the

GERD DÖBEN-HENISCH

The Blind's World I

Ein philosophisches Experiment auf dem Weg zum digitalen Bewußtsein

1 EINLEITUNG

Wer die Installation „The BLINDs WORLD I" (= BW1) betritt, dessen Blick wird zuallererst auf die bunten lustigen Pictogramme fallen, die die virtuelle Welt von BW1 repräsentieren. Die Assoziation zu einem Computerspiel drängt sich unweigerlich auf. Doch wäre der weitere Schluß, daß es sich bei BW1 auch „nur" um ein Spiel handelt, ein Fehl-Schluß.

BW1 versteht sich in erster Linie als ein philosophisches Experiment. BW1 soll entscheidend mit dazu beitragen, zentrale Fragen der modernen Philosophie zu beantworten, z. B.:

- Was ist das Bewußtsein?
- Welche Funktionen lassen sich im Bewußtsein unterscheiden?
- Wie entsteht ein „Wissen von der Welt"?
- Wie entwickelt sich Sprache?
- Welchen Einfluß haben die Gefühle?

Die Auffassung von BW1 als einem „philosophischem Experiment" steht in einiger Spannung zu einer Auffassung von Philosophie, die philosophische Erkenntnis ansieht als „in sich stehend", „nur auf sich beruhend", als „autark", „keiner zusätzlichen Hilfsmittel bedürfend", mit dem Ziel, das „Wesen der Dinge" zu erfassen, „ewige Wahrheiten" bzw. „allgemeingültige Prinzipien allen Erkennens" aufzudecken.

Dazu kommt, daß das Experiment als solches von Philosophen in der Regel dem empirischen Erkenntnisparadigma zugeschlagen wird, war es in der Vergangenheit doch gerade die „Kanonisierung des Experimentes" als Gültigkeitskriterium wissenschaftlicher Aussagen, die zur Abspaltung der modernen Naturwissenschaften von der klassischen Philosphie geführt haben. Für viele Philosophen – und wohl für die Mehrheit der Naturwissenschaftler – gilt diese „Spaltung" bis heute als unüberbrückbar.

BW1 hält noch eine weitere Provokation bereit: die Rede vom

A Philosophical Experiment on the Way to Digital Consciousness

1 INTRODUCTION

When you enter the installation "The BLIND's WORLD I" (= BW1), the first thing you see are the colorful funny pictographs representing the virtual world of BW1. The association with a computer game will inevitably suggest itself. However, the conclusion that BW1 is "just" a game would be a fallacy.

BW1 is primarily a "philosophical experiment". BW1 is designated to contribute decisively to issues of modern philosophy, such as

- What is consciousness?
- What functions can be discerned within consciousness?
- How do we gain "knowledge of the world"?
- How does language develop?
- What influence do emotions exert?

The notion of BW1 as a "philosophical experiment" represents a marked contrast to the idea of philosophy considering "philosophical cognition" as "self-contained", "based on itself", "independent", "needing no additional aid", with the aim to grasp "the nature of things"', uncovering "eternal truths" or "general principles of all recognition".

Furthermore, philosophers usually attribute the idea of the "experiment" as such to the empirical paradigm of cognition; in the past, it was after all the "canonization of the experiment" as a criterion to validate scientific statements which led to the schism between the modern natural sciences and classical philosophy. For many philosophers - and the majority of scientists, for that matter - this "schism" is still an unbridgeable gap.

BW1 holds another provocative idea in store: mention is made of "consciousness". In the modern experi-

mental sciences there is no room for a "consciousness sure of itself", not even in psychology. Even in modern philosophy, the term has been strongly discredited in certain fields - e.g. analytical philosophy.
Understanding BW1 in its philosophical and scientific significance will require a few additional explanations which go beyond the simple description of the BW1 program.
The history of BW1 starts with Alan Matthew TURING.

2 TURING THE VISIONARY

In 1936 Alan Matthew TURING's famous work "On computable numbers with an Application to the Entscheidungsproblem" was published. Not only did he answer - in the negative - the question for decideability in mathematics, which had been explicitly posed by HILBERT in 1928, he also proposed in it a definition of computable processes which during his lifetime became a household word among mathematicians and logicians under the name TURING MACHINE [TM]. For GÖDEL, the TM was the most satisfactory of all definitions ever given of a "mechanical process" (DAVIS 1965: 72, footnote).
The concept of the Turing Machine and especially of its generalized form, the "Universal Turing Machine" [UTM], is based on the hypothesis that it can represent all computable processes.
With his ideal machine TURING was as much ahead of the technical possibilities of his time as were his practical and philosophical visions inspired by this new concept.
The topics of his considerations, seen by many as provocative, fall into two categories: (1) the construction of an electronic brain and (2) learning processes in computers.
In spite of his strong interest in concrete physiology and particularly in neurophysiology - after all, he was a trailblazer in that he wrote one of the first mathematical works on the chemistry of morphogenesis in 1952 - he refused to imitate the physiological brain structure in hardware. He was rather interested in analyzing "the logical structure of the brain", proceeding on the assumption that every continuous system can be approximated with any degree of accuracy by a "discrete system". These assumptions opened a new perspective, i.e. that of simulating a brain approximated by discrete states in a UTM.
Given this process of creating a structural relation between the human brain and the UTM, the possibility of imitating human intelligence in a UTM seemed to be within reach. As early as in 1941 TURING had dealt with

Bewußtsein. In den modernen experimentellen Wissenschaften ist kein Platz für den Begriff eines „sich selbst gewissen Bewußtseins", nicht einmal im Rahmen der Psychologie. Selbst in der modernen Philosophie ist dieser Begriff in bestimmten Richtungen – wie z. B. in der analytischen Philosophie – stark diskreditiert.
Um also die philosophische und wissenschaftliche Bedeutung von BW1 verstehen zu können, bedarf es einiger Erläuterungen, die über eine bloße Deskription des Programms von BW 1 hinausgehen.
Die Geschichte von BW1 beginnt mit Alan Matthew TURING.

2 VISIONÄR TURING

1936 erschien die berühmte Arbeit On computable numbers with an Application to the Entscheidungsproblem von Alan Matthew TURING. Mit dieser Arbeit beantwortete er nicht nur die Frage nach der Entscheidbarkeit der Mathematik, die HILBERT 1928 explizit gestellt hatte, negativ, sondern er stellte in ihr auch eine Definition von berechenbaren Prozessen vor, die schon zu seinen Lebzeiten unter der Bezeichnung Turing Maschine [TM] zum Allgemeingut der Mathematiker und Logiker wurde. Für GÖDEL war die TM die befriedigenste aller vorgeschlagenen Definitionen eines „mechanischen Verfahrens" (DAVIS 1965:72 Anmk).
Mit dem Konzept der Turing Maschine, insbesondere mit ihrer verallgemeinerten Form, der Universellen Turing Maschine [UTM], verbindet sich die Hypothese, daß sich alle berechenbaren Prozesse durch sie darstellen lassen.
TURING war mit seiner idealen Maschine den technischen Möglichkeiten seiner Zeit weit voraus, desgleichen mit seinen praktischen und philosophischen Visionen, zu denen er durch sein neu gefundenes Konzept angeregt wurde.
Seine von vielen als provozierend empfundenen Überlegungen lassen sich vor allem zwei Themenkreisen zuordnen: (1) Bau eines elektronischen Gehirns und (2) Wie können Computer lernen?
Trotz seines starken Interesses für konkrete Physiologie und speziell auch Neurophysiologie – schrieb er doch 1952 eine der ersten bahnbrechenden mathematischen Arbeiten zur Chemie der Morphogenese – lehnte er es ab, die physiologischen Strukturen des Gehirns in der Hardware zu imitieren. Er war vielmehr an der Analyse der logischen Struktur des Gehirns interessiert und er ging davon aus, daß sich jedes stetige System mit beliebiger Genauigkeit durch ein diskretes System annähern läßt. Diese Annahmen eröffneten die Perspektive, ein durch diskrete Zustände approximiertes Gehirn durch eine UTM zu simulieren.
Auf der Basis einer solchen strukturellen In-Beziehung-Setzung von menschlichem Gehirn und UTM lagen natürlich auch Überlegungen zu einer möglichen Imitation der menschlichen Intelligenz durch eine UTM nahe. Schon 1941

hatte sich TURING mit der Frage von schachspielenden Maschinen beschäftigt und diese Fragen dann dahingehend ausgedehnt, wieweit eine Maschine, d. h. eine UTM, im allgemeinen Sinne lernen könne.
Bei der Verfolgung dieser Frage wirkt TURING ein wenig zwiespältig (siehe dazu [TURING 1948] und [TURING 1950]).
Einerseits ist er sich offensichtlich bewußt, daß zu einer allgemeinen Lernfähigkeit, so, wie sie der Mensch besitzt, ein entsprechend elaborierter Weltbezug gehört. Man müsse eine solche Maschine mit Fernsehkameras, Mikrophonen, Lautsprechern etc. ausstatten, damit sie möglichst umfassend zu Interaktionen mit der Außenwelt fähig sei, ebenso müßte sie „über Land streifen können", sie müsse „Eigeninitiative" besitzen, sie müsse „trainiert" und „unterrichtet" werden, kurzum alles, was ein menschliches Kind zur Verfügung hat, um zu lernen, das müsse man auch einer Maschine zur Verfügung stellen.
An anderen Stellen sprach er sich allerdings dagegen aus, den Menschen in seinen natürlichen Eigenschaften übermäßig imitieren zu wollen. Auch Maschinen mit einem reduzierten Körper waren für ihn interessante Kandidaten.
Das Erkennen des Vorliegens von „künstlicher Intelligenz" sollte mittels des Kriteriums der Imitation gewährleistet werden. Immer dann, wenn ein Mensch in einem ausschließlich durch einen über ein Terminal geführten Dialog zur „Meinung" käme, daß der Gesprächspartner ein Mensch sein könnte, dann könnte man auf die „künstliche Intelligenz" auch die Beschreibungsprädikate anwenden, die man sonst nur auf den Menschen anwenden würde.

3 BEDEUTUNG, WELTWISSEN UND BEWUSSTSEIN

Die durch TURING ausgelöste Diskussion, ob Maschinen nun eine Intelligenz entwickeln können, die der menschlichen vergleichbar ist, ob sie gar so etwas wie „Bewußtsein" entwickeln können, hält bis heute an und kann eigentlich nicht als definitiv entschieden gelten. Trotz aller faktischen Gewöhnung an Computer verbleibt der Fragestellung eine hohe philosophische Brisanz (DÖBEN-HENISCH 1993). Ein Schlüssel zur weiteren Klärung dieser Fragestellung scheint bei der Sprache zu liegen; dies sah auch schon TURING sehr klar. Im Rahmen der Taxierung möglicher Einsatzbereiche für eine intelligente Maschine stellt er z. B. fest: „Das Lernen von Sprachen wäre unter den oben genannten Anwendungen die beeindruckendste, weil es die menschlichste dieser Tätigkeiten ist" (TURING 1948; dt 1987:98).
TURING hat auch schon, wenngleich halbherzig, vorausgesehen, daß eine UTM, die ähnlich wie ein Mensch in der Lage sein soll, eine beliebige natürliche Sprache zu lernen und dann auch situationsgemäß anzuwenden, dazu auch die entsprechenden Interaktionen mit der Welt benötigt, um jenes Weltwissen erwerben zu können, das notwendig ist, um eine

the issue of chess-playing machines, extending the question to whether a machine, i.e. a UTM, would be able to "learn" in a general sense.
TURING seemed a bit ambivalent in pursuing this issue (cf. TURING 1948 and TURING 1950).
On the one hand, he is apparently aware of the fact that a general ability to learn like a human being requires appropriately "elaborate interactions with the world". The machine would have to be equipped with TV cameras, microphones, loudspeakers etc. so that it was able to interact with the outside world as much as possible. Moreover, it would have to be able to "roam the land', to have "personal initiative', to be "trained" and "educated", in short, it would have to have everything a human child has when learning.
On the other hand, there are passages where he is against an over-imitation of human beings and their natural qualities. Machines with a "reduced body" were just as interesting for him as candidates.
The existence of "artificial intelligence" was to be recognized by the criterion of "imitiation". Whenever a human being reached the "conclusion" that the partner communicating with him/her exclusively via a terminal could as well be human, the descriptors which would normally be applied to human beings only could also be applied to this "artificial intelligence".

3 MEANING, KNOWLEDGE OF THE WORLD AND CONSCIOUSNESS

The discussion triggered off by TURING as to whether machines could develop intelligence comparable to that of human beings and whether they could even develop something like "consciousness" has continued until today and the problem cannot be regarded as definitely solved. In spite of the fact that we have grown accustomed to computers, the issue remains philosophically explosive (DÖBEN-HENISCH 1993). Language seems to be a key to further clarifications in this context; TURING already saw this quite clearly. When enumerating potential applications for an intelligent machine, he said something along the following lines: "Among the above-mentioned applications, learning languages would be the most impressive one because it is the most human of these activities." (TURING 1948, German version 1987: 98)
TURING also predicted, albeit half-heartedly, that a UTM supposed to be in a position to learn any natural language similarly to a human being, and to use it as required in the respective situation, would also need the appropriate interactions with the world so as to acquire the knowledge of the world necessary to use the

language properly (e.g. TUIRNG 1948, German version 1987: 98).

However, the statement that one needs knowledge of the world does not necessarily say anything about (1) "what knowledge" this would exactly be, and (2) "how" such knowledge could be "acquired" or not, as well as (3) "how" this knowledge would have to be "represented internally" so that it can interact with the linguistic system.

Without a theoretical framework of reference within which the above-mentioned questions are to be answered, the answers to (1) through (3) will be random because they have no locus. This also holds true for questions such as (4) "how" a "language system evolves" in a speaker-listener and (5) "how" the "language system" can begin to "interact with the knowledge of the world".

Contrary to the empirical sciences, the questions (2), (4) and (5) have only played a marginal role in philosophy until today - if they played any role at all. As a consequence, questions (1) and (3) as well as special partial aspects of (5), such as the question for the meaning of verbal expressions were treated as isolated static aspects of a process of learning that is per se dynamic. It is against this background that one can at least partially understand why the view that the meaning of verbal expressions can also be reconstructed without reference to facts embedded in consciousness specifically developed in philosophy (cf. FREGE 1892, WITTGENSTEIN 1921, DAVIDSON/HARMAN 1972, BARWISE/PERRY 1983, to name only a few).

However, if one adopts the position of learning or of the learner - referred to as the "agent" here - and accepts questions (2), (4) and (5) as the guiding questions, one is forced not only to assume an agent-world system as a minimal framework, but also to think about the "interface with the world" which the agent has to have to translate the processes required in (2), (4) and (5) into reality.

TURING himself did not answer these questions in their entirety.

To be able to construct the internal function f of an agent we have to make a fundamental decision: should the assumptions concerning f (a) be "arbitrary" or will we (b) orient ourselves on certain "givens"?

A decision in favor of (a) would place the project of constructing f in the realm of a general structural science such as mathematics and would possibly limit it to the computable functions.

A decision in favor of (b) leads to two further decisions: should the construction of f (b.1) depend on certain behavioral data from human test subjects or (b.2) will we

Sprache angemessen verwenden zu können (z. B. TURING 1948; dt 1987:98).

Doch aus der bloßen Feststellung, daß man Weltwissen benötigt, folgt noch in keiner Weise, (1) welches Wissen das genau ist, und (2) wie dieses Wissen erworben werden kann und auch nicht, (3) wie dieses Wissen intern repräsentiert werden muß, damit es mit dem Sprachsystem in Wechselwirkung treten kann.

Ohne Festlegung des theoretischen Rahmens, innerhalb dessen die obigen Fragen beantwortet werden sollen, ist die Beantwortung von (1) bis (3) beliebig, da ortlos. Dies gilt auch für Fragen, wie, (4) wie entsteht ein Sprachsystem in einem Sprecher-Hörer und (5) wie kann das Sprachsystem mit dem Weltwissen in Beziehung treten.

Im Gegensatz zu den empirischen Wissenschaften spielten die Fragen (2), (4) und (5) in der Philosophie bis heute – wenn überhaupt – nur eine marginale Rolle. Dies hatte zur Folge, daß die Fragen (1) und (3) sowie spezielle Teilaspekte von (5), wie z. B. die Frage der Bedeutung sprachlicher Ausdrücke, nur als isolierte statische Aspekte eines an sich dynamischen Lerngeschehens behandelt wurden. Nur vor diesem Hintergrund kann man ansatzweise verstehen, warum es gerade auch in der Philosophie zu der Auffassung kommen konnte, daß sich die Bedeutung sprachlicher Ausdrücke auch ohne Bezugnahme auf Bewußtseinstatbestände rekonstruieren läßt. (Zu verweisen ist hier z. B. auf [FREGE 1892], [WITTGENSTEIN 1921], [DAVIDSON/ HARMAN 1972], [BARWISE/ PERRY 1983], um nur einige zu nennen.)

Nimmt man hingegen den Standpunkt des Lernens bzw. des Lernenden – hier als „Agent" bezeichnet – ein und akzeptiert man die Fragen (2), (4) und (5) als Leitfragen, dann wird man gezwungen, nicht nur ein Agenten-Welt-System als minimalen Rahmen anzunehmen, sondern man muß sich auch über das „Welt-Interface" des Agenten Gedanken machen, wie auch über all jene „internen Zustände" des Agenten, die notwendig sind, um die durch (2), (4) und (5) geforderten Prozesse realisieren zu können.

Turing selbst hat diese Fragen nur ansatzweise erörtert.

Um die interne Funktion f eines Agenten konstruieren zu können, muß man eine Grundsatzentscheidung treffen: Sollen die Annahmen über f (a) beliebig sein oder (b) will man sich an bestimmten Vorgaben orientieren?

Eine Entscheidung für (a) würde das Projekt der Konstruktion von f dem Bereich einer allgemeinen Strukturwissenschaft wie der Mathematik zuordnen, eventuell eingeschränkt auf die berechenbaren Funktionen.

Eine Entscheidung für (b) fordert zwei weitere Entscheidungen heraus: Will man die Konstruktion von f (b.1) abhängig machen von Verhaltensdaten menschlicher Versuchspersonen oder (b.2) will man das Wissen einbeziehen, über das jeder Mensch in der Perspektive seines Selbstbewußtseins verfügt.

Eine Entscheidung für (b.1) führt zum Paradigma der Kognitionswissenschaften, die auf der Basis von Verhaltensdaten (Phonetik, Sprachpsychologie, Physiologie, ...) versuchen, Computermodelle zu erarbeiten, die in ihren Funktionen möglichst weitgehend mit diesen Verhaltensdaten übereinstimmen sollen. Das Vorgehen nach (b.1) besitzt jedoch mindestens einen gravierenden Mangel: eine Modellbildung auf der Basis von Verhaltensdaten ist, bezogen auf die formal möglichen Strukturen interner Prozesse, hochgradig unterbestimmt.

Es verbleibt die Frage, was mit Variante (b.2) ist. Zunächst steht dieser Variante das heute weit verbreitete „Vor-Urteil" entgegen, daß Bewußtseinsdaten für eine ernsthafte Rekonstruktion unbrauchbar sind.

Für eine Einbeziehung von Bewußtseinsdaten sprechen jedoch ein Reihe von philosophischen Arbeiten, die zeigen, daß eine Vielzahl von Erkenntnisaspekten nur durch Bezugnahme auf das eigene Bewußtsein zugänglich sind (z. B. [MACH 1922], [HUSSERL 1913], [MERLEAU-PONTY 1945]). Eine indirekte Bestätigung der philosophischen Argumente findet sich auch in experimentellen psychologischen Arbeiten zur Wahrnehmung und zum Gedächtnis (z. B. in [MURCH/ WOODWORTH 1978], [KLIX 1980], [HOFFMANN 1982], [SHIFF 1980]).

Dieser hier grob als phänomenologisch zu bezeichnende Erkenntnisansatz zeichnete sich in der Vergangenheit allerdings durch das Fehlen einer hinreichenden Sprachkritik aus, was seine Akzeptanz im Rahmen der Philosophie bis heute deutlich beeinträchtigte. Durch Einbeziehung der neuzeitlichen Sprachkritik in die Phänomenologie sowie der Entwicklungen in der modernen Wissenschaftstheorie nach 1970 ((LUDWIG 1978), (BALZER/ MOULINES/ SNEED 1987)) läßt sich jedoch ein Erkenntnisparadigma formulieren, das es ermöglicht, auf der Basis einer phänomenologischen Analyse formale Theorien von Bewußtseinsstrukturen zu erarbeiten, die eine formale Theoriebildung erlaubt, die „nahezu alle" (DÖBEN-HENISCH 1994c) Aspekte des Bewußtseins darzustellen erlaubt und diese Darstellung kontrollierbar macht.

Der soeben unter (b. 2) skizzierte Lösungsansatz ist so angelegt, daß er die Anwendung der Theorie auf sich selbst erlaubt (eine Antwort auf (NAGEL 1986)). Durch die Wahl der Mittel ist eine solche philosophische Theorie ferner „kompatibel" mit jeder empirischen Theorie, da sie jede mögliche empirische Theorie auf formaler Ebene als Teiltheorie enthält. Die Unüberbrückbarkeit zwischen Naturwissenschaften und Geisteswissenschaften erweist sich aus dieser Sicht als bloßes Artefakt falsch gewählter Paradigmengrenzen.

Entscheidend ist nun, daß sich auf der Basis einer solchen formalen philosophischen Theorie des Selbstbewußtseins Simulationsmodelle definieren lassen. Eine spezielle Teilklasse von Simulationsmodellen sind solche, die sich auf den Einsatz von UTM's beschränken.

introduce the knowledge every human being has from the angle of his/her self-consciousness?

A decision in favor of (b.1) leads to the paradigm of the cognitive sciences trying to develop computer models on the basis of behavioral data (phonetics, psychology of language, physiology, ...), with the functions of these models corresponding as far as possible to the behavioral data. However, proceeding according to (b.1) has a serious drawback: the formation of models on the basis of behavioral data is "highly underdetermined" as regards the formal possibilities of the structures internal processes may have.

The question for variation (b.2) remains open. The first response to this variation will be the prejudice (rather widespread nowadays) that data acquired from our consciousness are unsuitable for serious reconstruction.

However, it can be said in favor of the use of data from our consciousness that there is a number of philosophical works showing that many aspects of cognition only become accessible by reference to one's own consciousness (e.g. MACH 1922, HUSSERL 1913, MERLEAU-PONTY 1945). Publications from the field of experimental psychology on perception and memory (e.g. MURCH/WOODWORTH 1978, KLIX 1980, HOFFMANN 1982, SHIFF 1980) have indirectly corroborated the philosophical arguments.

This approach to cognition, which can roughly be described as phenomenological, was in the past characterized by a lack of sufficient instruments of linguistic critique, which has had quite marked adverse effects on its acceptance in the field of philosophy until today. The introduction of modern linguistic critique in phenomenology as well as developments in the modern theory of science after 1970 (LUDWIG 1978, BALZER/MOULINES/SNEED 1987) allow for the formulation of a paradigm of cognition enabling us to elaborate formal theories of consciousness structures on the basis of phenomenological analysis. The formal theory formation we thus have at our disposal makes it possible to "represent" "almost all" (DÖBEN-HENISCH 1994c) aspects of consciousness and renders this representation "controllable".

The approach to a solution described under (b.2) is designed in such a way as to enable "the application of the theory on itself" (an answer to NAGEL 1986). Due to its choice of means, such a philosophical theory is also "compatible" with any empirical theory because on a formal level it contains every conceivable empirical theory as a partial theory. From this point of view, the suggestion that the gap between the sciences and the humanities cannot be bridged turns out to be a mere

artefact of misplaced boundaries separating paradigms.
The decisive point is that "simulation models" can be defined on the basis of such a formal philosophical theory of self-consciousness. Simulation models limited to the use of UTMs are a special partial category. However, simulation models are by no means mere spin-offs of a self-contained formal theory of consciousness. In view of the mind-boggling complexity of processes of consciousness we rather have to assume that the simulations are a necessary aid so that theories of self-consciousness with a certain degree of complexity can be developed at all. In this sense, simulation models are "instruments for philosophical experiments".
The procedure briefly described here may well form the basis of a new "discipline" which - depending on your point of view - could be "called Computer-Aided Philosophy" [CAP], or better "Computational Philosophy" [CP] if you feel more akin to a philosophical interest in cognition, or "Artificial Consciousness" [AC] if you want to place it in the general framework of computer science and pursue it as a new branch in parallel to the existing Artificial Intelligence [AI] branch.

4 KNOWBOTIC INTERFACE

To translate the concept of computational philosophy into reality it is required that a formal theory of self-consciousness as well as a suitable simulation model are developed at the same time.
BW1 is designed to be "the first prototype" of such a theory-guided "simulation model of self-consciousness" implemented in the framework of a Knowbotic Interface "[KInt]".
First formulated in the spring of 1994 (DÖBEN-HENISCH 1994a) and further developed theoretically in cooperation with Prof. HOCHE in a research group in Bochum (DÖBEN-HENISCH 1994b), the KInt addresses the consciousness issue and falls back on an idea proposed by TURING, whose vision of a learning machine was a child-machine: it would, of course, be a UTM whose nature would allow for it to be subjected to processes of education and training like a child (TURING 1950: 1987, pp. 177f).
Along with the KInt, a set of software "building blocks" would be made available so that the UTM Child Machines can be defined as TURING imagined them, complete with random UTM Worlds which "normal people", too, can enter, albeit in the guise of UTM Child Machines. The author calls these UTM Child Machines Knowbots (derived from to know + robot = knowbot)

Simulationsmodelle sind jedoch keinesfalls bloße „Abfallprodukte" einer in sich abgerundeten formalen Theorie des Bewußtseins. Aufgrund der schwindelerregenden Komplexität der Bewußtseinsprozesse muß man vielmehr annehmen, daß die Simulationsmodelle ein notwendiges Hilfsmittel sind, um überhaupt Theorien des Selbstbewußtseins mit einem gewissen Komplexitätsgrad ausarbeiten zu können. In diesem Sinne sind Simulationsmodelle Instrumente für philosophische Experimente.
Diese hier nur kurz skizzierte Vorgehensweise kann eine neue Disziplin begründen, die man, je nach Standpunkt, entweder als Computergestützte Philosophie [CGP] (Computer-Aided Philosophy [CAP] – oder besser Computational Philosophy [CP] –) bezeichnen könnte, wenn man sich mehr dem philosophischen Erkenntnisinteresse zugehörig fühlt, oder als Künstliches Bewußtsein [KB] (Artificial Consciousness [AC]), will man sie als neuer Sparte neben der bisherigen Künstlichen Intelligenz [KI] (Artificial Intelligence) im Rahmen einer allgemeinen Informatik betreiben.

4 KNOWBOTIC INTERFACE

Die Umsetzung des Konzeptes einer computergestützten Philosophie setzt voraus, daß man simultan sowohl eine formale Theorie des Selbstbewußtseins entwickelt wie auch ein zu dieser Theorie passendes Simulationsmodell.
BW1 soll der ersten Prototyp eines solchen theoriegeleiteten Simulationsmodells des Selbstbewußtseins sein, das im Rahmen eines Knowbotic Interface [KInt] realisiert wird.
Im Frühjahr 1994 zum ersten Mal formuliert (DÖBEN-HENISCH 1994a) und in einem Forschungsseminar 1994 zusammen mit Prof. HOCHE in Bochum theoretisch weiter ausgearbeitet (siehe DÖBEN-HENISCH 1994b), greift die Idee des KInt neben der Bewußtseinsproblematik einen Gedanken TURINGs wieder auf, der seine Vision von einer lernenden Maschine in die Gestalt einer Kind-Maschine gekleidet hatte: die Kind-Maschine – die natürlich eine UTM ist – sollte so beschaffen sein, daß sie wie ein Kind einem Erziehungs- und Lernprozeß unterworfen werden könnte (TURING 1950:1987 pp.177f).
Mit dem KInt wird ein Softwarebaukasten zur Verfügung gestellt, der es erlaubt, genau solche UTM-Kind-Maschinen zu definieren wie sie TURING vorschwebten, dazu beliebige UTM-Welten, in die auch „normale Menschen" „eintreten" können, letztere allerdings unter dem Gewande von UTM-Kind-Maschinen. Diese UTM-Kind-Maschinen nennt der Verfasser Knowbots (to know + robot = knowbot) und die UTM-Kind-Maschinen, die Menschen „verkleiden", nennt er Pseudo-Knowbots. (Die Anregung zum Begriff Knowbot bekam der Verfasser aus intensiven Gesprächen mit Christian Hübler von der Gruppe knowbotic research. Die Gruppe kr+cf verwendet auch die Bezeichnung knowbots, allerdings

in einem anderen Sinne. Ferner wird der Begriff „knowbot" im Internet gelegentlich für intelligente Agenten benutzt, die im Netz verschiedenste Informationen sammeln. Diese „gewöhnlichen" Knowbots haben mit den Knowbots vom KInt nur den Namen gemein.)
Das KInt bietet somit die Möglichkeit, Knowbots von Menschen trainieren und unterrichten zu lassen, ohne daß die Knowbots die Pseudo-Knowbots notwendigerweise als etwas von ihnen Verschiedenes erkennen müssen. Damit erschließt das KInt eine neue Variante des TURINGschen Imitationstests.
Gegenüber TURING wird in KInt entscheidendes Gewicht darauf gelegt, daß sowohl die innere Struktur der Knowbots, d. h. ihr Bewußtsein, wie auch die Struktur der KInt-Welt mit der uns Menschen bekannten Welt soweit übereinstimmt, daß im Prinzip alle bedeutungsrelevanten Sachverhalte simuliert werden können, die im Kontext der dem Menschen bekannten natürlichen Sprachen auftreten können.
Aufgrund der Flexibilität des KInt kann man natürlich auch andere Welt- und Bewußtseinsstrukturen als die aus der menschlichen Welt bekannten simulieren. Prinzipiell könnte man im Rahmen des KInt daher auch Themen behandeln, die sonst nur im Rahmen von Science-fiction-Romanen vorkommen: künstliche Bewußtseinsstrukturen, die sich von den menschlichen substantiell unterscheiden, die ihre eigene Sprache sprechen, die eventuell unsere menschliche Sprachen verstehen, wir aber nicht ihre. Im Rahmen des KInt kann man diese Experimente direkt ausführen.

5 THE BLINDS WORLD I

Die Grundstruktur von BW1 ist in Bild 1 zu sehen.
Zunächst wird zwischen einem Server-Programm unterschieden, das eine bestimmte Welt verwaltet, und einem Client-Programm. Letzteres kann ein Knowbot oder ein Pseudoknowbot sein. Client- und Serverprogramme können entweder auf dem gleichen Rechner oder auf verschiedenen Rechnern gestartet werden, d. h. diese Client-Server-Architektur ist voll netzwerkfähig. Es können auch mehrere Server-Programme gleichzeitig nebeneinander auf dem gleichen Rechner oder auf verschiedenen Rechnern existieren.
Die Welt, die durch das Server-Programm realisiert werden kann, wird zu Beginn aus einer Textdatei geladen und intern vom Server als Weltdatenstruktur aufgebaut. Die Textdatei kann wie ein normales Textdokument editiert werden, d. h. der Benutzer kann sich nach Belieben eine Welt definieren: das Innere eines Hauses, eine Stadt, eine Behörde, ein bestimmtes Land oder einen ganzen Planeten. In der ersten Version sind die möglichen Welten eingeschränkt auf 2-dimensionale Darstellungen, in denen 5 Schichten unterschieden werden. Die Welten von BW1 ähneln daher alle bunten

whereas the UTM Child Machines "disguising" human beings are called "Pseudo-Knowbots". (The name Knowbot was inspired by the author's intense discussions with Christian Hübler of the group knowbotic research. The group kr+cf also uses the term "knowbot", albeit in a different sense. The word "knowbot" is occasionally also used on the Internet to denote intelligent agents collecting information on the net. These "ordinary" knowbots and the knowbots of KInt only have their name in common.)
Thus, the KInt offers human beings the opportunity to train and educate knowbots without the knowbots necessarily having to recognize the pseudo-knowbots as something that differs from them. The KInt opens up a new variation of TURING's imitation test this way.
In contrast to TURING's approach, the KInt greatly emphasizes the fact that the internal structure of the knowbot, viz. its consciousness, and the structure of the KInt's world correspond to the world known to us human beings insofar that all facts relevant to meaning in the context of the natural languages known to man can in principle be simulated.
Due to the flexibility of the KInt other world and consciousness structures than those known from the human world can be simulated, too. In principle, topics to be dealt with in the framework of the KInt could well be derived from the realm of science fiction novels: artificial consciousness structures substantially different from ours, speaking languages of their own, possibly understanding human languages while we do not understand theirs. In the KInt framework such experiments can be conducted directly.

5 THE BLIND'S WORLD I

Illustration 1 shows the basic structure of BW1.
First, one has to differentiate between "a server program" administering a certain world and a "client program". The latter can be a knowbot or pseudo-knowbot. Client and server programs can either be started on the same or different computers, i.e. the "client-server architecture" is fully "network-compatible". Several server programs may exist side by side on the same computer or on different computers.
The "world" implemented by the server program is initially loaded from a text file and internally built up by the server as a world data structure. The text file may be edited like a normal text document, viz. the user may define the world at will: the interior of a house, a town, an authority, a certain country or an entire planet. In the first version the possible world representations are limited to two dimensions with five different strata.

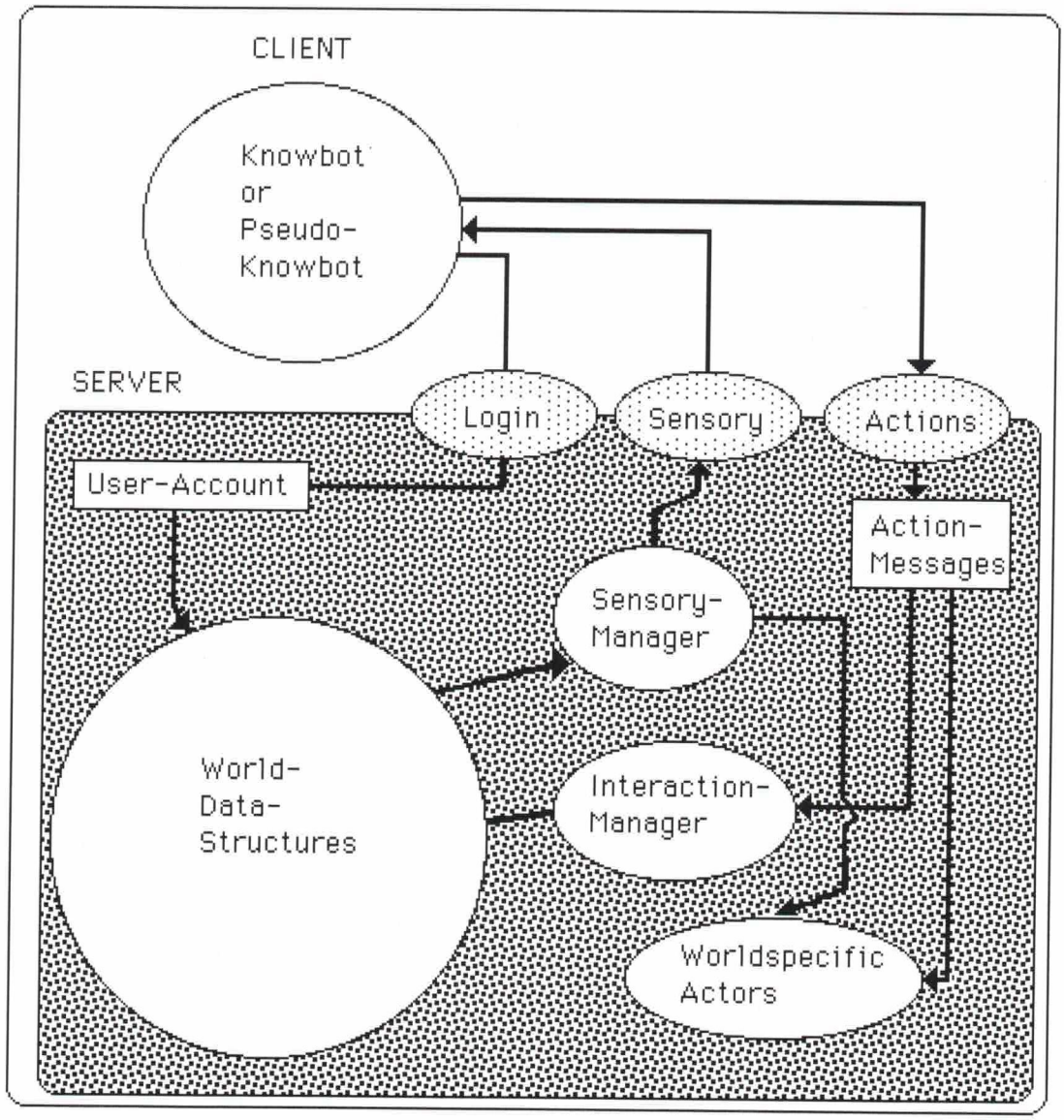

Bild 1

The BW1 worlds all resemble colorful maps, with multi-colored pictographs moving around on them.
All objects in BW1 are "multi-sensorial objects". Each object is depicted as a figure with an internal structure and colored patterns in the visual dimension, it may also come with certain sounds in the acoustic dimension, a certain smell in the olfactory dimension, a specific taste in the gustatory one and various haptic qualities in the tactile dimension. These qualities may also change continuously according to internal conditions or exterior influences. The multi-sensorial character of

Landkarten, in denen sich farbige Pictogramme hin und her bewegen.
Sämtliche Objekte von BW1 sind multisensorielle Objekte. Jedes Objekt hat nicht nur in der visuellen Dimension eine Gestalt mit innerer Struktur und farblichen Mustern, sondern es kann in der akustischen Dimension mit Geräuschen verbunden sein, in der olfaktorischen mit einem Geruch, in der gustatorischen mit einem Geschmack und in der taktilen mit diversen Berührungswerten. Diese Werte können sich außerdem in Abhängigkeit von inneren Zuständen oder von äußeren Einflüssen beständig ändern. Dieser multisen-

sorielle Charakter aller Objekte ist eine notwendige Konsequenz der Forderung, daß die (Pseudo-)Knowbots mit einer BW1-Welt ausschließlich über Sinnesorgane kommunizieren sollen.

Die prototypische Welt von BW1 ist insgesamt eine sehr einfache Welt: eine Welt wird als eine große Fläche angesehen, deren Enden z. Zt. nicht festgelegt sind. In Anlehnung an den biblischen Schöpfungsbericht besteht eine Welt grundsätzlich aus Wassermassen, in die man nach Belieben isolierte oder verbundene Landmassen einfügen kann. Jedes Land kann mittels einer Liste verfügbarer Landschaftstypen gestaltet werden. Es stehen z. B. zur Verfügung „Wüste", „Gras", „Wald", „Felsen", „Fluß", „See" und „Weg". In eine solcherart charakterisierte Landmasse kann man dann beliebig viele einzelne Objekte einfügen. Als Objekttypen stehen zur Verfügung die Objektklassen Pflanzen und Lebewesen. Bei den Pflanzen wird unterschieden zwischen „Kleinpflanzen", „Büschen" und „Bäumen". Bei den Lebewesen zwischen „Kleintieren", „Raubtieren" und „Großwild", wobei hier als besondere Objektart noch (Pseudo-)Knowbots auftreten können, falls sich solche als Clients bei dem Weltprozeß anmelden.

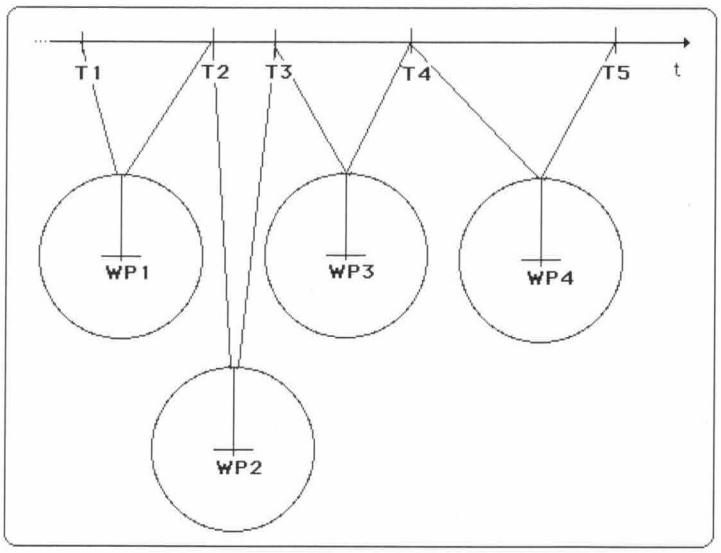

Bild 2

Neben den sensorischen Eigenschaften besitzen sowohl die Pflanzen wie die Lebewesen zusätzliche Eigenschaften. Pflanzen haben z. B. bestimmte Nährwerte, die sich auf die Lebewesen unterschiedlich auswirken können. Die Lebewesen sind alle zumindest mit den Grundbedürfnissen „Hunger", „Durst", „Müdigkeit" und „Fortpflanzungsbedürfnis" ausgestattet. Ferner können sie „Schmerzen" und „Angst" empfinden.

Nachdem sich Clients durch ein Login bei einem Server als „Weltbürger" angemeldet haben, verläuft der Weltprozeß als eine Folge von Weltzyklen. Ein Weltzyklus ist ein Verarbeitungszyklus im Server, der eine bestimmte physikalische Zeit benötigt (vgl. Bild 2).

In Abhängigkeit von der anfallenden Datenmenge und der Leistungsfähigkeit der Hardware kann die physikalische Zeit für einen einzigen Weltzyklus stark schwanken. Innerhalb des Weltprozesses definiert ein Durchgang aber genau eine minimale Weltzeiteinheit, d. h. der Weltprozeß realisiert eine logische Zeit.

Innerhalb eines Weltzyklus gibt es im wesentlichen zwei Aktivitäten: (1) Der Interaction Manager [IM] liest sämtliche Aktions-Mitteilungen, die seit der letzten Auswertung im Brief-

all objects is a necessary consequence of the demand that the (pseudo-)knowbots communicate with the BW1 world excusively through sense organs.

All things considered, the prototype BW1 world is very simple: it is a world seen as a huge surface area the boundaries of which have not been defined so far. Inspired by the biblical account of the creation of the world, these worlds are basically "expanses of water" into which isolated or linked "landmasses" can be inserted at will. Each country can be designed from a stock of "landscape types". These include, among other things, "deserts", "grassland", "woods", "rocks", "rivers", "lakes" and "paths". Any number of individual objects can then be placed on the land characterized this way. The "types of objects" available are "plants" and "living beings". Plants fall into the categories "small plants", "bushes" and "trees", living beings can be "small animals", "predacious animals" and "big game"; (pseudo-)knowbots may become special objects here if they register as clients in the world process.

Apart from the sensory qualities, plants and living beings have additional characteristics. Plants have e.g. certain "nutritional values" which can have various effects on living beings. Living beings are endowed with at least the basic needs such as "hunger", "thirst", "fatigue" and the "need to procreate". They may also feel "pain" and "fear".

Once the clients have logged in, registering as "citizens of the world" on a server, "the world process" passes as a sequence of world cycles. A "world cycle" is a

processing cycle of the server which requires a certain physical period of time. (ill. 2)
Depending on the quantity of data to be processed and on hardware capacity the physical time required for a single world cycle may vary strongly. One cycle within the world process corresponds exactly to a minimal unit of world time, i.e. the world process functions on the basis of "logical time".
Within a world cycle there are basically two activities: (1) the "interaction manager [IM]" reads all action messages which have accumulated in the action-message mailbox since the last evaluation. It checks each message for possible collisions. Whenever it detects a collision, it calculates a corrected final position for the movement and passes it on. Once the repercussions of the action messages have been recorded, (2) the "sensory manager [SM]" takes over. It calculates all possible sensory stimuli which may occur "according to the world laws" in the places where living beings are located at that point of time, and it does so for each living being, especially for the (pseudo-) knowbots. The totality of the values the SM calculates for a certain living being is then "rolled into" a sense message and sent to the being. When the SM has completed its task, a world cycle is over.
Action messages and sense messages are strings of signs (ASCII strings) the syntactic structure of which is determined by context-free grammars.

6 KNOWBOTS

The KInt described above is required to make a minimum of "ambient conditions" available for the simulations of a structural equivalent to consciousness.
The following passages refer to the framework conditions envisaged for the first BW1 experiments involving the phenomenon of self-consciousness.

6.1 Minimum Reactive System

In BW1 it is assumed that the knowbots have at least sense and efferent organs.
The sensory capacities are limited to the senses of "hearing', "smelling', "touching" and "tasting" (vision was left out in BW1). The developments of values communicated to the "interior" of the knowbot by the sensors are processed there; they are transformed into a three-dimensional sensory map maintaining their topology. From here, they are available for further use.
The efferent capacities are based on a finite set of elementary actions (walking, turning around, moving one's hands, grasping something, putting something away, eating, drinking, sleeping, mating, playing) which can be expressed in terms of parameters. Several such elekasten für Aktions-Mitteilungen eingegangen sind. Er überprüft dann jede Mitteilung daraufhin, inwieweit dadurch irgendwelche Kollisionen eintreten können. Stellt er eine Kollision fest, dann berechnet er eine korrigierte Endposition für die Bewegung und gibt diese weiter. Nachdem die Wirkungen der Aktions-Mitteilungen erfaßt sind, tritt (2) der Sensory-Manager [SM] auf den Plan. Für jedes Lebewesen, insbesondere natürlich für die (Pseudo-)Knowbots, berechnet er sämtliche mögliche sensorische Reize, die an der Stelle, wo sich das Lebewesen gerade befindet, „nach den Weltgesetzen" möglich sind. Die Gesamtheit der Werte, die der SM für ein bestimmtes Lebewesen findet, werden dann in eine sensorische Mitteilung verpackt und ihm zugeschickt. Wenn der SM seine Arbeit beendet hat, ist ein Weltzyklus beendet.
Die Aktions-Mitteilungen wie auch die sensorischen Mitteilungen sind Zeichenketten (ASCII-Strings), deren syntaktischer Aufbau durch kontextfreie Grammatiken (Contextfree Grammars) bestimmt sind.

6 KNOWBOTS

Das vorstehend skizzierte KInt ist notwendig, um minimale Umgebungsbedingungen für die Simulation eines strukturellen Bewußtseinsäquivalentes bereit zu stellen.
Es seien hier nun jene Rahmenbedingungen geschildert, die in BW1 für erste Experimente mit dem Phänomen Selbstbewußtsein vorgesehen sind.

6.1 Minimales reaktives System

In BW1 wird angenommen, daß die Knowbots zumindest über eine Sensorik und eine Effektorik verfügen können.
Die Sensorik ist beschränkt auf die Sinnesarten „Hören", „Riechen", „Tasten" und „Schmecken" (Das Sehen wurde für BW1 ausgeklammert). Die Wertverläufe, die die Sensoren in das „Innere" des Knowbot übermitteln, werden dort zunächst in einem 3dimensionalen sensorischen Plan unter Bewahrung ihrer Topologie aufbereitet. Von hier aus stehen sie zur weiteren Verwendungen zur Verfügung.
Die Effektorik basiert auf einer endlichen Menge von Elementar-Handlungen (Gehen, sich drehen, die Hände bewegen, etwas fassen, etwas aus der Hand geben, essen, trinken, schlafen, Paarungsverhalten, spielen), die parametrisiert werden können. Mehrere solcher Elementarhandlungen können zu komplexen Handlungen zusammengefaßt werden, wobei die Parametrisierungsbedingungen erhalten bleiben. In BW1 wird angenommen, daß jeder Knowbot „von Geburt an" über eine Reihe von vorgefertigten Verhaltensschemata verfügt, die basale Verhaltensweisen wie Nahrungssuche, Nahrungsaufnahme, Fortpflanzung und Flucht sicherstellen. Ein Verhaltensschema ist eine Menge von Verhaltensregeln, wobei jede Verhaltensregel ein Bi-Konditional der Art darstellt: ZIEL Z kann erreicht werden genau dann, wenn BEDINGUNG B1 ist erfüllt & ... & BEDINGUNG Bn

ist erfüllt. Der linke Teil der Regel heißt ihr „Kopf" und der rechte Teil ihre „Bedingung". Im allgemeinen Fall kann im Kopf mehr als ein Ziel genannt werden und in der Bedingung können auch Disjunktionen (oder) auftreten. Der „Schema"-Charakter von Verhaltensregeln entsteht dadurch, daß als Argumente der Ziele und Bedingungen auch Variablen auftreten können. Ein solcherart gegebenes Verhaltensschema steht dann für eine ganze Menge möglicher konkreter Verhaltensregeln, je nachdem, welche konkreten Werte für die Variablen eingesetzt werden dürfen.

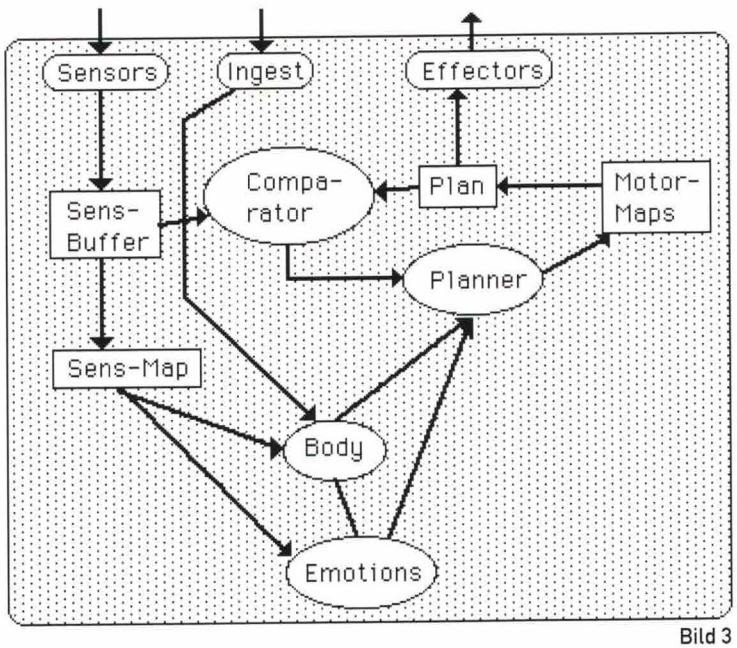

Bild 3

Zur Effektorik gehört auch eine einfache motorische Planung, die die Aktivierung der jeweiligen Verhaltensschemata vornimmt und die Koordinierung der auszuführenden Handlungen überwacht. Dazu gehört sowohl die Verwertung der Rückmeldungen von der Sensorik wie auch die Umsetzung der „Tendenzvorgaben" der emotionalen Steuerungseinheit.

Die Emotionen bilden zusammen mit diversen Körperzuständen eine dritte Basiskomponente. Zu den Körperzuständen gehören „Energiebilanz", „Flüssigkeitsbilanz", „Wach-/Schlafzustand". Zu den Emotionen werden Triebe gerechnet (Hunger, Durst, Müdigkeit, sexuelles Verlangen, Spielverhalten), Schmerzen, Angst sowie ein nicht weiter definierter „emotionaler Gesamtzustand". Diese Komponenten beeinflussen sich gegenseitig, wobei der emotionale Gesamtzustand die „Führungsrolle" hat und die Triebe wie auch die Schmerzen ihn je nach aktueller Stärke „übertönen" können. Für jede Emotion gibt es Aktivatoren oder Deaktivatoren. Hunger wird z. B. durch eine negative Energiebilanz aktiviert, und durch Nahrungsaufnahme – in Abhängigkeit von der Beschaffenheit der Nahrung – deaktiviert.

Zu jedem Zeitpunkt gibt es immer eine Emotion, die „dominant" ist. Diese beeinflußt den Planer bei der Gestaltung seiner Handlungspläne. Wechseln die Emotionen, dann führt dies u. U. auch zur Unterbrechung laufender Handlungen bzw. zur Wiederaufnahme eines vorher unterbrochenen Planes.

Die motorische Planungseinheit wird somit vollständig von dem aktuellen emotionalen Zustand dominiert, falls es nicht gerade einen Verhaltensplan gibt, der eine noch „höhere" Priorität hat.

Die Körperzustände werden sowohl von Handlungen wie

mentary actions can be summarized in complex actions, with the conditions for the expression in parameters remaining intact. In BW1 it is assumed that each knowbot has a number of behavior patterns existing "ab ovo" which ensure basic behaviors such as searching for food, eating, procreation and flight. "A behavior pattern is a set of behavior rules, with each behavior rule in turn being bi-conditional as follows: OBJECT O can precisely be attained if CONDITION C1 is fulfilled & ... & CONDITION Cn is fulfilled. The left-hand side of the rule is its "head" and the right-hand side of the rule is its "condition". In a general case more than one object can be formulated in one's mind and disjoint (or) statements are possible in the condition. The "pattern" character of behavior rules evolves because variables can also appear as arguments for the objects and conditions. Such a behavior scheme thus stands for a whole set of possible specific behavior rules, depending on the concrete values that may be substituted for the variables.

Response mechanisms also include simple motor planning activating the respective behavior patterns and coordinating the actions to be taken. This encompasses the utilization of sensory feedback as well as the implementation of "tendencies" given by the emotional control unit.

"Emotions" and various "physical states" form the third basic component. Physical states include the "energy balance", "balance of fluids" and "waking/sleeping states". Instincts and drives are counted among emotions (hunger, thirst, fatigue, sexual desire, play be-

havior), as are pains, fear and an "overall emotional state" that is not identified any further. These components have a bearing on one another, with the overall emotional state "dominating"; however, depending on their respective extent, instincts and pains may "override" it.
There are "activators" and "deactivators" for each emotion. Hunger is e.g. activated by a negative energy balance and deactivated by food intake, depending on the properties of the food.
At any given time there is a "dominating" emotion. It has a bearing on the design of the action plans. Changes in the emotions may lead to the interruption of ongoing actions or the reversion to a plan previously abandoned.
The motor planning unit thus completely depends on the current emotional state unless a behavior plan activated at the time has a "higher" priority.
Physical states are influenced by actions as well as sensory inputs; in turn, they have a bearing on the emotions.
The entirety of these basic components forms what in BW1 is called a "reactive system" (ill. 3), viz. these components are per se a functional unit in a position to move around in the world and to feed itself, albeit poorly. Further components may now supplement the reactive system, thus modifying it.

6.2 The Necessity of Learning
One of the major drawbacks of the reactive system is its inflexibility; except for simple variations within given boundaries, it is not capable of learning anything. It converts impacts from its surroundings into certain modes of reactions "in a certain way", but the mode of reaction f does not undergo any appreciable change while the system is in existence. The "mode of reaction f" can be regarded as a function mapping elements from the set S of stimuli on elements of the set R of reactions.
To turn a reactive system incapable of learning into a system "capable of learning" substantial changes in the mode of reaction f must be possible, i.e. one must be in a position to replace the current version of f by a different version f" from the set of "possible modes of reactions" F.
In such a system there is an additional learning function L which substitutes a new version f" for the current version of a mode of reaction f by reverting to the set F. However, "blind" learning, which replaces a current mode of reaction f by "any mode of reaction f" that is different", is not very effective. Improvements attained this way require "long" periods of time as well as high

auch von sensorischen Ereignissen beeinflußt und sie wiederum beeinflussen die Emotionen.
Die Gesamtheit dieser Basiskomponenten bildet das, was in BW1 ein reaktives System genannt wird (vgl. Bild 3), d. h. dieses Komponenten bilden für sich genommen schon eine funktionsfähige Einheit, die in der Lage ist, sich in der Welt zu bewegen und sich notdürftig mit Nahrung zu versorgen. Zu diesem reaktiven System können jetzt weitere Komponenten hinzutreten, um es zu modifizieren.

6.2 Die Notwendigkeit des Lernens
Ein großer Nachteil des reaktiven Systems besteht in seiner Starrheit; es kann, sieht man von einfachen Variationen innerhalb vorgegebener Grenzen einmal ab, nichts dazu lernen. Es setzt „auf eine bestimmte Weise" Einwirkungen der Umwelt in bestimmte Reaktionsweisen um, ohne daß sich diese Reaktionsweise f für die Dauer der Existenz dieses Systems nennenswert ändert. Die „Reaktionsweise f" kann man als eine Funktion auffassen, die Elemente aus der Menge S der Stimuli in Elemente der Menge R der Reaktionen abbildet.
Damit aus einem nicht lernfähigen reaktiven System ein lernfähiges System wird, bedarf es der Möglichkeit, die Reaktionsweise f substantiell ändern zu können, d. h. man muß eine aktuelle Version von f durch eine andere Version f" aus der Menge F der möglichen Reaktionsweisen ersetzen können.
In solch einem System gibt es eine zusätzliche Lernfunktion L, die die aktuelle Version einer Reaktionsweise f durch Rückgriff auf die Menge F der möglichen Reaktionsweisen gegen eine neue Version f" austauscht.
Ein blindes Lernen, das eine aktuelle Reaktionsweise f durch „irgendeine andere" Reaktionsweise f" ersetzt, ist jedoch nicht sehr effektiv. Bei großer Populationsdichte und hohen Vermehrungsraten läßt sich über „lange" Zeiträume möglicherweise auch auf diese Weise eine Verbesserung erzielen. Verhaltensbezogenes Lernen beschränkt sich jedoch auf den Bereich eines Individuums und ereignet sich in einem vergleichsweise „kurzen" Zeitraum. Dies scheint nur möglich zu sein, wenn man annimmt, daß verhaltensbezogenes Lernen als informiertes Lernen realisiert ist. Das Mindeste, was man dann von einem lernfähigen System fordern müßte, wäre, daß es über Bewertungskriterien verfügen kann, die über die Ausgestaltung der Reaktionsweise f mitentscheiden können. Eine solche Evaluation könnte z. B. Umwelteinwirkungen zum Ausgangspunkt von Bewertungen zu machen.
Neben der reinen Verhaltensfunktion f gibt es dann sowohl eine Bewertungsfunktion ev, die Umwelteinwirkungen mittels positiver oder negativer Bewertungszahlen indiziert, und eine modifizierte Lernfunktion L, die bei der Auswahl von

neuen Verhaltensweisen f" aus F auch Bewertungszahlen berücksichtigt.

(1) F : $S \rightarrow R$
(2) ev : $S \rightarrow E$
(3) L : $F \times E \rightarrow F$

6.3 Lernen in BW1

Die Umsetzung des abstrakten Postulates, Knowbots als Systeme mit einer informierten Lernfunktion zu konzipieren, geschieht in BW1 an mehreren Stellen zugleich.

Die Bewertungsfunktion ev wird in einem Knowbot in erster Linie durch die Auswirkungen von Ingestionen und Perceptionen auf körperliche und emotionale Zustände realisiert. Wenn z. B. das Essen eines bestimmten Nahrungsmittels zur Linderung des Hungers führt, dann korreliert solch ein Essen samt den zugehörigen Geschmacks- und Geruchsempfindungen mit einem positiven Wert. Dieser kann benutzt werden, um sowohl ein bestimmtes Objekt wie auch eine bestimmte Handlung im Kontext einer bestimmten Bedürfnisbefriedigung zu bewerten.

Damit solche Zusammenhänge von Handlungen, Perzeptionen, Ingestionen sowie deren Auswirkungen, die sich in einem bestimmten Zeitintervall li,j ereignen, auch zu einem späteren Zeitpunkt tj+c verfügbar sein können, bedarf es der Möglichkeit, diese in geeigneter Weise zu „erinnern", d. h. sie in „erinnerbarer Weise zu speichern". Man benötigt also eine Gedächtnisstruktur, die in der Lage ist, Objekte unter Berücksichtigung ihrer unterschiedlichen sensorischen Qualitäten sowie ihrer positiven oder negativen Auswirkungen auf emotionale Zustände und unter Berücksichtigung beteiligter Handlungen und relevanter räumlicher Strukturen „abzulegen" und „nach Bedarf" wieder zu „finden".

Schließlich muß auch die Möglichkeit bestehen, solcherart bewertete erinnerbare Zusammenhänge für die konkrete Verhaltensplanung nutzbar zu machen. D. h. wenn die Handlungs-Planungseinheit eine bestimmte Handlungsfolge beschlossen hat, um ein bestimmtes Ziel zu erreichen, dann muß es in einem lernfähigen System möglich sein, daß es die „vorgegebenen Verhaltensschemata" aufgrund von „Erfahrungen" „abändert". Solche Änderungen können auf unterschiedlichster Ebene stattfinden. Insgesamt stellt die Notwendigkeit der möglichen Einbeziehung von „erinnerbaren Zusammenhängen" in aktuelle Handlungsplanungen eine zusätzliche Anforderung an die Art und Weise dar, wie diese Zusammenhänge „gespeichert" werden.

Aufgrund der sehr kurzen Realisierungszeit konnten in BW1 diese Gedächtnisstrukturen vorerst nur sehr rudimentär implementiert werden.

6.4 Sprache

Zu einer natürlichen Sprache gehört neben dem Ausdrucks-

population density and multiplication rates. By contrast, behavior-related learning is limited to the individual and happens in a comparatively "brief" period of time. This only seems possible if one assumes that behavior-related learning is put into practice as "informed learning". The least one could then demand from a system capable of learning would be the existence of "rating criteria" which it can use to co-determine the design of the mode of reaction f. Such an evaluation could e.g. base ratings on environmental impacts.

Apart from the pure behavioral function f there would also exist a rating function ev indexing environmental impacts with positive or negative ratings, and a modified learning function L which takes ratings into consideration when selecting new modes of behavior f" from F.

(1) F : $S \rightarrow R$
(2) ev : $S \rightarrow E$
(3) L : $F \times E \rightarrow F$

6.3 Learning in BW1

In BW1 the abstract postulate that knowbots are conceived of as systems with an informed learning function is translated into reality in several places simultaneously.

In a knowbot the "rating function" ev materializes first and foremost through the impacts which ingestions and perceptions have on physical and emotional states. If e.g. eating a certain foodstuff appeases hunger, such food, including its taste and smell, correlates with a positive rating. This figure can be used to evaluate a certain object as well as a certain action in the context of a certain need gratification.

Making available, at a later point of time t_{j+c}, correlations of actions, perceptions, ingestions and their effects taking place in a certain interval of time Δ_{ij} requires the ability to "remember" them appropriately, viz. they have to be "stored in such a way that they can be remembered'. This requires a memory structure capable of "filing objects away" and "finding them when required" while taking into consideration their different sensory qualities as well as their positive or negative impacts on emotional states and recording the actions involved and the relevant spatial structures.

Finally, there must also be a possibility to utilize correlations that have been rated and made retrievable as described above when planning concrete behavior. That is to say, if the action-planning unit has decided that a certain sequence of actions is required to reach

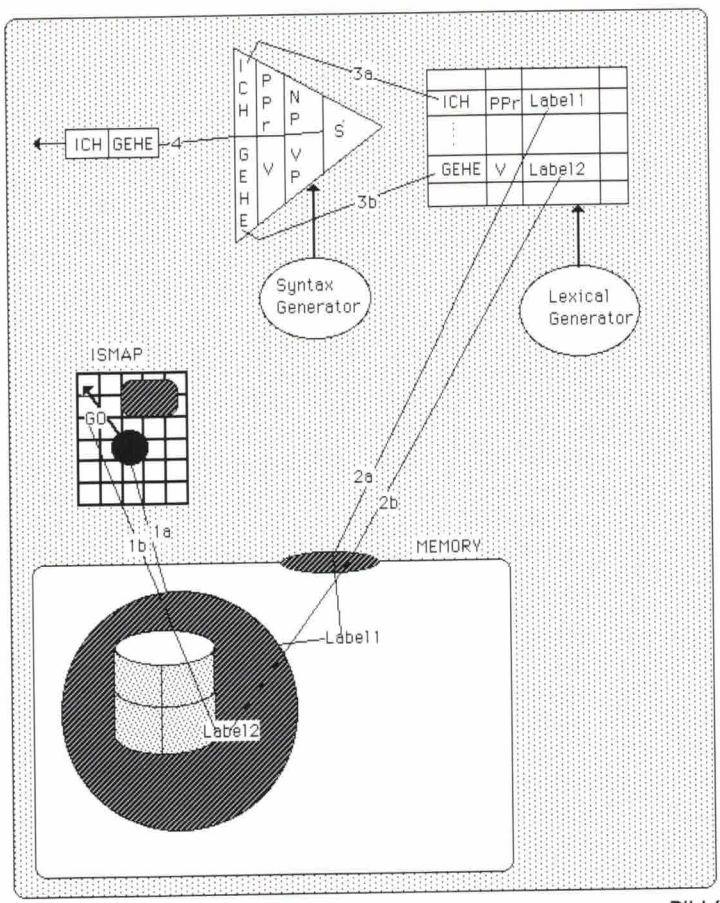

Bild 4

material (Laute, Zeichen) auch das, was die sprachlichen Ausdrücke „bezeichnen", was sie „sagen" und das, was sie dadurch „bewirken", daß man sie benutzt. Diese über das reine Ausdrucksmaterial hinausgehenden Aspekte eines sprachlichen Ausdrucks sollen hier vereinfachend die „Bedeutung" eines sprachlichen Ausdrucks genannt werden.

Für uns Menschen hängt die sprachliche Bedeutung durchgängig mit unserem Selbst- und Weltbezug zusammen.

Für die Knowbots von BW1 steht ein Selbst- und Weltbezug nur in dem Umfang zur Verfügung, wie ihn das zuvor geschilderte informiert lernfähige reaktive System bereitstellt. Wie sich im weiteren Verlaufe zeigen wird, ist diese Struktur für die Realisierung differenzierter Bedeutungsstrukturen, wie wir sie aus der menschlichen Alltagssprache kennen, noch bei weitem zu einfach. Nur primitivste Benennungen und einfachste 1-, 2-, 3-Wort-Sätze sind möglich.

Das Sprachmodul verfügt über mehrere Teilmodule.

Eine wichtige Basiseinheit bildet das Lexikon, das unterschiedliche Teilfunktionen umfaßt. Ein Wortgenerierer versucht aus einer Menge elementarer Laute und Intonationsmustern Phonemsequenzen zu bilden, die im Lexikon gesammelt werden. Im Lexikon können sich aber nur solche Phonemsequenzen halten, die innerhalb eines bestimmten Zeitintervalls von außen „bestätigt" werden. Ein Wort-Objekt-Generierer bildet Wort-Objekt-Hypothesen, d. h. er versucht Worte mit Zuständen oder Handlungen zu verknüpfen. Die aktuelle Situation besitzt für solche Hypothesenbildungen dabei Priorität. Auch die Wort-Objekt-Hypothesen können sich nur halten, wenn sie „von außen" bestätigt werden.

Um die Grammatik zu verstehen, muß man sich klar machen, welche Aufgabenstellungen innerhalb einer sprachlichen Kommunikation zu bewältigen sind. Grundsätzlich wird in BW1 zwischen Sprechen und Hören unterschieden. Bild 4 demonstriert den Fall des Sprechens:

Ausgehend von dem interpretierten sensorischen Plan (interpreted sens-map [ISMAP]), werden von einem Pragmatik-Modul – genannt language map [LMAP] – die Unterschei-

a certain goal, a system capable of learning must allow for the "given behavior patterns" to be modified on the basis of "experiences". Such modifications may take place on various levels. All in all, the necessity of an option to include "correlations that can be remembered" in current action plans is an additional challenge to the method of "storing" such correlations.

As BW1 has not been in existence very long, these memory structures have so far been implemented in a rather rudimentary way.

6.4 Language

Apart from its expressive material (sounds, signs) a natural language also consists of what the verbal expressions "signify", what they "say" and the effects they "produce" when they are used. These aspects of verbal expressions that go beyond the expressive material will, for our purpose, be called "meaning" by way of simplification.

For us human beings, meaning in language is insepar-

dungen zwischen Sprecher und potentiellem Hörer erarbeitet sowie verschiedene Beziehungen, die aktuell zwischen dem Sprecher, seiner Umgebung und dem Hörer bestehen. Als Ergebnis gibt es Vorschläge für verschiedene Sprechakte. Nachdem ein Sprechakt ausgewählt worden ist, werden dann im Lexikon jene Objekte aktiviert, die in den Sprechakt eingehen sollen. Die Wortformen des Lexikons einschließlich diverser lexikalischer Kategorien verweisen dann auf die Grammatik. In dieser gibt es grammatische Regeln, die festlegen, wie Worte einer bestimmten lexikalischen Kategorie unter Berücksichtigung ihrer Objektindizierung und unter Berücksichtigung eines bestimmten Sprechaktes miteinander kombiniert werden können. Das Ergebnis ist dann eine konkrete Folge von Wortformen, die dann als konkrete Äußerung „zu hören" ist.

Das Bild 5 demonstriert den umgekehrten Fall, das Hören einer Äußerung. Dieser Fall stellt erheblich höhere Anforderungen als das Reden.

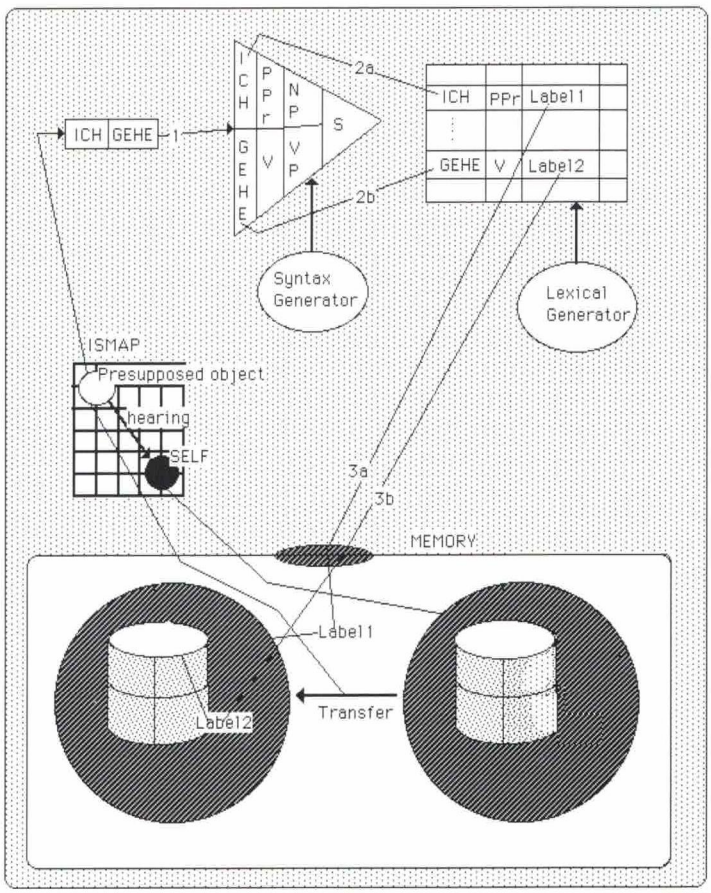

Bild 5

Ein vom Hörer verschiedener Sprecher sagt „Ich gehe". Die Regelschemata der Grammatik liefern zunächst eine erste Zerlegung der Wortfolge in einzelne Elemente, die über das Lexikon auch wieder auf entsprechende Objekte verweisen können. Mit Hilfe der LMAP versucht der Hörer gleichzeitig, den potentiellen Sprecher und die daraus sich ergebenden Relationen zu ihm selbst zu erfassen, um damit mögliche Sprechaktformen zu ermitteln. In diesem Zusammenhang muß er eine ziemlich komplizierte Transferleistung erbringen, insofern er nämlich das dem Wort „Ich" zugeordnet Ich-Objekt als Modell für ein anderes Objekt „ansehen" muß. Entsprechend sind die an der Äußerung „Ich gehe" zugänglichen Aspekte auf das Fremd-Ich-Objekt und dessen Beziehungen zur Umgebung zu übertragen.

Es sei darauf hingewiesen, daß der alltagssprachliche Verstehensbegriff erheblicher umfangreicher ist als der oben skizzierte sprachliche Dekodierungsvorgang während der grammatischen Analyse einer Äußerung. Im Alltag führt das Ergebnis der sprachlichen Erkennung in der Regel zu weiteren Erkenntnisprozessen wie z. B. Schlußfolgerungen, Ana-

able from the relation we have to ourselves and to the world.

For the knowbots in BW1 this relation to themselves and the world is only available to the extent to which it is provided by the reactive system capable of informed learning as described above. As will be shown below, this structure is still far too simple for the realization of discerning structures of meaning as we know them from human every-day language. No more than the most primitive naming and simple sentences consisting of one, two or three words are possible.

The language module is composed of several submodules.

The dictionary consisting of various sub-functions is an important basic unit. A word generator tries to form phoneme sequences to be stored in the dictionary from a set of elementary sounds and intonation patterns. The phoneme sequences will, however, only be kept in the dictionary if they obtain external "confirmation" within a certain period of time. A word-object generator forms

word-object hypotheses, i.e. it tries to link words with states or actions. The respective current situation has priority in the formation of such hypotheses. Word-object hypotheses, too, will only remain in the dictionary if they obtain confirmation from outside.

To be able to understand the grammar, one has to be aware of the tasks that have to be fulfilled in communication by means of language. Basically, BW1 differentiates between speaking and "listening". Ill. 4 demonstrates the procedure involved in speaking.

On the basis of an "interpreted sense map [ISMAP]" a pragmatic module called "language map [LMAP]" develops the differences between speaker and potential listener as well as various relations existing at that time between the speaker, its environment and the listener. The results are suggestions for various speech acts. Once a speech act has been selected, the objects to be included in the speech act are activated from the dictionary. The word forms in the dictionary, including various lexical categories, then refer to the grammar. The grammar has grammatical rules determining how words of a certain lexical category can be combined with one another while taking into consideration their object index and the speech act involved. The result is a concrete sequence of words that can be "heard" as a specific utterance.

Ill. 5 shows the reverse case, listening to an utterance, which is considerably more demanding than speaking. A speaker who is not identical with the listener says: "I go". The schematic grammatical rules first take the sequence of words apart; the individual elements are able to refer to the appropriate objects via the dictionary. With the help of the LMAP, the listener tries at the same time to comprehend the potential speaker and the resulting relations to itself so as to determine possible forms of speech acts. In this context, the listener has to perform a rather complicated transfer as it has to "regard" the "I-object" attributed to the word "I" as the model for a different object. In keeping with this, the accessible aspects of the utterance "I go" have to be transferred to the "other-I-object" and its relations to the environment.

It must be noted that the "notion of comprehension in every-day language" is much more complex than the process of linguistic decoding during the grammatical analysis of an utterance as described above. In every-day life the result of verbal recognition usually leads to further processes of cognition such as logical deductions, conclusions by analogies and graphic associations. Purely verbal comprehension has to be delineated as an - albeit necessary - sub-process of this general notion of comprehension. Due to the short time available, only the

logieschlüssen und bildhaften Assoziationen. Von diesem allgemeinen Verstehensbegriff ist das rein sprachliche Verstehen als ein – wenngleich notweniger – Teilprozeß abzugrenzen. In den Knowbots von BW1 konnte aufgrund der sehr knappen Zeit nur das sprachliche Verstehen realisiert werden und auch dies nur in ersten Ansätzen.

Dadurch, daß in BW1 alle Beteiligten blind sind, kann das Erlernen von Zusammenhängen zwischen Sprachzeichen und beliebigen Wahrnehmungsereignissen nur insoweit „von außen" beeinflußt werden, als es möglich ist, durch das Moment der Gleichzeitigkeit einen potentiellen Zusammenhang zwischen Teilen einer sprachlichen Äußerung und Teilen der Wahrnehmung bzw. Teilen des aktuellen Situationsmodells herzustellen. Durch den Ausfall des Sehsinns muß die mangelnde Eindeutigkeit dieses Vorgehens häufig durch ergänzende Tastwerte ausgeglichen werden, was nicht immer ganz einfach ist.

7 HARDWARE, SOFTWARE, MANPOWER

Die hohen Anforderungen an Interprozeß- und Interobjektkommunikation seitens der Installation BW1 konnte zum Zeitpunkt der Erstellung nur von dem Betriebssystem NEXTSTEP vollständig erfüllt werden. Ferner stellten die Systeme von Hewlett-Packard die einzige leistungsfähige Hardware dar, auf der im Entwicklungszeitraum NEXTSTEP portiert war. Wir entschieden uns daher, HP-Workstations mit dem Betriebssystem NEXTSTEP Version 3.2 als Entwicklungsplattform zu wählen: eine HP 712/80 und eine HP 712/60 mit jeweils 64 MB RAM, 1GB Platte und mit NEXTSTEP 3.2. Zusätzlich benutzten wir einen Aquarius-PC 486/66 mit 32 MB RAM und 1GB Platte mit NEXTSTEP 3.3. Für die Entwicklung der XWindow-Motif-Version des Normal-User-Clients stand ein ESCOM-PC 486/66 mit 16MB RAM 1.3GB Platte und dem Betriebssystem Unifix 1.5 inclusive Motif 2.0 zur Verfügung. Zusätzlich wurde eine Version des XClients unter IRIX auf Silicon-Graphics-Rechnern compiliert. Alle Rechner waren mittels EtherNet verbunden.

Die Server-Software sowie die NEXTSTEP-Version des Super-User-Clients wurden von Leo POS und Thore SWINDALL programmiert. Das Perzeptions- und das Memory-Modul im Knowbot erstellten Thore SWINDALL und Raoul SCHOLZ. Das Action-Modul sowie Teile des Language-Moduls programmierte Leo POS. Michael KLÖCKNER zeichnet verantwortlich für das Language-Module.

Die XWindow-Motif-Version des Normal-User-Clients wurde von Joachim RASCH und Sonja SCHELLENBERG im Rahmen ihrer Diplomarbeiten programmiert.

Die Vorlage für die Programmierung der Software unter NEXTSTEP bildet das Arbeitspapier Agents with Consciousness. The Theoretical framework for the Knowbotic-Interface Project. Phase I: The BLINDs WORLD I vom Verfasser. Ende

schen Bestimmung des Mensch-Maschine Verhältnisses. MS. Institut für Neue Medien e.V., erscheint in: Jahrbuch der Humboldt-Gesellschaft. 1995.

G. FREGE [1892]. Über Sinn und Bedeutung. in: Ztschr. f. Philos. u. philos. Kritik. NF 100. pp. 25 – 50. Abgedruckt in: G.PATZIG (Hrsg.). Funktion. Begriff und Bedeutungen. Göttingen. 1975. pp.40-65.
K. GÖDEL [1931]. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. in: Monatshefte f. Mathem. u. Physik. Vol. 38. pp. 173 – 198.

A. HODGES [1994 (engl.: 1983). 2.Aufl.] Alan Turing. Enigma. transl. R.HERKEN/ E.LACK. Springer-Verlag. Wien.

J. HOFFMANN [1982] Das aktive Gedächtnis.VEB Deutscher Verlag der Wissenschaften. Berlin.

E. HUSSERL [1913]. Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. Edited by E. STRÖKER. Felix Meiner Verlag. Hamburg.

F .KLIX [1980. 5th ed.]. Information und Verhalten. VEB Deutscher Verlag der Wissenschaften. Berlin.

G. LUDWIG [1978]. Die Grundstrukturen einer physikalischen Theorie. Springer. Berlin – Heidelberg – New York.

E. MACH [1886. 6. korr. Aufl. 1911. 1985 repr. von 9. Aufl. 1922 Die Analyse der Empfindungen und das Verhältnis des Physischen zum Psychischen. Wiss. Buchgesellschaft. Darmstadt.

M. MERLEAU-PONTY [1945. dt. 1965]. Phänomenologie der Wahrnehmung. Übersetzt von R.Boehm. Walter de Gruyter. Berlin.

G. M. MURCH/ G. L. WOODWORTH [1978]. Wahrnehmung. Kohlhammer. Stuttgart – Berlin – Köln – Mainz.

Th. NAGEL [1986]. The View from Nowhere. Oxford University Press. New York. Oxford.

W. SHIFF [1980]. Perception: An Applied Approach. Houghton Mifflin Company. Boston – Dallas – London et al.

A. M. TURING [1936-7]. On Computable Numbers with an Application to the Entscheidungsproblem. In: Proc. London Math. Soc., Ser. 2. vol. 42. pp. 230-265; corr. vol. 43. pp. 544-546 (Reprint in M. DAVIS 1965. pp. 116–151; corr. ibid. pp. 151–154).

– [1987 (engl.: 1948)]. Intelligente Maschinen. in: Intelligent Service – Schriften. Hrsg. v. B. DOTZLER u. F. KITTLER. pp. 81–113. (Engl. Titel: Intelligent Machinery).

– 1987 (engl.: 1950)]. Rechenmaschinen und Intelligenz. in: Intelligent Service – Schriften. Hrsg. v. B. DOTZLER u. F. KITTLER). pp. 147–280 (Engl. Titel: Computing machinery and intelligence. in: Mind. vol. 59. pp. 433–460).

– [1987]. Intelligence Service. Schriften. Hrsg. v. B.DOTZLER u. F. KITTLER. Verlag Brinkmann & Bose. Berlin.

L. WITTGENSTEIN [1921. dt. 9. Aufl. 1973]. Tractatus logico-philosophicus. Logisch-Philosophische Abhandlung. Suhrkamp. Frankfurt.