

# Der Südkoreanische Roboterkodex und die Robotergesetze von Isaak Asimov

Fabian Hundertmark Matrikel-Nummer: 1769284

5. September 2007

## Inhaltsverzeichnis

<b>1</b>	<b>Was werde ich tun?</b>	<b>1</b>
<b>2</b>	<b>Der Südkoreanische Roboterkodex</b>	<b>2</b>
<b>3</b>	<b>Die Robotergesetze von Isaak Asimov</b>	<b>2</b>
<b>4</b>	<b>Asimovsche Robotergesetze als moralische Regeln?</b>	<b>3</b>
4.1	Inkompatibilistische Ansätze für moralische Verantwortlichkeit . . . . .	3
4.2	Kompatibilistische Ansätze für moralischer Verantwortlichkeit . . . . .	5
4.2.1	Harry G. Frankfurts Position . . . . .	5
4.2.2	Daniel C. Dennetts Position . . . . .	6
4.2.3	John Martin Fischers „Reasons-responsive Mechanism View“ . . . . .	7
<b>5</b>	<b>Asimovsche Robotergesetze als Funktionen</b>	<b>7</b>
<b>6</b>	<b>Fazit</b>	<b>8</b>
<b>7</b>	<b>Quellen</b>	<b>8</b>

## 1 Was werde ich tun?

„Asimov und die Realität: Auf dem Weg zu einer Roboter-Ethik“<sup>1</sup>: unter dieser Überschrift findet sich seit 25. August 2007 auf der Website derStandard.at ein Artikel, in dem es darum geht, dass in Süd-Korea Ärzte, Psychologen und Roboterentwickler an einem „ethische[n] Kodex [für] die landeseigene Robotertechnologie“<sup>2</sup> arbeiten. „Der Kodex soll sicherstellen, dass der Mensch die Kontrolle über die Roboter behält und sie jederzeit identifizieren kann. Er soll verhindern, dass die Maschinen

---

<sup>1</sup>siehe: <<http://derstandard.at/?url=/?id=3000701>>, 25.8.2007, 12:00.

<sup>2</sup>ebenda

für illegale Zwecke genutzt werden. Daten, die von Robotern erfasst werden, sollen geschützt sein.“<sup>3</sup> Sogar eine Ausnahmeregelung für militärische Roboter soll es geben.

Auf dem Bild, welches zu dem Artikel gehört, ist ein Roboter abgebildet, welcher eine menschenähnliche Puppe auf den Armen hält. Die Bildunterschrift lautet: „Edel sei der Roboter, hilfreich und gut.“<sup>4</sup> Dieses und die Überschrift eines Artikels mit ähnlichem Inhalt („Roboter sollen ethisches Verhalten lernen“<sup>5</sup>) suggerieren, dass das Ziel des südkoreanischen Kodex oder auch der fiktiven asimovschen Robotergesetze – auf die in dem Artikel öfters Bezug genommen wird – darin liege, Robotern das moralische Handeln beizubringen.

Doch geht es den Südkoreanern wirklich darum? Und würden sie dieses Ziel erreichen, wenn sie den Robotern die asimovschen Robotergesetze „einbauen“ würden? Dieser Essay soll sich mit beiden Fragen beschäftigen. Zuerst werde ich zeigen, dass es den Südkoreanern nicht um moralische Roboter geht. Dann werde ich mich mit den Robotergesetzen von Isaak Asimov auseinandersetzen und zeigen, dass auch diese Gesetze nicht das Ziel moralischer Roboter haben können. Am Schluss werde ich zeigen, was die Asimovschen Robotergesetze tatsächlich darstellen.

## 2 Der Südkoreanische Roboterkodex

Der Südkoreanische Roboterkodex hat nicht das Ziel, dass künftige Roboter moralisch handeln. Es handelt sich viel mehr um Regeln, die beim Erbauen von Robotern beachtet werden sollen. So könnte beispielsweise beschlossen werden, dass künftig produzierte Roboter – wie Mobiltelefone – von der Polizei ortbar sein sollen. Vermutlich werden auch bestimmte Standards festgelegt, die verhindern sollen, dass Hacker die Kontrolle über die Roboter bekommen können.

Die in Südkorea erarbeitete Roboterethik lässt sich also wie folgt kategorisieren: „Roboterethik muss [...] Teil der menschlichen Ethik werden.“<sup>6</sup> So wäre es durchaus möglich, dass der erarbeitete Kodex in Südkorea Gesetz wird. In Deutschland würde ein solches Robotergesetz beispielsweise als bundesweit geltendes Gesetz, wie das Straßenverkehrsgesetz, umgesetzt werden.

Wenn ich auch hier von Robotergesetzen spreche, ist es wichtig, den Unterschied zu den Asimovschen Robotergesetzen zu beachten.

## 3 Die Robotergesetze von Isaak Asimov

Der Science-Fiktion-Autor Isaak Asimov beschrieb 1941 erstmals in der Kurzgeschichte „Run-around“ die sogenannten Robotergesetze. Diese stellen – in seiner Geschichte – hierarchisch aufgebaute Verhaltensregeln für alle Roboter dar:

---

<sup>3</sup>ebenda

<sup>4</sup>ebenda.

<sup>5</sup><<http://www2.onnachrichten.t-online.de/c/12/10/34/14/12103414.html>>, 25.8.2007, 19:00

<sup>6</sup>Asimov und die Realität: Auf dem Weg zu einer Roboter-Ethik. siehe: <<http://derstandard.at/?url=/?id=3000701>>, 25.8.2007, 12:00.

„ Fangen wir mit den drei Grundregeln des Roboterdienstes an – jenen Grundregeln, die am allertiefsten eingegraben in das positronische Gehirn eines jeden Robots! Wir haben also –

*erstens*: Ein Robot darf kein menschliches Wesen verletzen oder durch Untätigkeit gestatten, dass einem menschlichem 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 Regel Eins oder Zwei kollidiert. “<sup>7</sup>

Der größte Unterschied zwischen den asimovschen Robotergesetzen und den oben betrachteten möglichen südkoreanischen Robotergesetzen lässt sich wie folgt beschreiben: Die südkoreanischen Robotergesetze gelten für Menschen, die Roboter bauen. Die asimovschen Robotergesetzen gelten für Roboter im Einsatz.

Doch ist der Geltungsbereich der Gesetze nicht der einzige Unterschied. Der Gesetzesbegriff ist bei den asimovschen Robotergesetzen ein vollkommen anderer.

## 4 Asimovsche Robotergesetze als moralische Regeln?

Die drei Regeln des Roboterdienstes werden in der asimovschen Welt von jedem Roboter befolgt. Den Unterschied zu unseren menschlichen Gesetzen macht folgendes Beispiel klar:

Ein Roboter ist mit einem Menschen alleine, weit weg von jeglicher Zivilisation. Der Mensch will den Roboter mit einer Feuerwaffe zerstören. Doch er hält sie falsch herum, sodass er sich selbst töten würde, wenn er auf den Auslöser drückt. Befolgt der Roboter nun die Regel eins und warnt den Menschen, überlebt der Mensch und der Roboter wird zerstört. Befolgt er sie nicht, stirbt der Mensch und der Roboter existiert weiter. Auch wenn der Roboter für das Nichtbefolgen der Regel nie eine Strafe zu erwarten hätte, wird er sie befolgen.

Befolgt ein Roboter also die Gesetze, tut er dies in diesem Fall nicht aus Angst vor Strafe. Der Grund seiner Handlung scheint auf den ersten Blick viel eher etwas zu sein, was man bei einem Menschen Moral nennen würde. Doch können Roboter, die die Robotergesetze befolgen, überhaupt moralisch verantwortlich für diese Taten sein? Dieser Frage werde ich im Folgenden nachgehen, indem ich sie für die verschiedenen Ansätze zur Definition moralischer Verantwortlichkeit überprüfe.

### 4.1 Inkompatibilistische Ansätze für moralische Verantwortlichkeit

Da die philosophische Debatte über moralische Verantwortlichkeit eng mit der Willensfreiheitsdiskussion zusammenhängt, lassen sich auch hier die Positionen in kompatibilistische und inkompatibilistische

---

<sup>7</sup>Asimov, Isaac: Ich, der Robot [Dt. Übers.: Otto Schrag], München : Heyne , 1978. S. 34f. Hervorhebungen und Zeilenumbrüche vom Autor eingefügt.

bilistische einteilen.

So behaupten die Inkompatibilisten, dass ein Wesen nicht für eine Handlung verantwortlich gemacht werden kann, wenn es determiniert war diese Handlung durchzuführen. Was es mit diesem Determinismus auf sich hat, lässt sich anhand von John Conways „Game of Life“ verdeutlichen:

Bei dem „Game of Life“ handelt es sich um einen zellulären Automaten. Dieser besteht aus einem gitterförmigen, im Idealfall unendlich großen Spielfeld. Jedes Quadrat dieses Spielfelds ist entweder aktiv oder inaktiv. Die Aktivität aller Felder ändert sich rundenweise und nach folgenden Regeln:

1. Wenn exakt zwei Nachbarn eines Feldes im letzten Zyklus aktiv waren, behält es seinen Zustand bei.
2. Wenn exakt drei Nachbarn eines Feldes im letzten Zyklus aktiv waren, wird es aktiv.
3. Trifft weder Regel 1 noch Regel 2 zu, wird es inaktiv.

Hat man nun ein komplett deaktiviertes Spielfeld und stellt durch die Aktivierung einiger Felder einen Anfangszustand her, folgen aus diesem Zustand und den drei Regeln alle weiteren Zyklen.<sup>8</sup> Stellt man also im Zyklus  $T_1$  das einfache Muster eines sogenannten

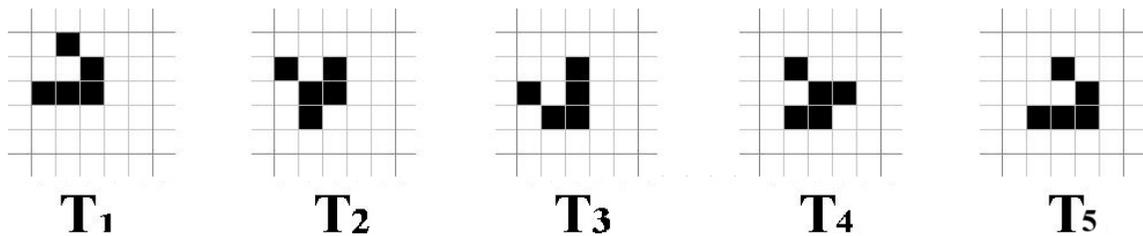


Abbildung 1: John Conways „Game of Life“: Ein Glider in Aktion (schwarze Felder: aktiv; weiße Felder: inaktiv)

Glider her, wird daraus im Zyklus  $T_5$  zwangsläufig dasselbe Muster erneut entstehen, nur um ein Feld nach rechts und ein Feld nach unten verschoben. Dies gilt natürlich nicht, wenn „von außen“ auf den Glider eingewirkt wird, da dieser beispielsweise mit einem anderen Glider „zusammenstößt“.

Ein determiniertes Wesen verhält sich genau wie ein solcher Glider: Kennt man den Zustand dieses Wesens zu einem beliebigen Zeitpunkt, die Gesetze, nach denen dieses Wesen funktioniert und alle äußeren Einflüsse, so ergibt sich der Zustand dieses Wesens zu einem späteren Zeitpunkt zwangsläufig.

Ein Roboter, der die Asimovschen Robotergesetze befolgt, ist in diesem Sinne determiniert: Befindet sich der Roboter in einer Situation, in der ein Mensch in Gefahr ist, so wird der Roboter

<sup>8</sup>vgl. Clark, Andy. *Mindware: An Introduction to the Philosophy of Cognitive Science*, Oxford: Oxford University Press, 2001. S. 52

zwangsläufig handeln – es sei denn er ist kaputt.

Ist man also der Meinung, dass nur indeterminierte Wesen moralisch verantwortlich sein können, spricht man Robotern, die den drei Regeln gehorchen, automatisch moralische Verantwortlichkeit ab.

## 4.2 Kompatibilistische Ansätze für moralischer Verantwortlichkeit

Im Gegensatz zu inkompatibilistischen Ansätzen moralischer Verantwortlichkeit, gehen kompatibilistische Ansätze davon aus, dass auch determinierte Wesen moralisch verantwortlich sein können. Ich werde im folgenden drei Beispiele aufführen und überprüfen.

### 4.2.1 Harry G. Frankfurts Position

Harry G. Frankfurt vertritt beispielsweise eine solche kompatibilistische Position. Für ihn gilt als einzige Bedingung für moralische Verantwortlichkeit, „dass die [betreffende] Handlung von einem Wunsch verursacht wurde, mit der sich der Handelnde identifiziert.“<sup>9</sup>

**Intentionale Zustände** sind zum Beispiel Überzeugungen, Wünsche, Hoffnungen und Befürchtungen. Für Frankfurts Theorie ist es also unabdingbar, dass ein moralisch verantwortliches Wesen tatsächlich solche Zustände hat. Im Folgenden werde ich kurz die drei gängigsten Theorien darstellen und überprüfen, welche es Robotern gestattet, Intentionale Zustände zu haben:

**Jerry Fodor** geht davon aus, dass Intentionale Zustände tatsächlich in den Wesen die sie haben, realisiert sind. Es soll sich bei ihnen um zusammengesetzte mentale Repräsentationen handeln.<sup>10</sup> Es ist nicht unplausibel anzunehmen, dass auch Roboter mentale Repräsentationen haben können. Hat also Fodor recht, kann auch ein Roboter Wünsche haben.

**Die Churchlands** gehen davon aus, dass es Intentionale Zustände gar nicht gibt, da sich solche mentalen Repräsentationen – wie sie Fodor fordert – in der Neurowissenschaft oder in der Physik nicht finden lassen.<sup>11</sup>

Geht man davon aus, dass die Churchlands mit ihrer Theorie recht haben, könnte man die These vertreten, dass *nur* Roboter oder Computer Intentionale Zustände besitzen, da sich die geforderten mentalen Repräsentationen in ihnen finden lassen. Beispielsweise könnte man einen Wunsch mit einem speziellen Speicherbereich auf einer Festplatte identifizieren.

---

<sup>9</sup>Schulte, Peter: Philosophie verständlich : Willensfreiheit : Harry G. Frankfurt. <<http://www.philosophieverstaendlich.de/freiheit/modern/frankfurt.html>>, 31.8.2007, 11:23.

<sup>10</sup>vgl. Clark, Andy. *Mindware: An Introduction to the Philosophy of Cognitive Science*, Oxford: Oxford University Press, 2001. S. 43f.

<sup>11</sup>vgl. ebenda S. 44ff.

**Daniel C. Dennett** geht, im Gegensatz zu Fodor und den Churchlands, davon aus, dass intentionale Zustände nicht als interne physikalische Zustände realisiert sein müssen, um zu existieren. Es reicht schon aus, dass sich das Verhalten eines Wesens mit intentionalen Zuständen erklären lässt um zu sagen, dass dieses Wesen tatsächlich intentionale Zustände hat.

Intentionale Zustände sind also, ähnlich wie der Äquator oder der Erdmittelpunkt, keine Teile der physikalischen Welt, sondern abstrakte, unverzichtbare Annahmen.<sup>12</sup>

Hat Dennett recht, reicht es aus, dass ein Roboter so handelt als hätte er Wünsche um zu sagen, dass er tatsächlich Wünsche hat. Es ist also auch nach dieser Theorie möglich, dass ein Roboter Intentionale Zustände hat.

Die gängigsten Theorien scheinen also nicht auszuschließen, dass Roboter Wünsche haben. Fraglich ist natürlich, ob es sich bei den Robotergesetzen oder bei konkreten, situationsbezogenen Handlungsplänen, die von diesen abgeleitet sind um Wünsche handelt.<sup>13</sup> Akzeptiert man dies aber, muss sich – nach Frankfurt – ein Roboter nur mit diesen Wünschen identifizieren, damit er für eine Handlung nach den drei Regeln moralisch verantwortlich ist. Für unser Beispiel von dem Menschen, der den Roboter zerstören will, heißt das:

Identifiziert sich der Roboter in der gegebenen Situation mit der Regel eins, ist er moralisch dafür verantwortlich, dass er den Menschen gewarnt hat. Identifiziert sich der Roboter in der gegebenen Situation nicht mit der ersten Regel und will den Menschen eigentlich nicht helfen, ist er auch nicht für seine Handlung verantwortlich.

#### 4.2.2 Daniel C. Dennetts Position

Im Gegensatz zu Frankfurt sieht Dennett die Identifikation eines Wesens mit dem handlungsverursachendem Wunsch nicht als notwendige Bedingung für moralische Verantwortlichkeit an. Er betrachtet eher folgende Frage: Wann ist es sinnvoll/vorteilhaft/zweckmäßig Wesen für ihre Taten verantwortlich zu machen? Der Zweck moralischer Verantwortlichkeit liegt ihm zufolge vor allem im Vorbereiten von Sanktionen, die wiederum positive Veränderungen im Verhalten verursachen sollen.

Lässt sich also das Verhalten eines Wesens positiv dadurch beeinflussen, dass man es für seine Taten sanktioniert, ist es auch moralisch für diese Taten verantwortlich.<sup>14</sup>

Da aber die Robotergesetze nicht durch Sanktionen des Roboters geändert werden können, ist es auch nicht sinnvoll einen Roboter für Taten, die er aufgrund der Robotergesetze durchführte, zu bestrafen. Folglich ist ein Roboter für solche Handlungen nicht verantwortlich. Für unser Beispiel heißt das:

Wenn man den Roboter dafür bestrafen oder belohnen würde, dass er den Menschen gewarnt hat, würde dies zu keiner Änderung im Verhalten führen. Demnach ist der Roboter im Beispiel nicht moralisch verantwortlich.

---

<sup>12</sup>vgl. ebenda S. 46-51

<sup>13</sup>Dies werde ich im Abschnitt „Asimovsche Robotergesetze als Funktionen“ bestreiten.

<sup>14</sup>vgl. Schulte, Peter: Philosophie verständlich : Willensfreiheit : Daniel C. Dennett. <<http://www.philosophieverstaendlich.de/freiheit/modern/dennett.html>>, 3.9.2007, 15:00.

### 4.2.3 John Martin Fischers „Reasons-responsive Mechanism View“

Eine ähnliche These wie Dennett vertritt auch John Martin Fischer mit seinem „Reasons-responsive Mechanism View“. Er sieht die sogenannte Führungskontrolle als notwendig für moralische Verantwortlichkeit an. Ein Akteur hat Führungskontrolle über eine Handlung, wenn diese Handlung aus einem Prozess resultiert, der zugänglich für Gründe ist.<sup>15</sup> Man ist also für eine Handlung verantwortlich, wenn ihr ein Prozess zugrunde liegt, welcher verschiedene Gründe abwägt.

Beim Roboter der nach den Robotergesetzen handelt, findet ein solches Abwiegen aber nicht statt: Auch wenn die meisten Gründe dafür sprechen gegen die Regeln zu handeln, wird der Roboter dennoch die Regeln befolgen, da sie absolut zwingend sind. Dies sieht man auch an unserem Beispiel: Selbst wenn dem Menschen ein qualvoller Tod direkt bevorstehen würde, da er beispielsweise vergiftet ist, würde der Roboter den Menschen warnen und damit auch seine eigene Zerstörung in Kauf nehmen.

Folglich ist auch nach dieser Position ein Roboter nicht für seine Handlungen verantwortlich, die auf den Robotergesetzen beruhen.

## 5 Asimovsche Robotergesetze als Funktionen

Auf die Frage, ob Handlungen die von Robotern nach den Robotergesetzen ausgeführt werden, von diesen moralisch verantwortet werden müssen, gibt, von den untersuchten Positionen, nur Harry G. Frankfurt eine positive Antwort. Dies bedeutet, dass nur Frankfurts Position nicht ausschließt, dass es sich bei den Robotergesetzen um moralische Regeln handeln könnte. Frankfurts Position behauptet dadurch noch lange nicht, dass dieser Zusammenhang tatsächlich besteht.

Wenn es sich aber bei den Robotergesetzen nicht um moralische Regeln handelt, was sind sie dann? Meine Antwort auf diese Frage ist folgende:

Die Robotergesetze lassen sich nicht verstehen, wenn man dem Roboter gegenüber eine intentionale Einstellung einnimmt, wenn man also davon ausgeht, dass ein Roboter Wünsche, Überzeugungen etc hat. Viel mehr sollte man eine funktionale Einstellung einnehmen, wie es bei Maschinen allgemein üblich ist.<sup>16</sup>

Tut man dies, kann man sagen, dass die Robotergesetze nur Funktionen eines Roboters sind, die dafür sorgen, dass dieser...

1. ...keinem Menschen schadet und sogar Schaden von Menschen abwendet.
2. ...funktioniert, wie es die Menschen wollen.
3. ...Schaden an sich selbst vermeidet.

---

<sup>15</sup>vgl. Ekstrom, Laura Waddel. *Free Will. A Philosophical Study*, Boulder: Westview Press, 2000. S. 169f.

<sup>16</sup>vgl. Beckermann, Ansgar. *Analytische Einführung in die Philosophie des Geistes*. 2. Auflage. De Gruyter, Berlin 2001. S. 307

## 6 Fazit

Sowohl der südkoreanische Roboterkodex als auch die asimovschen Robotergesetze haben nicht das Ziel, Roboter moralisch handeln zu lassen.

Der Roboterkodex soll sicherstellen, dass einige Funktionen, wie beispielsweise Ortbarkeit, Datensicherheit etc. in zukünftig produzierte Roboter eingebaut werden. Da es sich bei den asimovschen Robotergesetzen um genau solche Funktionen handelt, wäre es nichts grundsätzlich anderes, wenn man auch diese in zukünftige Roboter eingebaut würden.

Formulierungen wie ich sie in dem den Abschnitt „Was werde ich tun?“ zitiert habe, sind also unangebracht.

## 7 Quellen

- Asimov, Isaac: Ich, der Robot [Dt. Übers.: Otto Schrag], München : Heyne , 1978.
- Asimov und die Realität: Auf dem Weg zu einer Roboter-Ethik. <<http://derstandard.at/?url=/?id=3000701>>, 25.8.2007, 12:00.
- Beckermann, Ansgar. Analytische Einführung in die Philosophie des Geistes. 2. Auflage. De Gruyter, Berlin 2001.
- Clark, Andy. Mindware: An Introduction to the Philosophy of Cognitive Science, Oxford: Oxford University Press, 2001.
- Ekstrom, Laura Waddel. Free Will. A Philosophical Study, Boulder: Westview Press, 2000.
- Schulte, Peter: Philosophie verständlich : Willensfreiheit : Daniel C. Dennett. <<http://www.philosophieverstaendlich.de/freiheit/modern/dennett.html>>, 3.9.2007, 15:00.
- Schulte, Peter: Philosophie verständlich : Willensfreiheit : Harry G. Frankfurt. <<http://www.philosophieverstaendlich.de/freiheit/modern/frankfurt.html>>, 31.8.2007, 11:23.