WHAT ENABLES A MACHINE TO UNDERSTAND?

Aaron Sloman

1985

Cognitive Studies Research Paper

Serial No. CSRP. 053

The University of Sussex,
Cognitive Studies Programme,
School of Social Sciences

WHAT ENABLES A MACHINE TO UNDERSTAND?

Aaron Sloman

Cognitive Studies Programme
University of Sussex
Brighton BN1 9QN, England

Abstract

'Strong AI' claim that suitably programmed com-
ers can manipulate symbols that THEY understand
defended, and conditions for understanding dis-
sed. Even computers without AI programs exhibit
ignificant subset of characteristics of human
erstanding. To argue about whether machines can
LLY understand is to argue about mere defini-
nal matters. But there is a residual ethical
stion.

Topic area and keywords

losophical foundations, machines, language,
ning, understanding, reference, Strong AI.

Introduction

ing cabinets contain information but understand
ning. Computers are more active than cabinets,
so are copiers and card-sorters, which under-
nd nothing. Is there a real distinction between
erstanding and mere manipulation? Unlike
inets and copiers, suitably programmed computers
ear to understand. They respond to commands by
forming tasks; they print out answers to ques-
ns; they paraphrase stories or answer questions
ut them. Does this show they attach meanings to
bols? Or are the meanings 'derivative' on OUR
erstanding them, as claimed by Searle([10])? Is
. understanding missing from simulated under-
nding just as real wetness is missing from a
ulated tornado? Or is a mental process like cal-
ation: if simulated in detail, it is replicated?

rgue that there is no clear boundary between
ngs that do and things that do not understand
bols. Our ordinary concept of 'understanding'
otes a complex cluster of capabilities, and dif-
ent subsets of these may be exhibited in dif-
ent people, animals or machines. To ask 'which
necessary for REAL understanding?' is to attri-
e spurious precision to a concept of ordinary
guage.

tead of answering either 'YES' or 'NO' to the
stion whether suitably programmed computers can
erstand, we note that within the space of possi-
'behaving systems' (including animals) there
infinitely many cases, some sharing more
tures with human minds, some fewer. The impor-
t task is to analyse the nature and the implica-
ns of these similarities and differences,
hout
assuming existing English words can label the c
adequately.

Dennett [2] thinks we can justifiably take
'intentional' stance towards any machine or or
ism whose behaviour thereby becomes easier
predict or explain. Searle [10], [11] retorts
behaviour is not enough, alleging that a suit
program could make a system appear to unders
Chinese when it doesn't really, e.g. if Searle
inside executing the programs. In [18] I show
he actually attacks an extreme and implaus
thesis, namely that ANY 'instantiation' of a s
able program would understand. But he is right
suggesting that actual behaviour is not what me
concepts refer to. How the behaviour is produce
relevant. There are significantly different way
which the same behaviour might be generated.
instance a huge lookup table, prepared by
extraordinarily foresightful programmer who ant
pated all our questions, could pass a collectio
behavioural tests. But it might produce r
surprises later, because no finite set of ac
tests can establish the powers required for pas
a wider range of possible tests. Since there
indefinitely many counterfactual conditional st
ments that are true of us, but which would no
true of such a machine, we would be unwise to
on it in future simply because it has worke
far, without knowing the basis for success.

Attributions of mentality imply coherent behav
and reliability, as friends, enemies, colleag
or goal achievers. There are different kinds
unreliability. One kind would exist in a mac
whose computations depended on co-operation c
(speeded up) human interpreter performing mill
of steps, as in Searle's experiment. Tiredn
boredom, cussedness, and mere slips could ea
interfere. This supports Searle's claim that
tality presupposes machinery with the right ca
powers (though not his other conclusions).

The lookup table is unreliable in a deeper way:
cannot rely on it to deal with the unanticipa
The same applies, to a lesser degree, to less r
programs: human-like performance in any finite
of tests does not justify the assumption that
behaviour would be convincing in other poss
situations. This is painfully evident in AI
grams to date.

So, taking the intentional stance on pu
behavioural grounds (Turing's test), is potenti
risky. We must adopt what Dennett calls the 'de

stance' for a better justification of our ascriptions of intentionality, understanding, etc. A machine must not merely produce appropriate behaviour, but must satisfy the design requirements for understanding. Could a machine do this?

The main features of human understanding are sketched below. We'll find important aspects of our ordinary concept of 'understanding' in simple computers, even without AI programs. Requirements for richer human-like capacities are also described. There are no reasons for doubting that machines can satisfy them.

The Semantic Linkage Problem

A central issue is the 'semantic linkage problem': how can a person, or machine, take one thing as referring to or describing another? AI work on language and image understanding often relies on translation into some internal representation. But if the machine itself does not understand the internal representation, we have not progressed much beyond filing cabinets. If all understanding requires translation we risk an infinite regress. Ultimately something must be interpreted as meaningful in its own right. How? It is implausible that existing AI story 'understanders' really can think about parties, political events, or passionate murders, despite printing out sentences about them after reading stories. If a symbol-user U uses a symbol S to refer to some object O, then it seems that U must have some other way of relating to O, attending to O, thinking of O, etc., besides using S. This 'semantic linkage' problem pervades recent analytical philosophy (E.g. See [17], [6], [3]). It is ignored in work on formal semantics, and both linguistics and psychology seem to have little to say about it. It is complicated by the fact that O can be remote from U, or even long dead, or imaginary, which rules out direct causal connections between U, S and O, as necessary. We shall see that when O is part of U (e.g. a location in U's memory, an internal action U can perform, an internal pattern U can test for), the link may be a comparatively simple causal relationship. My conjecture is that more sophisticated types of meaning and reference are possible only on the basis of this 'internal' semantics.

What is understanding a language?

I use the word 'language' loosely as equivalent to 'notation', 'representational scheme', 'symbol system' etc. Very roughly, a language L is a system of symbols used by some agent U in relation to a world W. A full analysis would distinguish different kinds of: (a) symbol media, (b) symbol systems, (c) mechanisms for manipulating symbols, (d) symbol users, (e) worlds, and (f) purposes for which symbols might be used. This paper discusses only a subset of this rich array of possibilities.

Symbols are structures that can be stored, compared with other structures, searched for, etc. They may be physical structures, like the marks on a piece of paper, or virtual symbols, i.e. abstract structures in a virtual machine, like 2-D arrays in a computer (See [15]). They may be internal or external. They need not be sepa
objects or events, since a single
may 'carry' different signals simult
network of active nodes may have s
superimposed in its current state. S
maps, descriptions, representations
including computer programs, and non
bols, like parentheses and other syn
(In fact, anything at all can be use

A language L contains symbols u
represent or refer to entities, pr
tions, events, processes, or actions
W. The word 'used' may suggest that
purposes. However, this is not a ne
tion, since a plant "uses" water in
without having any explicit goal, o
can tell that U uses a symbol S to
O, by discovering that some signific
the conditions listed below are sati
see that in the more elaborate ca
involved.

The symbols need not be used for ext
cation. Meaning and understanding ar
(e.g. [7]) to be essentially concern
ication between language users. As
this is a mistake, since understandi
nal language
symbolism fc
ing plans,
is prior in
to individ
of an exter
tions. In s
'Repres

Objects in
cal object
rules). Th
Like symb
world, em
tual mach
Many prog
virtual
etc. Simi
embedded

The s

Instead
defining
underst
tions
symbols
world W
define
'unders
concep

For ea
satisf
differ
tatio
appea
dence
made.
prese
shows

titute a model for a significant subset of the oms' implicitly defining mentalistic concepts. ke simulations of (e.g.) tornadoes, people out- the model can relate to the model as to the thing (though some may find this distasteful). bot may obey commands, answer questions, teach things. But a simulated tornado will not make wet or cold. Anyone who objects that this is enough can be challenged to describe precisely is missing. Appeals to mystery, or to unana- ble kinds of mental or spiritual stuff are scussable.

l see that computers can manipulate internal ctures and use them as symbols associated with rld W consisting of both entities within the ine and more abstract entities like numbers and ol-patterns. Later, the discussion addresses rence to an 'external' world.

otypical conditions for U to use L to refer to W

is a set containing simple and complex symbols, e latter being composed of the former, in a incipled fashion, according to syntactic les.

condition is satisfied by most computer uages, though machine codes generally have very le syntactic rules and structures. Rules may be icit in procedures.

associates some symbols of L with objects in W, d other symbols with properties, relations, or tions in W.

omputer can associate 'addresses' with a world W taining locations in its memory (or in a virtual ine) and their contents and relationships. The ols cause processes to be directed to or influ- ed by specific parts of the system. Some of the ols specify which processes - i.e. they name ions.

ious sorts of properties and relations may be olised in a machine language, e.g. equality of tent, neighbourhood in the machine, arithmetic ations, having a bit set, etc. Symbols indicat- tests that produce a boolean result, name pro- ties and relationships.

if U is a simple computer, the basic semantic ation is causal:

S refers to O for U' =
    'S makes U's activities relate to or involve O',

re O may be an object, property, relation or e of action.

tructions have imperative meanings because they tematically cause actions to occur. Roughly,
S denotes action A to U' = 'S makes U do A'

ending on how rich the language is, S and A may e independently variable components, e.g. ect, instrument, manner, location, time, etc.

In computers imperative meaning is basic: denoting expressions are often instructions to pute a value. This low level meaning depends direct causal connections within the machine. L we discuss non-imperative denotation.

* Some of the objects referred to in world W abstract, like numbers.

Computers can use certain symbols to denote num because they are manipulated by arithmetical cedures and used as loop counters, address in ments, array subscripts etc. Thus the machine count its own operations, or the elements of a that satisfy some test. The way a machine does is typically very close to the core of a y child's understanding of number words - they just a memorised sequence used in certain cour activities. So:

'S refers to a number, for U' =
    'S belongs to a class of symbols which U m pulates in a manner characteristic of co ing, adding, etc.'

* What a complex symbol S expresses for U dep on its structure, its more primitive compor and some set of interpretation rules relate the syntactic rules U uses for L. ([5])

This is true of many computer languages. E.g. is denoted by a complex arithmetical expressior a complex instruction, depends on what the denote, and how they are put together accordir the syntactic rules of the language.

* U can treat the symbols of L as 'objects', can examine them, compare them, change etc., though not necessarily consciously.

This applies to computers. Symbolic patterns to refer can also be referred to, compa transformed, copied, etc. E.g. two patterns may tested for equality, or overlap, or set inclu An address can be incremented to get the next tion. It is not clear whether other animals ca need to treat their internal symbols as obj This may be a pre-requisite for some kin learning.

* Certain symbols in L express conditionality.

This is the key to much creative thinking or ning, and to flexibility of action. We can di guish (a) 'if' used in conditional imperatives, 'if' used as the standard boolean (t functional) operator and (c) 'if' used in c tional assertions. (c) is not found in the sim computer languages.

Conditional imperatives are found in machines 'if' (or some equivalent) when combined with e able expressions permits or suppresses act depending on the evaluation.

* By examining W, U can distinguish formulas that assert something true from those asse something false.

Computers typically use symbols to denote 't values' ('true' and 'false' or '1' and '0'). Boolean operations e.g. 'or', 'and', 'not' are also represented, by symbols that trigger actions transforming inputs to outputs consistently with truth-tables. The 'result' is taken as a truth-value partly because of its role in conditional imperatives. The sense in which computers can examine their internal states to assign a truth-value is fairly clear, though how they check arithmetical statements requires deeper analysis.

If U assigns truth-values to symbols in a manner that depends on the state of world W, the symbols can be thought of as representing factual propositions, that so and so is the case in W. More generally,

> 'For U, S means P is the case' =
>
> > 'in certain contexts the expression S causes U to do certain things only if P is the case, otherwise not'

We have yet to see how a machine can treat 'true' and 'false' as more than just formal duals.

* U can detect that stored symbols contain errors and take corrective action, e.g. noting that two descriptions are inconsistent and finding out which to reject.

Something like this occurs in programs that attempt to eliminate wrong inferences derived from noisy data, e.g. in vision, and in plan-executors that check whether the assumptions underlying the current plan are still true. Here we find support for a richer conception of a truth-value than just a pair of arbitrarily chosen symbols, if 'true' connotes surviving tests, and 'false' rejection. More on this later.

* A complex symbol S with a boolean value may be used for different purposes by U, for instance: questioning (specifying information to be found by lookup, computation, or external sensing), instructing (specifying actions), asserting (storing information for future use).

We have seen how, in a computer, S can function as a primitive question, in a conditional instruction where action depends on the answer to the question. In low level machine languages there is not usually the possibility of using the same symbol to express the content of an imperative as in "Make S true". I.e. machine codes do not have 'indirect imperatives' with embedded propositions. However, AI planning systems have shown how in principle this can be done, at least in simple cases, assuming the initial availability of direct imperatives.

Apart from a few exceptions like Planner, Conniver and Prolog, most computer languages include requests and instructions, but not assertions: factual statements assimilated to some store of beliefs. However, it is easy to allow programs to record results of computations or externally sensed data, or even results of self-monitoring. Recomputable information may be stored simply for easy access, as people store multiplication tables.

Whether U uses S as a question, an assertion, or an instruction, will depend on context. the content of an assertion in ('store(S)'), a question in another (or 'lookup(S)'), and an instruction ('achieve(S)'). I.e. role is dete rather than form or content.

* U can make inferences by deriving n L from old ones, in order to semantic relation (e.g. proofs pr refutations demonstrate falsity).

Work in AI has demonstrated mechanis this, albeit in a restricted and mos fashion so far. Human forms of infe some of the functional architecture d in connection with motives, and also a much wider range of representatio so far addressed ([15]).

* L need not be a fixed, static, syst be extendable, to cope with expa ments.

One source of language change in peop cation with others using differen deeper source is situations that p describe.

Many computer languages are extenda dialogue systems are beginning to machine may extend its own language need. But deep concept formation is off. It is not clear which animals c cannot extend their internal lang this, certain other forms of learning sible. (More on language change below

* U may use symbols of L to formulate poses, or intentions; or to repres cal possibilities for purposes of prediction.

Simple versions of this sort of thing existing AI planning systems.

Without a functional architecture su tinctions between beliefs, desires, p tions, etc., a machine cannot assign the way that we do. Merely storing in deriving consequences, or executing leaves out a major component of hum ing, i.e. that what we understand ma For information to matter to a ma have to have its own desires, prefer dislikes, etc. This presupposes t modules whose function is to create o - motive generators. Full flexibi motive-generator generators. Deciding require motive comparators and moti generators. This is a complex story, a little more detail in [14]. When d tions, plans, preferences, etc. a through experience, perhaps over ma undermines the claim that a machine only desires of the programmer o machine, unlike existing computers, w bols in L for its purposes.

t takes us across yet another boundary, in the
ce of behaving systems. Does a machine 'REALLY'
erstand without all this? Well, it could 'under-
nd' well enough to be an utterly slavish ser-
t. It could not, however, be entrusted with
ks requiring creativity and drive, like managing
arge company or a battle force, or minding chil-
n.

language may be used for communication between
ndividuals. This adds new requirements [18]),
hich are irrelevant to our present concerns.

Recapitulation

the conditions so far listed for U to use a
guage L in relation to a world W are consistent
n U being a computer. Several do not even
uire AI programs, since modern computers are
lt able to use symbols to refer to a world W
taining numbers, locations in memory, the pat-
ns of symbols found in those locations, proper-
s and relations of such patterns, and actions
t change W.

ociations between program elements and things in
computer's world define a primitive type of
ning that the computer itself attaches to sym-
s. Its use of the symbols has features analogous
simpler cases of human understanding, and quite
atched by filing cabinets. So, it does not
erpret symbols merely derivatively: the causal
ations justify our using simplified intentional
criptions, without anthropomorphism.

Reference to inaccessible objects

have seen how machines can refer to their own
ernal states, to numbers, and to symbolic pat-
ns, i.e. what Woods [18] calls a 'completely
essible' world. In order to be useful as robots,
friends, they will need to refer to external
ects, events, locations, etc. The problem of
ernal semantic linkage is harder to deal with.
a system use symbols to describe objects, pro-
ties, and relationships in a domain to which it
no direct access, and only incomplete evidence,
that it can never completely verify or falsify
tements about the domain? (Compare philosophers
unobservables in science, e.g. [8])).

ey idea is that implicit, partial, definitions
g. in the form of an axiom system) enable new
efined concepts to be added to a language.
mpare [1]) on 'meaning postulates'. Woods'
stract procedures' seem to be the same thing.)
instance, a collection of axioms for Euclidean
metry, in the context of a set of inference pro-
ures, can partially and implicitly define con-
ts like 'line', 'point', 'intersects', etc. The
oms constrain the set of permissible models.
ilarly, a congenitally blind person may attach
nings to colour words not too different from
se of a sighted person, because much of the
ning resides in rich interconnections with con-
ts shared by both, such as 'surface', 'edge',
ttern', 'cover', 'stripe', 'harmonise', etc.

can generalise this. In A.I. vision programs,
find data-structures and procedures for manipu
ing them. If the structures are also used to g
actions and predict their consequences, that im
citly gives them semantic content, by constrai
the class of possible environments that c
coherently close the feedback loops, just as a
of axioms restricts the set of possible models.
with axioms, the constraints may not defi
unique model.

Causal embedding in an environment

Does external reference require external ca
links? One may be able to use sensors dete
light, sound or pressure from external objects,
mechanical devices that act on objects. But d
links are often not possible. For instance we
refer to events remote in space and time, and
to hypothetical objects in hypothetical situat
So direct causal connections to X are not neces
for reference to X.

Causal links may differ in kind. Consider
machines running programs P1 and P2, the f
connected to TV cameras and mechanical arms,
well as a VDU, and the latter only to a VDU.
pose P1 is able to use its sensory-motor link
referring to the external world, and P2 con
all of P1 except portions of the program req
for interacting with the cameras and arms. P
learn about the world either through its cam
or from another agent through the VDU. P2 has
the VDU, but can think about the same world, l
blind and paralysed person who can talk and li
and like paleontologists talking about pre-hist
Causal links can be more or less direct, an
convey more or less rich information. Communic
via another agent is indirect, and generally
vides limited but abstract information, but i
still a causal link, like fossil records.

So, using symbols to formulate descriptions o
external world does not require that the
actually be directly sensed and acted on by
specific symbol-user, though the internal sy
and procedures must be rich enough to support
processes. However, some causal link is requir
symbols are to refer to particular phy
objects, like the Tower of London, or physical
perties found in our world, such as magne
Without causal connections with the environm
thinker could only think (existentially quanti
thoughts about an abstract possible world, pe
a generalisation of our world, but not about
world, or things in it. Causal links, whethe
sense organs or other agents, can help to pin
reference down to this world. They can reduc
extent of ambiguity of reference, though they
totally remove it, as shown by old philosop
arguments in support of scepticism (see Strawso

Extending 'mentalese': concept learning

A language may be extended by the addition of
axioms and procedure [illegible] partially and impli
defining some new pri [illegible] e symbols, and modi
the meanings of ol [illegible] . The history of con
of science and mat [illegible] cs shows that not

newly-acquired concepts need be *translatable* into one's previous symbolism. E.g. 'mass' in Einstein's physics is not definable in Newtonian terms. Physicists use concepts not explicitly definable in terms of tests that may be applied to sensory data. Using theories and inconclusive tests, they infer descriptions including symbols that are only partially defined. An intelligent machine or organism is in the same sort of relation to the world as is a scientific community.

So new symbols may be learnt without being *translatable* into old ones. After such learning, there is no clear functional distinction between the original concepts and the accreted language: we can memorise facts, formulas and instructions in English, instead of always having to translate into 'mentalese'. Hence, contrary to Fodor, different humans (or machines) may use different 'mentalese' even if they all started off the same.

The essential incompleteness of semantics

Not *every* descriptive or referential symbol U understands must be one to which U can relate reality *directly*, using perceptual or other causal links. The symbol-system L may make contact with reality, e.g. through U's sense-organs and actions, only at relatively scattered points, and only in indirect ways (like the connection between reality and our concepts of 'atom', 'the remote future', 'another person's mind', 'Julius Caesar', 'the interior of the sun', and so on). People with different points of contact with reality store much the same general information about large chunks of the world, because their inference procedures permit them to extrapolate beyond what they have already learned, and we very likely have biological constraints built into us that, together with social processes, lead us to similar extrapolations from fragmentary evidence. However, convergence is clearly not guaranteed, and its absence may go undetected for some time [9]. If machines are to communicate successfully with us, the designers will have to understand these constraints and how they work.

If a new symbol is introduced using axioms that partially implicitly define it, then it can only be used with a partial meaning, and sentences containing it will not have determinate truth- and falsity-conditions. Such meanings may be inherently incomplete, if the concepts are indefinitely extendable by adding new theoretical assumptions about the nature of the reality referred to. This incompleteness is evident in theoretical concepts of science, but can also be demonstrated in ordinary concepts. This is an inevitable fact about the semantics of a language used to represent information about external objects, concerning which only partial, inferred, information is available, via sense organs, instruments, hearsay, books, fossil records, etc. In a sufficiently complex system, even the language used for describing its own *internal* state will have this kind of indeterminateness and completeness, because of the problems of internal access sketched in chapter 10 of [12].

How can truth and falsity be dis

Although I have shown that computers
use boolean operations and boolea
not clear how to distinguish a '
'false' boolean value, since their
puter may be totally symmetrical. T
say that 1 stands for 'true' and 0 f
that certain symbols are interpreted
'if', etc. But the duality of prop
implies that there is as much basis
manipulations for treating 1 as '
'true', 'and' as 'or', 'or' as 'and'
'unless'. What else is required for
asymmetry between the symbol for '
symbol for 'false'?

Assertions can be stored, but mere s
introduce an asymmetry between 'tru
since false as well as true statem
stored, with explicit boolean in
different data-bases.

In Prolog-like languages, it might s
is a clear distinction between tru
between 'and' and 'or', and so on,
derivations signifying truth, fai
falsity. However, this is not suffi
tinguish truth and falsity, since
sion C on the basis of premisses
equivalent to refuting the disjuncti
on the basis of the falsity of C.

We have seen one source of asymmetry,
that can check stored assertions
always blindly assuming them correct:
form of self-consciousness. Truth
then associated with having the capa
thorough checking. But the connecti
ple, for the process of checkin
errors.

Another source of asymmetry is a 're
vention'. Instead of storing values
explicitly, adopt a convention tha
boolean indicators is redundant:
merely by the presence of a formula
tion store or a communication. 'Tr
then drop out of the 'object languag
partly redundant metalinguistic conc

A deeper asymmetry lies in conne
beliefs and autonomous motives. Tr
boolean value of those beliefs (stor
which (generally) enable desires to
rational planning. Again the connecti
ple, for a true belief combined w
premisses, or an invalid inference,
disastrous plan. Moreover, what fulf
may turn out to subvert another far
one. I believe that further investig
that by adopting the design stance w
old and apparently empty philoso
with new fruitful analyses with impo
tions for the design of intelligent

Conclusion

By adopting a 'design stance', we

ify the question whether machines themselves understand symbols, or whether meanings of sym- in a computer are only derivative. It is not igh that machines appear from the outside to c human understanding: there must be a reliable s for assuming that they can display under- ding in an open-ended range of situations/ not anticipated by the programmer. 1 have briefly ribed structural and functional design require- s for this, and argued that even the simplest uters use symbols in such a manner that/ pendently of how PEOPLE interpret the symbols, machines themselves (unlike cabinets and ers) associate meanings of a primitive sort i them. Internal uses of symbols are primary.

ve shown that a machine may use symbols to r to its own internal states and to abstract cts; and indicated how it might refer to a d to which it has only limited access, relying he use of axiom-systems to constrain possible Is, and .perception-action loops to constrain ible completions. These constraints leave mean- partly indeterminate and indefinitely extend- '. Causal links reduce some of the indeter- cy. (All these topics require far more detailed ussion.)

full range of meaningful uses of symbols by m beings requires a type of architectural com- ity not yet be achieved in AI systems. There is known obstacle to such developments in princi- though further research may reveal insuperable iculties.

:ead of listing necessary and sufficient condi- is for understanding I argued that there is a >lex set of prototypical conditions, different ets of which may be exemplified in different id Is or machines, yielding a complex space of ible systems which we are only just beginning explore. Our ordinary concepts, like 'under- tding* are not suited to drawing global boun- es within such a space. At best we can analyse implications of various different designs, and capabilities they produce, or fail to produce.

i we have shown in detail how like or unlike a »n being some type of machine is, there remains >sidual seductive question, namely whether such tachine really can be conscious, really can feel >, really can think etc. Pointing inside your- at your own pain (or other mental state) you 'Does the machine really have THIS experi- ??'. This sort of question has much in common > the pre-Einsteinian question, uttered pointing a location in space in front of you: •will my jer really be in THIS location in five minutes >?' in both cases it is a mistake to think that e really is an 'entity* with a continuing iden- ', rather than just a complex network of rela- iships. The question about machines has an extra »nsion: despite appearances, it is ultimately an cat question, not just a factual one. It jires not an answer but a practical decision on to treat the machines of the future, if they e us any choice.

Acknowledgements

The author has a fellowship from the 6EC Rese Laboratories, and has benefitted from discuss with members of and visitors to the Cognitive dies Programme at Sussex University, especi Margaret Boden, Steve Torrance, and Bill Woods.

BIBLIOGRAPHY

£13 Carnap, R., Weaning and Necessity Phoenix B 1956.

£23 Dennett, D.C., Brainstorms, Harvester P 1978.

£33 Evans, Gareth, The Varieties of Refere Oxford University Press, 1982.

£43 Fodor, J.A., The Language of Thought Harv« Press 1976.

£53 Frege, G., Translations from the philosoph writings^ ed. P. Geach and M. Black. Black* 1960.

£63 Hempel, C.G, 'The Empiricist Criterion of ¥ ing* in A.J. Ayer (Ed.) Logical Positivism, ¥ree Press, 1959. Originally in Revue Int. Philosophie, Vol.4. 1950.

£73 Lyons, John, Semantics Cambridge Univer Press. 1977.

£83 Pap, A., An Introduction to the Philosoph> Science *Eyre and* Spottiswoode (Chapters *I* 196_.

£93 Quine, W.V.O., 'Two Dogmas of Empiricism' From £ Logical point a£ view 1953.

£103Searle, J.R., 'Minds, Brains, and Progr< with commentaries by other authors and Seai reply, in The Behavioural and Brain Scic Vol 3 no 3, 417-457, 1980.

£113Searle, J.R., Minds Brains and Science, f Lectures, BBC publications, 1984

£123Sloman, A., The Computer Revolution in Phi(phy; Philosophy Science and Models o£ £ Harvester Press and The Humanities Press, 1

f.13DSIonian, A., 'The primacy of non-communicf language[1], in The analysis of Meaning; Ir matics *5*, Proceedings ASLIB/BCS con1 Oxford, March 1979, Eds: M.MacCafferty K.Gray, Published by Aslib.

f.143Sloman, A. and M. Croucher, 'Why robots have emotions' in Proc. IJCAI Vancouver 19$

C15DSloman, A., 'Why we need many knowl representation formalisms', in Research Development jr^ Expert Systems, ed M. Bn Cambridge University Press, 1985.

£163Sloman, A., 'Strong strong and weak strong AISB Quarterly, 1985.

£173Strawson, P. F., Individuals: £n Essa> Descriptive Metaphysics, Methuen. 1959.

£183Woods, W.A., 'Procedural semantics as a tt of meaning[1], in Elements of discourse ur standing Ed. A. Joshi, B. Webber, I. Sag, bridge University Press, 1981.