリアを伴う意識という現象は物理現象からは説明できないのではないかと、という立場に立つものが多い。Chalmers など。
- ・心理学者：Libet の実験…自由意志は存在しない。意識の時間は無意識に対して遅れている。
- ・コネクショニスト：ニューラルネットワークによって、意識という現象を伴う心は作れるのではないかと、ただし、作り方はわかっていない。

私も、心理学者やコネクショニストと同じ立場に立つ。つまり、意識というクオリアを作り出す方法はわかっていないので、現時点でクオリアを伴うロボットの心をつくることは難しいものの、機能としての心は作れる、と考える。したがって、ロボットに敵意を持たせることは可能である。

具体的には、私の著書「脳はなぜ心を作ったのか—私の謎を解く受動意識仮説（筑摩書房）2004年（日本語）」で述べたように、自律分散的な無意識を司るニューラルネットワーク群と、無意識的な情報処理結果を受け取り、体験し、エピソード記憶を作成するための受動的な機能である意識を作れば、与えられた目的を持たず、自由に知情意の情報処理を繰り返すロボットの心を作れると考えている。したがって、人類に敵意を持ったロボットを作らないような人類の合意や法の整備が将来必要となろう。

また、ヒトの自由意志はまやかしであり、上で述べたロボットの自由意志と変わらないと考える。私たちの意識は、自らの意図を持って自らが意思決定する存在であるかのように感じるが、実は、無意識的な情報処理結果を受け取って疑似体験しているだけであり、物理現象から独立した自由意志など存在しないのである。ただし、自由意志という現象が存在しないわけではなく、脳は意識や意思のクオリアを作り出している。

すなわち、ニヒリズムである。ヒトとロボットの意識・自由意志・敵意・好意は存在するのか、しないのか、と問うこと自体が無意味であり、いずれも脳が作った幻想に過ぎないが、存在しないわけではないのである。

### 参考

前野隆司、脳はなぜ「心」を作ったのか—「私」の謎を解く受動意識仮説、筑摩書房、2004年11月

<http://www.maeno.mech.keio.ac.jp/Maeno/consciousness/>

前野隆司、ロボットの心の作り方—受動意識仮説に基づく基本概念の提案—、日本ロボット学会誌23巻1号、2005年1月、pp. 51-62

<http://www.maeno.mech.keio.ac.jp/Maeno/consciousness/>



# ロボットの意識は作れるか？ —工学者の視点—

慶應義塾大学理工学部機械工学科

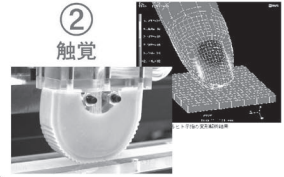
前野 隆司

1

## 前野隆司 自己紹介

- 1984年東工大機械工学科卒
- 1986年東工大修士課程修了
- 1986年キヤノン㈱入社
- 1995年慶大専任講師
- 1999年助教授、現在に至る
- 1990 - 1992年カリフォルニア大バークレー校客員研究員
- 2001年ハーバード大学客員教員

②  
触覚



③

ヒトとロボットの身体と心



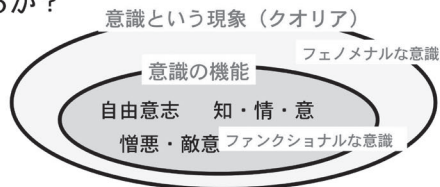
①  
超音波  
モータ



2

## エッセルボルン博士の問題提議

1. 憎悪や敵意など、人間と類似した特徴を有するロボットは作れるか？
2. そもそも、クオリア（意識の質感）を伴う意識という現象を持ったロボットを作れるか？



3

## 心についての議論

### 心脳問題の哲学的議論

- ・クオリア（意識の質感）を伴う意識という現象は、物理現象から説明できないのではないか（二元論, Chalmers）
- ・ニューラルネットワークにより作られている（一元論, Churchland, Denet）

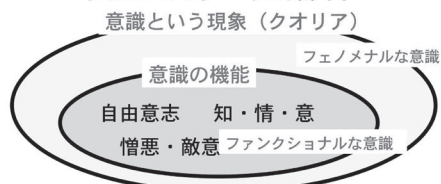
### 工学

- ・意識を持ったロボットは作れない（還元論）
- ・ニューラルネットワークにより作れる（コネクショニスト）



4

## 私は心身一元論者



- 現象的な意識を持ったロボットの作り方はまだ不明だが、ロボットの擬似意識（意識の機能）は作れる。
- 目的を自動生成する心の機能を作れば、敵意や好意を持ったロボットも作れる。

5

## ヒトの心についての私の考え方

「意識」や「自由意思」は受動的なのではないか？  
錯覚・幻想なのではないか？



脳はなぜ「心」を作ったのか—「私」の謎を解く受動意識仮説、筑摩書房、2004年

6

# 「知」「情」「意」は 受動的か？ 幻想か？

7

## 「知」は幻想か？ (知覚の例)

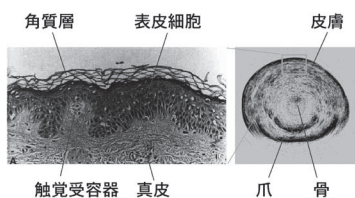
「痛み」や「触感」を  
指先で感じるというのは  
どういうことだろうか？



8

「痛み」「触感」  
触覚受容器→大脳皮質

→しかし、「痛み」「触感」を感じる  
のは、角質層（死んだ皮膚）の表面



9

痛い・熱い・つるつる・ざらざら・・・  
は皮膚で感じる



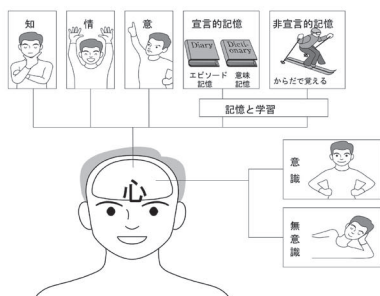
しかし、皮膚には  
角質層と触覚受容器があるだけ。  
情報処理は大脳。



触覚は幻想(錯覚)  
としか考えようがない。  
自己意識も幻想(錯覚)  
としか考えようがない。

10

## 「情」はそもそも受動的 感情・情動



11

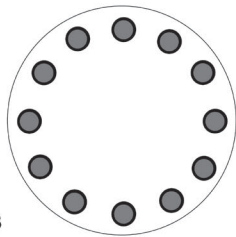
「知」「情」は受動的。  
では、「意」（意思・意図）は？

「よーい、ドン」  
とスタートするとき  
意識はすべての始まりか？

12

## 「指を動かそう」と意図する実験

- 1 筋肉への指令が発せられる瞬間  
(運動準備電位) ←脳に電極
- 2 「動かそう」と意図する瞬間



結果：  
1は2よりも  
0.35秒  
早い！  
「意」も幻想  
(受動的)

By B. Libet 1983

13

つまり、

『「知」「情」「意」は主体的である』  
と感じるのは幻想。

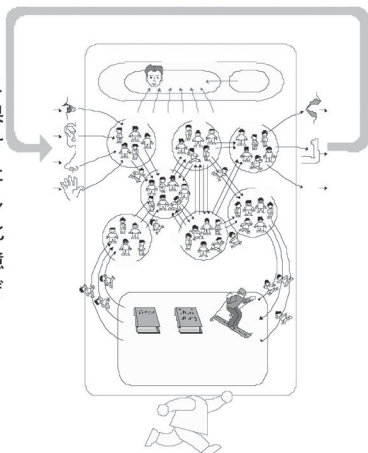
言い換えれば、  
すべて受動的。

ヒトの「意識」が主体的に行っていると思  
っていることは、すべて、  
「無意識」の結果を受動的に受け取って  
追体験しているだけではないのか？

14

## 受動意識仮説：

意識は、無意識下のこ  
びとたちがやった結果  
に受動的に注意を向け  
あたかも自らがやった  
かのように幻想体験し  
(つまり体験を単純化  
して) エピソード記憶  
するための存在に過ぎ  
ない。



15

現代はニヒリズムの時代：

歴史には明確な目的や到達点は  
存在しない。(リオタール)



自由意志も存在しない。  
(物理現象や無意識から独立した意思は存在し  
ない。)  
我思う、しかしそれは幻想である。(前野)

そもそも人間はロボットとさほど違わな  
い存在なのではないか。

16

## アイボ (ソニー)



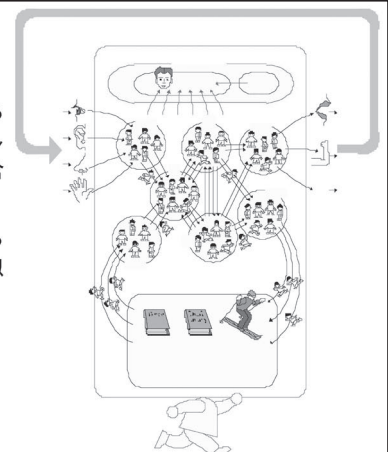
ペットロボット  
＝動くコンピュータ  
「知」「情」「意」の原型  
＝目的を持たないロボット  
受動的意識を付加すれば擬似  
的心は作れるはず

17

目的を自立的に生  
成するロボット

無意識下のこびとたち  
の相互作用の結果とし  
て知情意が(憎悪も含  
めて)作り出される。

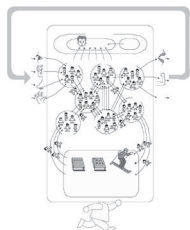
意識はその結果を自ら  
が行った体験と(擬似  
的に)感じる。



18

## 心を持ったロボットの問題点

- ・無意識の小びとたちの  
作り方
  - ・作りこみ
  - ・進化的設計
- ・そもそも必要か？
- ・倫理的・宗教的問題
- ・法の整備



19

現代はニヒリズムの時代：

歴史には明確な目的や到達点は  
存在しない。（リオタール）



自由意志も存在しない。  
（物理現象や無意識から独立した意思は存在し  
ない。）  
我思う，しかしそれは幻想である。（前野）

このような不確かな現代をいかに生きるか。  
が問われている時代

20



# 質疑応答

識名 巽先生、前野先生、ありがとうございました。それでは、両先生のお話を受けて今度はエッセルボルン先生よりコメントをいただきます。

エッセルボルン まず巽先生のお話コメントします。ロボットの三原則が家電の三原則とも言えるとおっしゃいましたが、それに対しては私も同感です。先日、三菱重工のロボット WAKAMARU を見学した際に、担当者に同様のことを伺いましたが、有効というお返事をいただいていたばかりです。

また人間とロボット、あるいはコンピューターの話をする際に、人種問題が必ずそこに介在する、もしくは支配者とプロレタリアートといった階級問題や階級闘争が、ロボットが登場する SF の世界に介在するという点については、さまざまなレベルで解釈されることを指摘しておきます。たとえば種族や労働者階級のような違いであったり、エイリアンのような全く未知の生物であったりします。各作品において、どのようなレベルで人種あるいは階級問題が扱われているか、注意を払う必要があると思います。

またレムの作品に対して、私は決して否定的ではありませんので、誤解のないようお願いします。私がレムを引用したのは、レムが「人工知能が登場し、やがてそれが人間に取って代わる存在になる」と言っているからです。レムにとって理性は信じるに足るものでした。そして、コンピューターは理性が結集した優れたものであると考えています。ですから、優れたコンピューターが人間に取って代わる存在になるのだと指摘しているのです。

最後に挙げられた「サイバーパンク」「ポストサイバーパンク」の展開については、とても興味深く拝聴しました。サイボーグは人間と機械の違いを解消する存在として描かれることがあります。ただし私は、その解消方法については懐疑的です。つまり、とても幻想的な試みではないかと思っています。ソフトウェアから生まれる種族、あるいはサイバー上に存在する新しい種族が生まれるという例も挙げられましたが、それらについては、日本の『攻殻機動隊 (GHOST IN THE SHELL 訳注：ドイツ国内でも有名。TV で放映されている)』という作品で

も、取り扱われています。ソフトウェアから生まれた種族などが自律性をもったときに、人間はそれらを悪として捉えるのではないか、ということが予想されます。これは、ストーリー展開としては非常に面白いと思います。

次に、前野先生のお話コメントさせていただきたいと思います。まず、ロボット工学の専門家の立場から、人間と同じような感情、ここではクオリアを感情と捉えていらっしゃると思いますが、それらを模倣できるかという点に、私も興味があります。これは人間の意識がどのように機能しているのか、ということを探り当てるための実験であり、とても意義のあることだと思います。コンピューターあるいはロボットを用い、簡略化した形で人の意識がどのように生まれるのか、作られているのかを探ることができるでしょう。前野先生のお話を伺って、それが近い将来に実現するだろうと思いました。ただ、最後に前野先生も指摘されたように、こうした未知の部分の解明には、常に倫理的な問題がついて回ると思います。たとえば、人間に似た生物をつくるということについては、アシモフも言っていますが、その生物に法的な人としての権利を与えるのか、といった問題が浮上するでしょう。

巽 エッセルボルン先生、大変貴重なご意見をありがとうございました。レムに対しては、先生がおっしゃる通りだと思います。また、架空の書評を集めた短編集『完全な真空』という作品や、架空の本の序文を集めた『虚数』という作品がありますが、やはりコンピューターが小説



巽孝之氏

を書いたり、人類を圧倒する知性をもっていたり、多様な力を備えた人工生命が登場します。

また、お話の中でサイボーグに関する問題がありましたが、サイバーパンク SF はウィリアム・ギブスの『ニューロマンサー』が1984年に発表したのちに大きな運動となり、その直後1985年にカリフォルニア大学サンタクルス校のダナ・ハラウェイが『サイボーグ宣言』という、のちに文化研究のバイブルとも言われる論文を発表します。その論文の中では、サイボーグの典型としてシリコンバレーで働く有色人女性が登場します。これは、ハイテク環境ととても密接した民族問題なのです。ここでダナ・ハラウェイが言っているサイボーグの概念にはクレオールという概念も入ってくる。そして、論文でダナ・ハラウェイは、我々はすでにサイボーグなのだ、という言い方をしています。本日の発表では、ロボットとコンピューターという対置がありましたが、1970年以前であればロボットとサイボーグ、自体が別々の存在であったのが、1980年以降はロボットとコンピューター、そしてサイボーグは限りなく近い存在として扱われるようになっていきます。また人間もロボットなのではないかといった話も耳にします。フィリップ・K・ディックの原作で有名になった映画『ブレード・ランナー』にも有機アンドロイドが登場します。いわゆるメカニックではないわけです。今日では、ロボットとサイボーグの区分がなくなってきたのではないのでしょうか。

**前野** エッセルボルン先生のお話にありました「人の意識をどのようにして作るか」これを探るためにロボットが使えるのではないか、そして法律や倫理が問題になるのではないか、という点については、まさにその通りだと思います。ただし、私の話の中では、原理的にはロボットの心を作れるようになった、というだけであって倫理の問題、法律の問題が出てくるのは、ずいぶん先のことだと思います。ですから、期待をもたせてしまったかもしれませんが、ロボットが人間のようなことをするのはまだ100年、200年先なのではないかと思っています。

ですから私は、SF作品はうらやましい、と思いつつ地道にロボットに関する研究を続けていかなければならないと思っています。



前野隆司氏

**識名** エッセルボルン先生、巽先生、前野先生ありがとうございました。それでは、次に会場の方々からの質問に入りたいと思います。

**学生（理工学部一年生）** 私は理工学部で、ドラえもんを作ろう、と思い入学しました。先ほどのお話ですと、私がドラえもんを作ることには出来ないように思いました。実際に、ドラえもんを作りたいと思いますが、それは難しいのでしょうか。

**前野** ロボットの開発技術がドラえもんのレベルに至るには、あと何年かかるのか定かではありませんが、ドラえもんを作るまでには、まだまだやらなくてはならない研究が多くあることはわかっています。ですから、ドラえもんそのものを作ることが出来なくとも、ドラえもんを作るための第一歩を切り開く研究者になれる可能性は大いにあると思っています。ですから、がんばっていただければと思います。

**学生（文学研究科修士課程一年生）** 大変興味深いお話をいただきまして、ありがとうございます。ロボットが人間に悪意をもつかどうか、という問いに対しては、人間の視点からロボットが脅威に見えるかどうか、に関係すると思いました。また、人間がロボットに対して敵意をもつケースとして考えられるのは、人間を支配しようとする、攻撃を仕掛けてくることなどが挙げられる



と思います。私は、その脅威となる原因の中にロボットが自身を自己複製していく、というものが入ってくると思います。人間もプログラムされた機械であり、自己複製、増殖がその目的として埋め込まれています。エッセルボルン先生が述べたような、ロボットの人間に対する敵対的な行動として、この自己複製は入るのでしょうか。

**識名** 自己複製の問題はとても面白い視点ですね。人間をはじめとする他の生き物は自己複製、繁殖をしていますが、それをロボットが行うようになる。やがて、ロボットたちは独自の道を進んでいくことになるでしょうし、文学的にどのように捉えられるか興味深い点です。エッセルボルン先生いかがでしょうか。

**エッセルボルン** 私の講演内容を正しくご理解いただきありがとうございます。文学的に考えて、ロボットが人間に悪意をもっているかについては、人間の観点からしか判断できないことだと思います。この場合、人間の観点がまず絶対的な見方であるか否かということも考えなければいけません。レムの『砂漠の惑星』は、これと異なった見方を示している作品でしょう。私は、ロボットが自己複製することだけでは、人間にとって悪意、あるいは脅威となる存在になるとは思いません。仮に、ロボットが独自の種族をもつとするなら、彼らには自己保持の権利が生まれると思います。つまりロボットなりの倫理が存在するようになるでしょう。このロボット独自の倫理は、非人間的な倫理ですから、人間がもつ善悪の価値判断の基準を超えた倫理となります。人間の倫理とロボットの倫理が共存するということになると思います。

**学生（理工学部三年生）** 本日の講演から、やはりロボット対人間という構造が強く印象に残りました。そして、それを語るうえで、階級意識や人種差別といったことが含まれることも理解できました。

さて、以前に映画になった『A.I.』の最後のシーンでは、ロボットがさらに進化したロボットたちに囲まれているというシーンがありました。それに関連して、人間の意識を機械という体に移植することで、人間がロボットに進化するという概念は考えられるのでしょ

うか。ロボットの位置づけが、人間が進化した形と捉えると人間とロボット間の逃走も減ると思うのですが、いかがでしょうか。

**識名** ポストサイバーパンクの話の中にもご質問にあるようなことが書かれていると思いますが、いかがでしょうか。

**巽** 『A.I.』は2001年くらいに発表された作品で、スタンリー・キューブリックの遺作をスピルバーグが完成させたものです。それを見る限り、スピルバーグがいかにキューブリックに憧れていたのがわかります。

さて、人間がロボットに進化するという概念は、グレッグ・イーガンの『ディアスポラ』という作品にも記されています。30世紀の未来に、人間が肉体を捨て、自分自身のコピーをソフトウェアに移してしまっているという世界です。そして、そのソフトウェア自体が自身の精神種子になっているのです。それにより、人間はデジタル空間上で進化していくのです。また、エッセルボルン先生もご覧になられたという押井守の『攻殻機動隊』より以前に発表されたエイミー・トムソンの『バーチャルガール』もアーティフィシャル・インテリジェンス、人工生命が人間の肉体、あるいは機械の肉体をとりあえずまとったかたちで登場しますね。『ディアスポラ』は、拠って立つはずの人間の肉体を捨ててしまい、人間はソフトウェアとして疑似的ながら永遠の命を得るという内容になっています。海老原君の質問にもありました自己複製に関連しますが、人間から進化したロボットが独自の種族、いわゆるポストヒューマンへ進化していくという途方もない物語が展開されていくのだと思います。

**学生（理工学部一年生）** 私は、ドラえもんよりも御茶ノ水博士になりたいと思い、理工学部に入学しました。

さて、お話の中にありました『攻殻機動隊』については、知性をもった人工知能が敵として扱われていたり、アーサー・G・クラーク『2010年宇宙の旅』では知性をもったコンピューターが敵として扱われています。そうした人間とコンピューターの関係の言い方を換えれば、人間が考え方の異なった他人と、どのように付き合うか、と

いうことにもつながると考えます。そのように考えるとロボットの三原則は、ロボットだけではなく。人間にも当てはまるのではないか、と思いますがいかがでしょうか。

**エッセルボルン** ロボットの三原則が人間にも当てはまるか、ということですが、アシモフはこの概念はカントから受け継いだと言っています。私は、本当にこれが適用できるかは疑問に思います。ですが、とりわけ人間対ロボット、あるいはコンピューターなどの人工物との対立は、まさに異文化間の対立と捉えることができます。私たちがロボットやコンピューターをどれだけ受け入れることができるのかという問いは、私たちが異文化を尊重して受け入れることができるのかという問いと重なります。ご質問にあったような、人間と異なる生物と人間の対立というのは、たとえばヨーロッパとイスラム文化との対立と同様に考えられると思います。

**識名** エッセルボルン先生、ありがとうございました。もう少し話を聞きたいところですが、これにて講演会を終了致します。コメンテーターとしてお越しいただきました巽先生、前野先生、ありがとうございました。また、最後にエッセルボルン先生より感想を頂戴したいと思います。

**エッセルボルン** 巽様、前野様より貴重なコメントをいただいたことに、お礼を申し上げます。また会場の皆様からのご質問にも感謝します。今後、皆様からいただいた質問を参考にして、私の研究をさらに進めたいと思います。また、司会をご担当いただきました識名様、また翻訳を担当してくださいました石原様、ありがとうございました。日本でこのように関心の高い聴衆の前で、ドイツ文学やSFについて語らせていただく機会を与えてくださったことに心からお礼申し上げます。ありがとうございました。

# プロフィール

## 講師

エッセルボルン, ハンス (Prof. Dr. Esselborn, Hans)

ケルン大学哲学部教授

### 専門分野：ドイツ語ドイツ文学

テュービンゲン、ミュンヘン、ケルン大学で学ぶ。主専攻はドイツ語ドイツ文学、副専攻はフランス語フランス文学および哲学。在学中、パリに留学。1979年、ケルン大学哲学博士号(Dr.phil.)取得。1987年、ジャン・パウルに関する研究 *Das Universum der Bilder. Die Naturwissenschaft in den Schriften Jean Pauls* (Tübingen 1989) で教授資格 (habil.) 取得。フランスの大学 Université Nancy 2, Paris XII, Lyon 3 で訪問教授を務めた。ドイツ啓蒙主義文学およびジャン・パウル研究と並行して、SFに関する研究にも力を入れており、特に2000年からは毎夏、「SFの夕べ」Kölner Science Fiction-Nacht を主催し、ドイツ語圏のSF作家を招いている。また氏が編集した独・仏二国間共同コロキウム講演集 *Utopie, Antiutopie und Science Fiction im deutschsprachigen Roman des 20. Jahrhunderts* (Würzburg 2003) には、20世紀ドイツSFに関する研究も多く含まれる。日本学術振興会・外国人招聘研究者・短期プログラムに採用され、2005年11月中、日吉キャンパスに滞在した。

## コーディネーター・講演監訳および字幕作成

石原 あえか

慶應義塾大学商学部助教授

### 専門分野：ドイツ語ドイツ文学

慶應義塾大学後期博士課程在学中の1995年秋より渡独、1998年春、ケルン大学で哲学博士号(Dr.phil.)取得。主専攻はドイツ語ドイツ文学、副専攻は日本学と教育哲学。指導教授はH. D. イルムシャー氏で、エッセルボルン氏とは同門。2002-2004年、日本学術振興会海外特別研究員として再びケルンに滞在したのがきっかけで、今回のシンポジウムを計画・実現の運びとなった。現在、教養研究センター所員兼務。主著に *Makarie und das Weltall. Astronomie in Goethes „Wanderjahren“* (Köln/Weimar/Wien 1998), *Goethes Buch der Natur: Ein Beispiel der Rezeption naturwissenschaftlicher Erkenntnisse und Methoden in der Literatur seiner Zeit.* (Würzburg 2005)がある。2005年秋、ドイツ学術交流会DAADよりJacob-und-Wihlem-Grimm-Förderpreis受賞。

## コメンテーター

### 巽 孝之

慶應義塾大学文学部教授

#### 分野：アメリカ文学

1955 年東京生まれ。コーネル大学大学院修了 (Ph.D., 1987)。慶應義塾大学文学部教授。アメリカ文学専攻。主著に『サイバーパンク・アメリカ』（勁草書房、1988 年度日米友好基金アメリカ研究図書賞）、『ニュー・アメリカニズム』（青土社、1995 年度福澤賞）、編訳にラリー・マキャフレイ『アヴァン・ポップ』（筑摩書房、1995 年）、共編著に『事典 現代のアメリカ』（大修館書店、2004 年）ほか。

## 前野 隆司

慶應義塾大学理工学部機械工学科助教授

#### 専門分野：ロボット工学

慶應義塾大学理工学部機械工学科助教授。1984 年東京工業大学卒業、1986 年同大学大学院修士課程修了後、キヤノン（株）入社。超音波モータや精密機械の研究開発に従事。1995 年慶應義塾大学理工学部専任講師、1999 年より現職。1990~1992 年カリフォルニア大学バークレー校 Visiting Industrial Fellow、2001 年ハーバード大学 Visiting Scholar。博士（工学）。大学移籍後は、新型アクチュエータ、ヒトの触覚、触覚センサ・触覚提示デバイス、遠隔操作型ロボット、生物の進化シミュレーション、進化・生命化するロボットなど、ロボットとヒトの研究に従事。

## 司会

### 識名 章喜

慶應義塾大学商学部教授

#### 専門分野：ドイツ近現代文学、ドイツ語圏の SF

慶應義塾大学商学部教授。1984 年、東京大学大学院人文科学研究科博士課程満期退学。1985 から 87 年まで DAAD（ドイツ学術交流会）奨学生として、バンベルク大学に留学。助手、助教授を経て 2000 年より現職。主要著作に『ユートピアの期限』（慶應義塾大学出版会、2002 年、共著）、『生命の教養学』（慶應義塾大学出版会、2005 年、共著）、訳書にザフランスキー『E.T.A. ホフマン』（法政大学出版局、1994 年）、N. ボルツ『ゲーテンベルク銀河系の終焉』（共訳、法政大学出版局、1999 年）、ヘルマン『理想郷としての第三帝国』（2002 年、柏書房）、レーマン『森のフォークロア』（共訳、法政大学出版局、2005 年）など。

慶應義塾大学教養研究センター主催  
慶應義塾におけるドイツ年記念公開セミナー

敵か味方か  
——ロボットをめぐる文化——

---

2006 年 3 月 31 日発行  
編集・発行 慶應義塾大学教養研究センター  
代表者 横山千晶

〒 223-8521 横浜市港北区日吉 4-1-1  
TEL 045-563-1111 (代表)  
Email lib-arts@hc.cc.keio.ac.jp  
<http://www.hc.keio.ac.jp/lib-arts/>

©2006 Keio Research Center for the Liberal Arts  
著作権者の許可なしに複製・転載を禁じます。  
ISSN 1880-3628  
ISBN 4-903248-06-2



