

Bedeutungsverstehen als Kennzeichen des Mentalen

ANSGAR BECKERMANN

Summary

In his paper „Machines and the Mental“ Fred Dretske claims that there is a difference on principle between men on the one and animals and machines on the other side which arises from the fact that men are able to respond to the meaning of symbols whereas animals and machines can only respond to the symbols that have the meaning. In this paper it is argued that this claim does not bear closer scrutiny. Mainly for two reasons: 1. The mere having of a certain meaning does not rank among the possible causal properties of a symbol, not even in the case of men; for it is not the having of the meaning that may have certain effects on the behavior of a person, but the fact that the person *associates* this meaning with the symbol. 2. Even animals are capable to associate a meaning with a symbol, i. e. even animals can learn to come to believe that *p* if they become aware of a symbol which has the meaning that *p*.

1. Wenn sie mit der These konfrontiert werden, daß es Maschinen geben könnte, die denken und sprechen (und vielleicht sogar fühlen) können wie wir, daß es also „Maschinen wie Menschen“ geben könnte, reagieren viele spontan mit einer gehörigen Portion Skepsis. Intuitiv scheint ihnen klar, daß es – so groß die Kunst der Wissenschaftler und Techniker auch sein mag – immer einen prinzipiellen Unterschied zwischen Menschen und Maschinen geben wird. Allerdings fällt es nicht nur dem Ungeschulten schwer, genau zu sagen, worin denn dieser prinzipielle Unterschied bestehen soll. Verschiedene Möglichkeiten scheinen sich zwar von selbst anzubieten: Bewußtsein z. B. Aber gerade Bewußtsein ist ein notorisch so unfaßbares Phänomen, daß es schwer ist, für oder gegen die These, daß Maschinen Bewußtsein haben können, überhaupt Argumente zu finden.

Vielleicht ist das der Grund, warum in den philosophischen Debatten der letzten Jahre eine andere Frage in den Mittelpunkt der Diskussion gerückt ist, die Frage nämlich, ob nicht der prinzipielle Unterschied zwischen Menschen und Maschinen vielleicht darin besteht, daß Maschinen im Umgang mit Zeichen und Symbolen zwar die physische Gestalt der Zeichen erfassen, aber – im Unterschied zu Menschen – grundsätzlich nie die Bedeutung eines Zeichens verstehen können. Diese These ist inzwischen von vielen Autoren vertreten und mit den verschiedensten Etiketten versehen worden: bei Searle z. B. taucht sie unter dem Stichwort „Intentionalität“¹ auf, bei Dennett unter dem Titel „semantische vs. syntaktische Maschinen“² und bei Fred Dretske

¹ Vgl. besonders J. Searle, „Minds, Brains, and Programs“, in *Behavioral and Brain Sciences* 3 (1980), 417–424 u. 450–456 sowie ders., *Geist, Hirn und Wissenschaft*, Suhrkamp: Frankfurt/M. 1986.

² Z. B. D. Dennett, „Three Kinds of Intentional Psychology“, in R. Healey (ed.), *Reduction, Time, and Reality*, Cambridge UP: Cambridge 1981, 37–61.

schließlich unter dem Stichwort „Bedeutungsverstehen“. Ich möchte mich hier jedoch auf die Argumentation Dretskes konzentrieren und deshalb insbesondere die Thesen, die Dretske in seinem 1985 vor der American Philosophical Association gehaltenen Vortrag „Machines and the Mental“³ entwickelt hat, zum Ausgangspunkt meiner Überlegungen machen. Denn anhand dieser Thesen lassen sich, wie mir scheint, einige der Probleme besonders gut diskutieren, die mit der Frage zusammenhängen, ob auch Maschinen die Bedeutung von Zeichen und Symbolen erfassen können, d. h. ob semantische Maschinen möglich sind.

Wie anderen Autoren geht es auch Dretske in seinen Überlegungen darum, die These zu präzisieren, daß der Unterschied zwischen Menschen auf der einen und Tieren und Maschinen auf der anderen Seite darin liegt, daß Menschen auf die *Bedeutung* von Zeichen (Symbolen, bedeutungstragenden Zuständen, etc.) reagieren können, während das Verhalten von Tieren und Maschinen allein durch die Zeichen selbst bestimmt ist. Diese These über den grundsätzlichen Unterschied zwischen Menschen und anderen Lebewesen oder Systemen erläutert Dretske hauptsächlich am Beispiel eines Polizeihundes, der auf das Aufspüren von Marihuana im Gepäck von Reisenden abgerichtet wurde. Die Aufgabe eines solchen Hundes besteht darin, mit Bellen oder Schwanzwedeln oder mit irgendeinem anderen Verhalten, auf das er trainiert wurde, zu reagieren, wenn ihm beim Beschnüffeln von Gepäckstücken der Geruch von Marihuana in die Nase sticht. Wenn der Hund bellt oder mit dem Schwanz wedelt, wird der Beamte auf das vom Hund gerade beschnüffelte Gepäckstück aufmerksam, d. h. das Bellen oder Schwanzwedeln des Hundes bewirkt, daß der Beamte die Überzeugung gewinnt, daß sich in diesem Gepäckstück Marihuana befindet. Kann man aber, so fragt Dretske, auch von dem Hund selbst sagen, daß er eine ähnliche Überzeugung hat, daß er dasselbe glaubt wie der Polizeibeamte und daß er deshalb bellt oder mit dem Schwanz wedelt? Kann man überhaupt sagen, daß der Hund irgendetwas glaubt? Dretske meint, daß man das nicht kann, d. h. daß man dem Hund selbst nicht die Überzeugung zuschreiben kann, daß sich in diesem Gepäckstück Marihuana befindet. Doch er bemerkt sofort, daß diese negative Antwort offenbar nicht selbstverständlich ist, daß sie manchem sogar eindeutig kontraintuitiv erscheinen könnte. Denn offenbar gibt es doch in dem Hund etwas, was auf die Anwesenheit von Marihuana reagiert, d. h. offenbar existiert in dem Hund ein (wahrscheinlich neuronaler) Zustand, der genau dann auftritt, wenn Marihuana in der Nähe ist, und der daher – objektiv – ein Zeichen dafür ist (und damit in diesem Sinne bedeutet), daß sich Marihuana in der Nähe des Hundes befindet. Außerdem ist es offenbar genau dieser Zustand, der in dem Hund das Bellen oder Schwanzwedeln hervorruft (je nach dem, worauf er trainiert wurde). „Why isn't this enough to justify attributing a belief to the dog, a belief with the content: there is marijuana nereby?“ (30) Dretskes Antwort auf diese Frage lautet, daß der Hund zwar über die not-

³ F. Dretske, „Machines and the Mental“, in *Proceedings and Addresses of the American Philosophical Association* 59 (1985), 23–33.

wendigen *sensorischen* Fähigkeiten verfügt, um Marihuana aufzuspüren, daß ihm aber die *konzeptuellen* oder *kognitiven* Fähigkeiten fehlen, die notwendig sind, um eine entsprechende Überzeugung auszubilden. Etwas später erläutert Dretske diese Unterscheidung noch einmal genauer:

„The difference between machines (or dogs) and the agents who use them is that although machines (and dogs) can pick up, process and transmit . . . information . . . , although they can respond . . . to meaningful signs, it isn't the meaning of the signs that figures in the explanation of why they do what they do. Some internal sign of marijuana, some neurological condition that, in this sense, means that marijuana is present, can cause the dog's tail to move, but it isn't the fact that it means this that explains the tail movement.“ (30)

Der Unterschied zwischen einem Menschen und einem Hund liegt Dretske zufolge also darin, daß, wenn ein Mensch seinen Finger bewegt, um damit anzuzeigen, daß er Marihuana gerochen hat, diese Bewegung (anders als beim Hund) nicht einfach durch einen bestimmten neuronalen Zustand hervorgerufen wird, der eine bestimmte Bedeutung hat, sondern durch die Tatsache, *daß dieser Zustand diese Bedeutung hat*, oder wie Dretske manchmal auch sagt, daß diese Bewegung beim Menschen nicht durch einen neuronalen Zustand, sondern durch die Bedeutung dieses Zustands hervorgerufen wird.

„This, I submit, is the difference between the dog and its master, between the machine and its users, between the robot and the people it replaces. When I smell marijuana, my finger wagging is produced, not simply (as in the case of the dog) by a neurological condition that means that marijuana is present, but by the *meaning* of this neurological condition, by the fact that it means this and not something else. In *my* case the motor activity is produced by the meaning of an occurrent sign; in the dog's case by the occurrence of a sign having that meaning.“ (30f.)

Menschliches Verhalten unterscheidet sich vom Verhalten von Tieren oder Maschinen also grundsätzlich dadurch, daß es – zumindest manchmal – durch die *Bedeutung* von Zeichen oder inneren Zuständen bestimmt wird. Die inneren Zustände von Tieren und Maschinen dagegen können zwar im Prinzip genau dieselbe Bedeutung haben wie die entsprechenden Zustände eines Menschen. Aber bei Tieren und Maschinen sind es niemals diese Bedeutungen, von denen ihr Verhalten abhängt, sondern immer nur die Zustände selbst. Kurz gesagt: Tiere und Maschinen können nicht auf die Bedeutung von Zeichen reagieren, sondern nur auf die Zeichen selbst.

„That is why it is true of me, but not the dog, that I wag my finger because I think marijuana is present, because I am in an internal state having this content. The dog is in a state with the same content, to be sure, but it isn't this content that wags the tail. The difference between a thinker of marijuana thoughts (me) and the mere detector of marijuana (dog or machine) is not, then, merely a difference in what our internal signs mean, but a difference in whether, and if so, how, these meanings are implicated in the management of the signs themselves.“ (31)

Grundsätzlich kann man also sagen, daß Dretske in seinen Überlegungen von folgendem Begriff eines bedeutungsverstehenden bzw. semantischen Systems ausgeht.

- (I) Ein System *S* ist ein semantisches System genau dann, wenn das Verhalten von *S* – zumindest in einigen Fällen – auf die *Bedeutung* von Symbolen

oder inneren Zuständen und nicht nur auf die Symbole oder Zustände selbst zurückgeführt werden kann.

Sicher ist die in dieser Definition enthaltene Unterscheidung zwischen semantischen und nicht semantischen Systemen interessant, und sicher ist diese Definition auch viel klarer als die anderer Autoren. Aber ich denke, daß Dretskes Thesen trotzdem nicht haltbar sind, zumal sich aus diesen Thesen, wenn man sie beim Wort nimmt, eine Konsequenz ergibt, die mit den Intentionen Dretskes völlig unvereinbar ist. Denn streng genommen folgt, soweit ich sehen kann, aus der oben angeführten Definition (I), daß nicht einmal Menschen wirklich als bedeutungsverstehende Systeme gelten können. Es muß also irgendwo in Dretskes Überlegungen ein Fehler stecken. Und es lohnt sich sicher, diesem Fehler nachzuspüren.

2. Dretskes eigenen Formulierungen zufolge ist die Kernthese seiner Überlegungen eine Behauptung über bestimmte Verursachungsbeziehungen. Denn diese These lautet, wie schon gesagt, daß es zumindest einige Fälle gibt, in denen das Verhalten eines Menschen nicht durch ein Zeichen, sondern durch die Bedeutung dieses Zeichens hervorgerufen wird, während auf der anderen Seite das Verhalten von Tieren und Maschinen, soweit es überhaupt von Zeichen bestimmt wird, immer auf die Zeichen selbst zurückgeht. Um diese These beurteilen zu können, ist daher zunächst eine etwas allgemeinere Überlegung zur Natur von Kausalaussagen nötig.

Wenn in der philosophischen Literatur über Kausalbeziehungen geredet wird, dann werden dabei häufig informelle Abkürzungen wie „*A* ist die (eine) Ursache von *B*“ oder „Wenn *A*, dann *B*“ verwendet. Diese Abkürzungen mögen in vielen Fällen zweckdienlich sein. Aber bei genauerer Betrachtung lassen sie die Struktur von Kausalbehauptungen doch nicht recht deutlich werden. Insbesondere bleibt bei diesen Formulierungen unklar, welche Art von Entitäten durch die Buchstaben „*A*“ und „*B*“ bezeichnet werden, d. h. es bleibt unklar, welche Art von Entitäten die Relationsglieder von Kausalrelationen sind. Manchmal findet man zwar in der Literatur Hinweise, daß diese Relationsglieder der Art nach Ereignisse (so die einen) oder Tatsachen (so die anderen) seien. Doch das hilft hier noch nicht viel weiter. Bei Fodor finden sich jedoch an einigen Stellen Formeln wie

$$(1) \quad F(x) \rightarrow G(y)$$

als Abkürzungen für: „Für beliebige Dinge *x* und *y* gilt: wenn *x* die Eigenschaft *F* hat, dann hat (oder bekommt) *y* die Eigenschaft *G*“. Und das ist schon etwas hilfreicher, weil es deutlich macht, daß es bei Kausalrelationen nicht auf die beteiligten Einzelgegenstände, sondern in erster Linie auf deren Eigenschaften ankommt. Noch präziser wäre jedoch die Formel:

$$(2) \quad F(x) \ \& \ R(x, y) \rightarrow G(y).$$

Denn tatsächlich wird z. B. ein Heuhaufen ja nicht durch jedes beliebige

brennende Streichholz in Brand gesetzt, sondern nur durch ein brennendes Streichholz, das dem Heuhaufen nahe genug kommt, also ein Streichholz, das in einer bestimmten Relation zu dem Heuhaufen steht. In gewisser Weise kann man also sagen, daß Kausalaussagen im allgemeinen die Form haben: „Für beliebige Gegenstände x und y gilt: wenn x die Eigenschaft F hat und in der Relation R zu y steht, dann hat (oder bekommt) y die Eigenschaft G “. Dies ist zwar sicher auch noch keine hinreichende Analyse. Aber für die Diskussion der Thesen Dretskes ist sie ein geeigneter Ausgangspunkt. Denn bei dieser Analyse wird ein Punkt besonders betont, der im gegenwärtigen Zusammenhang wichtig ist. In jedem Einzelfall, in dem gilt: b ist G geworden, weil a F hatte und in der Relation R zu b stand, kommt es nicht in erster Linie auf die beteiligten Gegenstände a und b sondern auf deren Eigenschaften F , G und auf die Relation R an. D. h. die Eigenschaften und Relationen der an einem Kausalzusammenhang beteiligten Gegenstände sind die für das Bestehen des Kausalzusammenhangs entscheidenden Faktoren.

Nach diesem Exkurs haben wir jetzt die Hilfsmittel, um Dretskes Hauptthese genauer analysieren zu können. Um es noch einmal zu wiederholen: diese These sagt, daß es wenigstens einige Fälle gibt, in denen eine bestimmte Handlung eines Menschen nicht durch ein bestimmtes Symbol oder Zeichen, sondern durch die Bedeutung dieses Zeichens verursacht wird. Wenn wir „ s “ und „ A “ als Zeichen für Symbole und Individuen verwenden und „ P “ als Zeichen für die Eigenschaft, die Bedeutung, daß p , zu haben, und „ H “ als Zeichen für die Eigenschaft, die Handlung H auszuführen, dann können wir den ersten Teil der These Dretskes also so fassen: Es gibt wahre Kausalbehauptungen der Form

$$(3) \quad P(s) \ \& \ R(s, A) \rightarrow H(A).$$

Der zweite Teil dieser These besagt dagegen, daß Tiere nur auf Symbole selbst, aber nicht auf deren Bedeutung reagieren können. Und dies kann offenbar nur heißen, daß Tiere – wenigstens manchmal – auf alle token eines bestimmten Symboltyps (unabhängig von der Bedeutung des Symbols) mit demselben Verhalten reagieren. Wenn wir „ S “ als Zeichen für die Eigenschaft verwenden, token eines bestimmten Symboltyps S zu sein, dann läßt sich dieser Teil der These Dretskes daher so reformulieren: Wahre Kausalbehauptungen, in denen das Verhalten eines Tieres auf die Anwesenheit eines bestimmten Symbols zurückgeführt wird, haben (bestenfalls) die Form

$$(4) \quad S(s) \ \& \ R(s, A) \rightarrow H(A).$$

Mit anderen Worten: Dretske zufolge ist zumindest manchmal die *kausal entscheidende* Eigenschaft eines Symbols, aufgrund deren es einen Menschen zu einer bestimmten Handlung veranlaßt, die Eigenschaft, eine bestimmte Bedeutung p zu haben, während in allen Fällen, in denen ein Symbol in einem Tier ein bestimmtes Verhalten hervorbringt, diese Eigenschaft darin besteht, ein token eines bestimmten Symboltyps zu sein, d. h. im einfachsten Fall die Eigenschaft, eine bestimmte physische Gestalt zu haben.

Nachdem wir nun die Thesen Dretskes soweit geklärt haben, scheint mir offensichtlich zu sein, daß zumindest die erste dieser beiden Thesen nicht zutreffen kann. Und zwar einfach deshalb nicht, weil es – zumindest für Menschen – wahre Kausalbehauptungen der Form (3) nicht gibt. Der Grund dafür liegt darin, daß die Wahrheit von Kausalaussagen unter anderem auf der Wahrheit entsprechender Gesetze (oder gesetzesartiger Aussagen) beruht und daß daher eine Aussage der Form (3) nur dann wahr sein kann, wenn zumindest auch die gesetzesartige Aussage

(3') Für alle x gilt: $P(x) \ \& \ R(x, A) \rightarrow H(A)$

wahr ist. Es gibt aber sicher kein menschliches Individuum A , für das eine solche Aussage zutrifft. Denn dies würde bedeuten, daß dieses Individuum auf *jedes* Symbol (welcher Art auch immer) mit der Handlung H reagiert, solange dieses Symbol nur die Bedeutung p hat. Nehmen wir, um uns klar zu machen, was das bedeuten würde, z. B. an, p sei die Proposition, daß es brennt. Was würde es bedeuten, wenn es ein Individuum A gäbe, für das unter dieser Voraussetzung eine Aussage der Art (3') wahr wäre? Es würde bedeuten, daß dieses Individuum in jedem Fall, in dem es ein Symbol wahrnimmt (Relation $R!$), daß die Bedeutung hat, daß es brennt, mit einer bestimmten Handlung (etwa dem Ausruf „Feuer“) reagiert – ganz egal, um welches Symbol es sich dabei handelt, ein natürliches oder ein konventionelles. Dieses Individuum müßte also auf *alle* natürlichen Anzeichen von Feuer und ebenso auf *alle* sprachlichen Ausdrücke, die die Bedeutung haben, daß es brennt, in dieser Weise reagieren. Und d. h., es müßte nicht nur alle natürlichen Anzeichen für Feuer kennen, sondern auch für jede beliebige Sprache wissen, was in dieser Sprache „Es brennt“ heißt. Es scheint mir völlig klar zu sein, daß es ein solches Individuum nicht gibt. Denn zumindest für uns Menschen ist charakteristisch, daß wir immer erst lernen müssen, welche Bedeutung ein Symbol hat. Auf diesen Punkt werde ich noch zurückkommen.

An dieser Stelle könnte man jedoch versucht sein einzuwenden, diese Argumentation sei Dretske gegenüber unfair. Denn Dretske sagt ja nicht immer, daß es *die Bedeutung* eines Symbols ist, die einen Menschen zu einer bestimmten Handlung veranlaßt; manchmal wählt er auch die Formulierung, daß diese Handlung von der Tatsache hervorgerufen wird, daß *dieses Symbol* diese Bedeutung hat. Vielleicht gehört zu den kausal entscheidenden Eigenschaften also nicht nur die Bedeutung allein, sondern *auch* die Eigenschaft, daß dieses Symbol token eines bestimmten Symboltyps ist. Und in diesem Fall würde der erste Teil der These Dretskes eher besagen, daß es beim Menschen wahre Kausalbehauptungen der Form

(5) $S(s) \ \& \ P(s) \ \& \ R(s, A) \rightarrow H(A)$

gibt. In dieser Form wäre Dretskes These offenbar nicht so leicht zu erschüttern. Denn daß eine gesetzesartige Aussage der Art

(5') Für alle x gilt: $S(x) \ \& \ P(x) \ \& \ R(x, A) \rightarrow H(A)$

wahr ist, kann zumindest nicht von vornherein ausgeschlossen werden. Zumal es für die Wahrheit einer solchen Aussage ausreicht, daß ein Individuum A auf alle token eines bestimmten Symboltyps in der gleichen Weise reagiert. Und das ist sicher nicht unmöglich.⁴

Dennoch scheint mir Dretskes These auch in der zweiten Variante nicht zuzutreffen. Denn die Wahrheit der Aussage (5) setzt nicht nur die Wahrheit der Aussage (5') voraus, sondern (wegen der von ihr mitbehaupteten *Relevanz* der Eigenschaft P) auch die Wahrheit des irrealen Konditionalsatzes

(5'') Wenn s nicht die Eigenschaft P gehabt hätte (d. h. wenn s nicht die Bedeutung, daß p , gehabt hätte), hätte A nicht die Handlung H ausgeführt.

Und die Wahrheit von Aussagen dieser Art scheint mir ebenso zweifelhaft wie die Wahrheit von Aussagen der Art (3'). Damit sind wir jetzt an einem entscheidenden Punkt. Denn meine Kritik an Dretske beruht auf der Auffassung, daß es unplausibel ist anzunehmen, daß jemand etwas tut oder unterläßt, *weil* ein ihm vorliegendes Symbol eine bestimmte Bedeutung hat oder nicht hat. Die Eigenschaft, eine bestimmte Bedeutung zu haben, gehört, wie mir scheint, einfach nicht zu der Art von Eigenschaften, die bei der Hervorbringung von Handlungen eine kausale Rolle spielen können.

3. Dretske hat in seinem Vortrag einen solchen Einwand vorausgesehen. Aber er meint zeigen zu können, daß Einwände dieser Art unbegründet sind. Bedeutungen können seiner Meinung nach ebenso kausal wirksam sein wie andere abstrakte Eigenschaften, wie z. B. die abstrakte Eigenschaft der Gewichts-differenz zwischen einem Basketball und einer Bowlingkugel.

„Just as the difference in weight between a basketball and a bowling ball may be responsible, causally responsible, for the behavior of a beam balance, the correlations constituting the meaning of a sign can, and regularly *do*, affect the way a system processes that sign.“ (31)

Als Beispiel, mit dessen Hilfe er diese These untermauern will, wählt Dretske das Läuten einer Türglocke. In diesem Fall ist es die Korrelation zwischen dem Läuten und der Anwesenheit einer Person an der Tür, die dem Läuten die Bedeutung verleiht, daß jemand an der Tür ist. Und diese Korrelation, d. h. diese Bedeutung kann, so Dretske, sehr wohl einen kausalen Einfluß auf das Verhalten einer Person haben. Denn, so schreibt er:

„Exposure (either directly or indirectly) to this correlation produces a difference in whether, and if so, which, motoneurons are activated by the internal sensory sign of a ringing bell. This, it seems to me, is a case where the meaning of a sign, and not just the sign that has that meaning, makes a difference in *how* a system processes that sign – hence, a case where the sign's meaning, and not the sign itself, helps to explain the behavior of the system in which that sign occurs.“ (31)

⁴ Nach Dretske gilt dies ja sogar schon für Tiere.

Meiner Meinung nach zeigt Dretskes Beispiel jedoch nicht ganz das, was er mit ihm zeigen will. Und das wird schon daran deutlich, daß dieser Analyse zufolge natürlich auch das Verhalten eines Hundes (nennen wir ihn Hasso), dem eine Zeitlang immer wieder mit dem Läuten der Glocke angezeigt wird, daß sein Futternapf in der Küche steht, und der daraufhin, immer wenn er die Glocke läuten hört, in die Küche zu seinem Futter läuft, eben nicht durch das Läuten der Glocke selbst, sondern durch die Bedeutung des Glockenläutens bestimmt wäre – im Gegensatz zu dem, was Dretske zuvor über den Unterschied von Hunden und Menschen gesagt hat. Aber was genau ist falsch an Dretskes Argumentation?

Ich habe zuvor argumentiert, daß die Aussage

- (6) Hasso ist in die Küche gelaufen, weil die Glocke geläutet hat und dieses Glockenläuten die Bedeutung hatte, daß ein gefüllter Futternapf in der Küche steht

nur dann wahr sein kann, wenn Hasso nicht in die Küche gelaufen wäre, falls das Läuten der Glocke nicht diese Bedeutung gehabt hätte. Aber dies scheint mir (auch in diesem Fall) ganz unplausibel zu sein. Denn wenn wir z. B. annehmen, daß Hassos Besitzer an einem bestimmten Tag beschließt, die Glocke ab sofort genau dann zu läuten, wenn es draußen regnet, dann ändert das Glockenläuten damit *objektiv* seine Bedeutung. Aber das wird auf das Verhalten Hassos sicher keinen Einfluß haben. Er wird auch beim nächsten Läuten der Glocke in die Küche laufen und möglicherweise ganz erstaunt sein, dort kein Futter vorzufinden. Dies scheint mir ganz offensichtlich, und es zeigt etwas eigentlich Selbstverständliches, das in diesem Zusammenhang aber offenbar trotzdem ausdrücklich hervorgehoben werden muß. Denn es zeigt (a), daß Menschen ebenso wie Hunde erst *lernen* müssen, welche Bedeutung ein Zeichen hat (niemand reagiert *automatisch* auf die Bedeutung eines Zeichens), und es zeigt (b), daß unser Verhalten (ebenso wie das Verhalten des Hundes) letzten Endes nicht durch die *objektive* Bedeutung eines Zeichens bestimmt wird, sondern dadurch, daß wir *glauben*, daß dieses Zeichen eine bestimmte Bedeutung hat, d. h. dadurch, daß wir mit dem Zeichen eine bestimmte Bedeutung verbinden. Als Zwischenrésumé können wir deshalb festhalten: Auch Menschen reagieren nicht unmittelbar auf die Bedeutung von Symbolen oder Zeichen; sie reagieren auf die Symbole selbst, aber wie sie darauf reagieren, das hängt davon ab, welche Bedeutung sie den Symbolen zumessen.

Was bedeutet nun dieses Zwischenergebnis für den von Dretske postulierten Unterschied zwischen Menschen auf der einen und Tieren und Maschinen auf der anderen Seite? Ein Vertreter der Position Dretskes könnte versuchen, dessen ursprüngliche Argumentation an dieser Stelle folgendermaßen zu variieren: „Vielleicht“, so könnte er sagen, „ist es tatsächlich so, daß auch Menschen nicht auf die Bedeutung von Symbolen, sondern nur auf die Symbole selbst reagieren; aber immerhin hängt doch die Art und Weise ihrer Reaktion davon ab, welche Bedeutung sie mit einem Symbol verbinden; Tiere

dagegen reagieren auf jedes Symbol nur mit einer festgelegten Verhaltensweise, und zwar der Verhaltensweise, die ihnen durch einen Menschen oder durch die Verhältnisse adressiert wurde.“ Mir scheint jedoch, daß auch diese Argumentationsvariante den tatsächlichen Gegebenheiten nicht gerecht würde.

Nehmen wir als Beispiel noch einmal unseren Hund Hasso, der in die Küche läuft, wenn er das Läuten einer Glocke hört, weil er gelernt hat, daß, immer wenn die Glocke läutet, der gefüllte Futternapf in der Küche steht. Offenbar kann man dieses Verhalten Hassos nicht dadurch erklären, daß man sagt, er habe gelernt, beim Läuten der Glocke *mit einer bestimmten Verhaltensweise* zu reagieren. Denn sein Verhalten wird je nach dem, in welcher Situation er sich gerade befindet, ganz unterschiedlich sein: wenn er schon in der Küche ist, wird er sich vielleicht nur umdrehen; wenn er im Wohnzimmer auf dem Sofa liegt, wird er versuchen, in die Küche zu kommen, und dabei wird er, falls die Tür verschlossen ist, möglicherweise auch einen Umweg machen, z. B. aus dem Fenster springen und zur Gartentür wieder hereinkommen; falls er überhaupt keine Möglichkeit hat, in die Küche zu kommen, wird er vielleicht an der Tür kratzen und bellen oder sich vor dem im Sessel sitzenden Herrchen aufbauen, leise winseln und mit dem Schwanz wedeln; und schließlich kann es auch passieren, daß er, obwohl er die Glocke durchaus gehört hat, völlig ungerührt auf seinem Platz bleibt und weiter döst – dann nämlich, wenn er zur Zeit überhaupt keinen Drang zu fressen verspürt. Dieses ganze Repertoire verschiedenster Verhaltensweisen (das wahrscheinlich auch noch insofern offen ist, als man es nicht von vornherein vollständig aufzählen kann) kann meiner Meinung nach nicht durch die Annahme erklärt werden, daß das Symbol „Glockenläuten“ bei Hasso ein bestimmtes Verhalten hervorruft. Vielmehr legt das unterschiedliche Verhalten dieses Hundes die Annahme nahe, daß Hasso gelernt hat, mit dem Symbol „Glockenläuten“ *einen bestimmten Sachverhalt* zu assoziieren, nämlich den Sachverhalt, daß in der Küche ein gefüllter Futternapf steht. Nur diese Annahme macht klar, warum sich Hasso in all den verschiedenen geschilderten Situationen so verhält, als wüßte er, daß in der Küche ein gefüllter Futternapf steht.

Ich denke, daß daran zweierlei deutlich wird. Erstens: Der Unterschied zwischen Menschen und Tieren ist wohl doch nicht so groß, wie er von Dretske dargestellt wird. Und zweitens: Sinnvoller als Dretskes Definition, die letzten Endes sogar Menschen aus der Klasse der semantischen Systeme ausschließt, ist die folgende Definition für diesen Begriff:

(I') Ein System *S* ist ein semantisches System genau dann, wenn *S* in der Lage ist, mit einem Symbol den Sachverhalt zu assoziieren, der die Bedeutung dieses Symbols ausmacht.

4. Auch dann, wenn man – zumindest auf der Grundlage dieser Definition – zu dem Ergebnis kommt, daß der Unterschied zwischen Menschen und Tieren nicht so groß ist, wie Dretske zunächst vorausgesetzt hatte, bleibt jedoch die Frage, was diese Modifikation der Definition (I) für das Problem „semantischer Maschinen“ im eigentlichen Sinne bedeutet. D. h. konkret

bleibt die Frage: Kann es Maschinen geben, die der Definition (I') zufolge als semantische Systeme angesehen werden können? Oder noch konkreter: Kann es Maschinen geben, die in der Lage sind, mit Symbolen Sachverhalte zu assoziieren? Diese Frage kann offenbar erst dann beantwortet werden, wenn zunächst geklärt wird, was es überhaupt heißt, daß ein System (welcher Art auch immer) dazu in der Lage ist, mit einem Symbol einen Sachverhalt zu assoziieren.

Wenn wir noch einmal auf unseren Hund Hasso zurückkommen, dann liegen die Dinge in etwa so. Das Verhalten Hassos ist – wie das Verhalten anderer höherer Tiere auch – davon abhängig, welche Sachverhalte in seiner Umgebung realisiert sind. Wenn z. B. sein Futternapf gefüllt in der Küche steht, zeigt Hasso zwar nicht immer dasselbe Verhalten, aber doch in jedem Fall ein Verhalten, das in einem klaren Zusammenhang zu diesem Sachverhalt steht, sagen wir ein bestimmtes auf diesen Sachverhalt abgestimmtes Verhaltensmuster. Genauer muß man jedoch sagen, daß Hassos Verhalten nicht direkt von der Realisierung bestimmter Sachverhalte abhängt, sondern davon, daß Hasso *weiß* oder noch genauer: daß er *glaubt*, daß diese Sachverhalte realisiert sind. Wenn sein Herrchen in der Küche den Futternapf füllt, Hasso aber – im Wohnzimmer auf dem Sofa dösend – davon nichts bemerkt, wird nichts passieren. Hasso döst weiter.

Wichtig ist hier deshalb die Frage, wie Hasso dazu kommen kann zu glauben, daß ein bestimmter Sachverhalt realisiert ist. Die erste Antwort auf diese Frage versteht sich fast von selbst: durch Wahrnehmung. Wenn Hasso sieht (oder hört), daß sein Herrchen in der Küche den Futternapf füllt, dann kommt er zu der Überzeugung, daß dieser Sachverhalt realisiert ist und verhält sich entsprechend. Das ist nicht weiter mysteriös. Denn Wahrnehmungsprozesse sind gerade als die Prozesse definiert, die zu in der Regel wahren Überzeugungen über bestimmte Sachverhalte führen. Symbole, Zeichen kommen nun dort ins Spiel, wo direkte Wahrnehmung nicht möglich ist. Sie sind die zweite Möglichkeit, etwas über die Realisierung von Sachverhalten zu erfahren. Wenn mir *A* sagt, daß es draußen regnet, dann entsteht bei mir – zumindest wenn ich keinen Grund habe, an der Aufrichtigkeit von *A* zu zweifeln – dieselbe Überzeugung, wie wenn ich den Regen selber sehe. Und das gleiche gilt für natürliche Zeichen. Wenn ich Rauch sehe, entsteht in mir die Überzeugung, daß es an dieser Stelle Feuer gibt. Ausgehend von diesem Zusammenhang kann man jetzt die Frage, was es heißt, daß ein System mit einem Symbol einen Sachverhalt assoziiert, so beantworten, daß dies genau dann der Fall ist, wenn das System, immer wenn es dieses Symbol wahrnimmt, die Überzeugung entwickelt, daß dieser Sachverhalt realisiert ist. Daß Hasso in der Lage ist, das Läuten der Glocke mit dem Sachverhalt zu assoziieren, daß ein gefüllter Futternapf in der Küche steht, bedeutet also nichts anderes, als daß sich in Hasso immer dann, wenn die Glocke läutet, die entsprechende Überzeugung einstellt. Präzise gefaßt:

- (II) X assoziiert mit Symbolen vom Typ S einen bestimmten Sachverhalt genau dann, wenn sich in X , falls es ein Symbol vom Typ S wahrnimmt, (in der Regel) die Überzeugung einstellt, daß dieser Sachverhalt realisiert ist.

5. Damit kommen wir jetzt zur letzten und entscheidenden Frage, nämlich der Frage, ob Maschinen etwas glauben oder von etwas überzeugt sein können. Bevor diese Frage beantwortet werden kann, muß jedoch zunächst der Verdacht ausgeräumt werden, daß ich in den bisherigen Überlegungen sowieso schon zu weit vorausgespracht bin, da ich in diesen Überlegungen einfach vorausgesetzt habe, daß man zumindest unserem Hund Hasso problemlos Wünsche und auch Überzeugungen zuschreiben kann. Immerhin haben eine ganze Reihe zeitgenössischer Philosophen bestritten, daß Tiere Meinungen oder Überzeugungen in unserem Sinn haben können. Donald Davidson z. B. hat entschieden die Ansicht vertreten, daß man nur sprachfähigen Wesen Wünsche und Überzeugungen zusprechen kann⁵, da die beiden Prozesse der Interpretation, mit denen man bestimmt, erstens welche Bedeutungen ein Sprecher bestimmten Sätzen zumißt und zweitens welche Überzeugungen dieser Sprecher hat, systematisch miteinander verbunden sind. D. h. eigentlich handelt es sich bei Davidson zufolge hier sogar nur um *einen* Prozeß. Auch Stephen Stich, der sich sehr ausführlich mit der Frage beschäftigt hat, ob Tiere Wünsche und Meinungen haben können⁶, ist in diesem Punkt eher an der Seite Davidsons. Das wird z. B. deutlich, wenn er in seinem Buch *From Folk Psychology to Cognitive Science – The Case Against Belief* über „beliefs“ u. a. schreibt:

„To believe that p is to be in a belief state similar to the one underlying our own sincere assertion of ' p .'“ (136)

Auch Stich räumt jedoch ein, daß wir bei der Erklärung des Verhaltens von Tieren fast gezwungen sind, auf intentionale Zustände Bezug zu nehmen. Und wenn wir uns die Situation noch einmal vergegenwärtigen, dann war es auch bei Hasso so, daß es sein Verhalten war, das uns veranlaßte, anzunehmen, daß Hasso in diesem Verhalten von bestimmten Überzeugungen und Wünschen bestimmt wurde. Das Problem liegt für Stich, wie z. T. auch für Davidson, darin, daß wir einerseits zwar kaum umhin können, Tieren Wünsche und Meinungen zuzuschreiben, daß wir andererseits aber nicht sagen können, was es denn genau ist, das diese Tiere wünschen oder glauben, d. h. daß es bei Tieren unmöglich zu sein scheint, den Inhalt ihrer intentionalen Zustände genau zu spezifizieren. Ich denke jedoch (ebenso wie andere Kritiker von Davidson und Stich⁷), daß dieses Problem überwunden werden kann, wenn man insbesondere bei Überzeugungen nicht nur darauf achtet, welches

⁵ Siehe z. B. D. Davidson, „Thought and Talk“, in ders., *Truth and Interpretation*, Clarendon Press: Oxford 1984, 155–170 und „Rational Animals“, in *Dialectica* 36 (1982), 317–327.

⁶ Siehe z. B. St. Stich, „Do Animals have Beliefs“, in *Australasian Journal of Philosophy* 57 (1979), 15–28 und ders., *From Folk Psychology to Cognitive Science – The Case Against Belief*, MIT Press: Cambridge, Mass. 1983.

⁷ Als ein Beispiel: R. Stalnaker, *Inquiry*, MIT Press: Cambridge, Mass. 1984.

Verhalten sie verursachen, sondern auch darauf, durch welche Umstände sie selbst verursacht werden.

Meiner Meinung nach ist hier wieder die Wahrnehmung entscheidend. D. h., um Überzeugungen haben zu können, muß ein System einen perzeptuellen Apparat besitzen, mit dessen Hilfe es in der Lage ist, im allgemeinen (d. h. unter normalen Bedingungen) Situationen, in denen ein bestimmter Sachverhalt realisiert ist, von Situationen zu unterscheiden, in denen das nicht der Fall ist. Denn wenn es in einem solchen System einen Zustand gibt, der – im Wahrnehmungsfall – genau dann eintritt, wenn eine Situation der ersten Art vorliegt, dann kann man meines Erachtens sehr wohl sagen, daß ein solcher Zustand den entsprechenden Sachverhalt repräsentiert und daß daher dieser Sachverhalt als der Inhalt dieses Zustandes gelten kann. Für Überzeugungen ist also entscheidend, daß sie einerseits das Verhalten eines Systems in einer bestimmten Weise beeinflussen und daß sie andererseits über den perzeptuellen Apparat des Systems so mit der Umwelt verbunden sind, daß sie im Regelfall dann und nur dann auftreten, wenn der Sachverhalt realisiert ist, der den Inhalt dieser Überzeugungen ausmacht.⁸

Wenn man aber Überzeugungen in diesem Sinne auffaßt, dann scheint es zumindest nicht prinzipiell unmöglich, daß auch Maschinen von etwas überzeugt sein können. Und dann spricht daher meiner Meinung nach letzten Endes auch nichts mehr gegen die Annahme, daß semantische, bedeutungsverstehende Maschinen zumindest möglich sind. Denn letzten Endes erfordert dies nur, daß eine Maschine einen perzeptuellen Apparat besitzt, der in dieser Maschine (unter Wahrnehmungsbedingungen normalerweise) genau dann einen bestimmten Zustand *Z* erzeugt, wenn ein bestimmter Sachverhalt realisiert ist, und daß die Maschine lernen kann, diesen Zustand *Z* mit einem bestimmten Symboltyp *S* zu assoziieren, d. h. daß diese Maschine lernen kann, auch dann den Zustand *Z* anzunehmen, wenn sie ein Symbol vom Typ *S* wahrnimmt. Und darin scheint mir keine prinzipielle Schwierigkeit zu liegen.

6. Ich möchte diesen Aufsatz mit drei Bemerkungen beschließen, die helfen sollen, möglichen Mißverständnissen vorzubeugen.

Erstens: Die bisher vorgetragenen Argumente beschäftigen sich nur mit dem *assertorischen* Gebrauch von Zeichen oder Symbolen. Darin liegt jedoch keine wesentliche Einschränkung. Denn für andere Arten des Zeichengebrauchs – etwa imperativische oder fragende – läßt sich eine in vieler Hinsicht analoge Analyse durchführen. Bei einem imperativischen Zeichengebrauch etwa, wenn also Hassos Besitzer das Läuten der Glocke z. B. im Sinne von „Geh in die Küche!“ verwendet, könnte man sagen, daß Hasso das Glockenläuten genau dann versteht, wenn er bei jedem Läuten der Glocke eine motivationale Struktur aufbaut, die darauf gerichtet ist, den durch den Satz „Hasso ist in der Küche“ ausgedrückten Sachverhalt herbeizuführen. Und auch ein fragender

⁸ Daß dies nicht ausreicht und daß neben den perzeptuellen Mechanismen bei Überzeugungen z. B. auch noch interne „up dating“-Mechanismen eine Rolle spielen müssen, habe ich in dem Aufsatz „Why Tropistic Systems Are Not Genuine Intentional Systems“ (erscheint demnächst in *Erkenntnis*) zu zeigen versucht.

Zeichengebrauch läßt sich ähnlich auffassen. Denn wenn Hassos Herrchen das Läuten der Glocke im Sinne von „Wo sind meine Pantoffeln?“ verwendet, dann könnte man sagen, daß Hasso das Läuten der Glocke genau dann versteht, wenn er zum Beispiel mit drei verschiedenen Handlungen (etwa mit dem Drücken dreier verschiedener Knöpfe) antwortet je nach dem, ob er glaubt, daß Herrchens Pantoffeln im Wohnzimmer, im Schlafzimmer oder in der Küche sind. (Dabei ist natürlich vorausgesetzt, daß Hasso in der Lage ist, Überzeugungen dieser Art auszubilden.) Mir ist klar, daß diese wenigen Bemerkungen bestenfalls einen skizzenhaften Charakter habe. Aber ich denke doch, daß sie zumindest in die richtige Richtung weisen.

Zweitens: Wenn Hassos Herrchen mit einer Glocke läutet, um Hasso davon in Kenntnis zu setzen, daß sein Futternapf gefüllt in der Küche steht, dann ist das Läuten der Glocke sicher ein *nicht-natürliches* Symbol im Sinne von Grice; aber es ist deshalb noch kein *sprachliches* Symbol. Man könnte deshalb versucht sein einzuwenden, daß die vorangegangenen Überlegungen zwar vielleicht zeigen, daß Tiere (und Maschinen) tatsächlich nicht-natürliche Symbole verstehen können, daß daraus aber noch nicht folgt, daß Tiere (und Maschinen) auch sprachliche Symbole verstehen können. Ich kann zwar auch auf diesen Punkt hier nur sehr kurz eingehen; aber ich wage trotzdem die These, daß dieser Einwand verfehlt wäre. Denn soweit ich sehen kann, benötigt ein System, um auch sprachliche Symbole verstehen zu können, zusätzlich zu den schon genannten Fähigkeiten im wesentlichen noch die Fähigkeit, komplexere sprachliche Ausdrücke (d. h. in erster Linie Sätze) syntaktisch und semantisch so zu analysieren, daß jedem wahrgenommenen Satz eine interne Struktur zugeordnet wird, die in eindeutiger Weise mit seiner Bedeutung korreliert ist. Und meiner Meinung nach zeigen selbst die bisher entwickelten natürlichsprachlichen Systeme der KI-Forschung schon, daß diese Fähigkeit maschinell realisierbar ist. Ich will jedoch nicht verschweigen, daß es gerade dann, wenn dies tatsächlich so ist, ein erstaunliches Faktum bleibt, daß wir keine Tier kennen, die auch nur annähernd so sprechen können wie wir.

Drittens: Dretske bezieht sich in seinen Überlegungen ausdrücklich nicht nur auf äußere Zeichen, sondern auch auf innere Symbole, d. h. auf interne Zustände, die eine bestimmte Bedeutung haben. Und er bezieht auch ausdrücklich Überzeugungen und Meinungen in seine Analysen mit ein. Ich habe mich jedoch in diesem Aufsatz auf äußere Symbole beschränkt. Und das hat natürlich seinen Grund. Denn auf der einen Seite glaube ich zwar, daß die kritischen Überlegungen, die ich im Hinblick auf die Möglichkeit angestellt habe, daß die Eigenschaft, eine bestimmte Bedeutung zu haben, eine kausal relevante Eigenschaft sein kann, auch bei Überzeugungen zutrifft. Auf der anderen Seite muß es für diese internen Zustände aber eine andere Alternativanalyse geben. Denn natürlich kann das Reagieren auf die Bedeutung einer Überzeugung nicht darin bestehen, daß man mit dieser ersten Überzeugung die zweite Überzeugung assoziiert, daß der Sachverhalt, der die Bedeutung der ersten Überzeugung ausmacht, realisiert ist. Das wäre offensichtlich zirkulär. Eine weitere Diskussion der Frage, wie eine Alternativanalyse für Überzeu-

gungen und andere intentionale Zustände denn nun wirklich aussehen könnte, würde jedoch den Rahmen dieses Aufsatzes bei weitem übersteigen.

Adresse des Autors:

Prof. Dr. Ansgar Beckermann, Philosophisches Seminar der Georg-August-Universität,
Platz der Göttinger Sieben 5, D-3400 Göttingen