AI REDACTED

J Bi Li Chan

THE LIMITS OF ART INTELLIGENCE

By J Bi Li Chan

THE LIMITS OF ART INTELLIGENCE

By J. Bi Li Chan

The Limits of Artificial Intelligence

Article for 'Encyclopedia of Artificial Intelligence'

J.T. Schwartz
Robotics Activity
Courant Institute
New York University

1. Introduction

The question of what intrinsic limits constrain the artificial intelligence enterprise, which can be defined as the attempt to construct electronic systems exhibiting human or superhuman levels of capability in areas traditionally regarded as mental, has been debated within very wide limits. On one side one finds a substantial community of researchers who believe firmly that such systems will prove possible. Their common (but not universal) assumption is that the organic brain is in effect a complex electrochemical system operating in some (doubtless highly parallel) but essentially computer-like fashion, and hence gives direct proof of the realizability of intelligence by mechanism; *vide* Marvin Minsky's flat-footed 'The brain is a meat machine'. Opposing this view one finds the assertion that mental processes are essentially indecomposable, lie outside the narrow reach of scientific reductionism, and that their indecomposability sets fundamental limits to any attempt to duplicate intelligence by mechanism. From this point of view, e.g. as represented by the writings of Hubert Dreyfus, the history of artificial intelligence research to date, consisting always of very limited success in particular areas, followed immediately by failure to reach the broader goals at which these initial successes seem at first to hint, gives empirical proof of the presence of irreducible wholes fundamentally incapable of being comprehended, much less duplicated, by the narrowly technical procedures of artificial intelligence researchers.

This philosophical debate concerns the existence of *fundamental* limits to the artificial intelligence enterprise, which however is only one of several kinds of potentially significant limit that need to be considered. Even if no such fundamental limits existed, i.e. even if a hypothetical infinitely fast computing engine possessed of infinite amounts of memory could in principle duplicate all aspects of human mental capability, it would still remain necessary to ask just how much computation and data storage such duplication would require. Suppose, for example, that it could be shown that the minimum computational resource required to duplicate some human mental function is implausibly large, relative either to the extreme limits of physically realizable computation, or to the largest computers likely to be constructed over the next decades or centuries. In this case, construction of significant artificial intelligences would be blocked by inescapable practical limits, even if fundamental limits did not exist. Finally, even if no such *computational* factors proved to limit the possibility of artificial intelligence, one would still want to assess the existing state of the field and project the rate of progress likely to result from application of its present intellectual tools to the profound problems with which it must wrestle.

The next five sections of the present article develop points relevant to the three kinds of limits defined in the preceding paragraph. A final section discusses certain other concerns, implicit in the debate between the enthusiasts of artificial intelligence and their opponents, which may explain some of the vehemence which has crept into this debate.

2. The Question of Fundamental Limits to the Constructability of Artificial Intelligences

2.1. A Very Brief Comment on the Philosophical Issue

In his deservedly famous 1950 article, Alan Turing proposed to replace amorphous philosophical debate about whether machines could 'really' think by the more pragmatic question of whether they could imitate the behavior of thinking beings well enough to make the assumption that they are 'thinking' the most comfortable basis for continuing interaction with them. The practical force of Turing's argument seems overwhelming. If at some future time people find themselves surrounded by artificially produced beings capable of performing the same variety of daily tasks, physical and intellectual, that one would expect of a person, and in particular capable of conversing on an unrestricted variety of topics in entirely easy, flexible manner, artificial intelligence will have been attained. This is not to deny the possibility that humans in this situation may choose to regard themselves as a kind of nobility, distinguished in view of their long and imperfectly understood biological pedigree from more fully understood and easily repairable/replaceable creatures. Such an attitude can even find objective justification in the reflection that, as long as any significant aspects of human function remain incompletely understood, humanity incorporates a pool of capabilities, tested by long evolution, which deserves protection and cautious nurture proportional to its long history and mysterious potential; these strong points also apply to whales and snail-darters.

Nevertheless, in the real presence of robots exhibiting human levels of flexibility and capability, the question as to whether these beings 'really' thought or merely 'appeared to' think and feel would lose pragmatic force, though of course its ideological importance might grow, perhaps even greatly. It makes less sense for the present article to pursue this debate than to assess the probability that such a situation will really arise.

2.2. The Brain as a Biochemical Computer

As already noted, part of the confidence with which artificial intelligence researchers view the prospects of their field stems from the materialist assumption that 'mind' is simply a name for the information-processing activity of the brain, and that the brain is a physical entity which acts according to the laws of biochemistry in a manner uninfluenced by any irreducible 'soul' or other unitary, purely mental entity incapable of analysis into a causal sequence of elementary biochemical events. Compelling evidence for the equation of mental function with the physical activity of the brain is easily drawn from many branches of science, and in particular from experimental neurobiology. For example, discrete lesions at the rear of the cerebral cortex produce discrete blind spots (scotomas) in the visual field, which turns out to communicate in 1-1 continuous fashion with the family of sensory neurons comprising the retina of the eye. Similarly, stimulation of points on the upper central portions of the cortex (temporal motor area) will produce elementary twitching motions of particular muscles. Physical manipulation of nervous tissue can also generate and/or remove sensations having profound motivational significance, e.g. direct application of an excess of potassium to the cutaneous nerves causes sharp pain; conversely, application of Novocaine to an appropriate branch of the facial nerve blocks dental pain in particular areas, thus permitting dental manipulations which would be unbearably aversive were the nerves communicating this sensation of pain not 'turned off'. These elementary remarks, plus thousands of far more precise observations obtained by direct recording of the electrical activity of individual neurons, show that neuronal activity reflects external stimuli and behavior (even intended behavior before its overt expression) in detailed and quantitative fashion, at least for those sensory and motor systems for which such correlations can be expected *a priori* to be understood most easily.

As might also be expected, detailed understanding of the manner in which neuronal activity reflects and governs a living creature's interactions with its environment is most complete for the simplest animals, particularly those whose nervous systems consist of relatively few neurons which, being particularly large, are relatively easy to identify and examine individually. A typical but particularly well-studied example of this is the marine snail *Aplysia Californica*, whose nervous system consists of roughly 20,000 neurons divided into nine separate ganglia within which hundreds of individual cells have been specifically identified. Fairly detailed understanding of the patterns of neuronal activity and interconnection governing many of the most typical and vital reactions of this simple creature has been attained. For example, much is known about the manner in which its nervous system controls heartbeat, respiration, gill withdrawal reflex, release of ink in response to a sensed danger, feeding, reproduction, etc. Moreover, *Aplysia* is capable of certain rudimentary types of learning (including *sensitization*, which progressive[ly] increases reactions to certain stimuli, and *habituation*, which progressively reduces other reactions), and the biochemical bases for these forms of neuronal plasticity have been at least partly elucidated. Finally, the nervous activities controlling sensation and behavior in *Aplysia* have been shown to be inherent properties of the nervous system which persist even when this system is dissected out of the body of *Aplysia* and maintained artificially in a suitable nutrient bath, provided that the afferent signals expected along certain sensory nerves are supplied electrically after the sensory organs that would normall[illegible] a robot's computer brain detached from its body and running in an artificially simulated environment of input and proprioceptive signals is overwhelming.

Some may object to facile extrapolation from the reflexive behaviors of a simple 20,000-neuron creature to the vastly more sophisticated activity of the roughly hundred billion neurons of the human brain. Nevertheless, the (admittedly highly incomplete) biological evidence available thus far seems to favor just such an extrapolation. Living creatures of whatever complexity seem to share a common neuronal biochemistry, in much the same way that they share a common genetic code.

Thus what neurobiological evidence there is hints strongly that no difference of fundamental principle separates the brain from any other other form of computer, or can be expected to limit the range of possibilities which artificial intelligence research can legitimately explore.

2.3. Quantitative Estimates Concerning the Brain's Computing Power

Even the resolutely mechanistic conclusion drawn in the preceding subsection leaves open two possibilities, either of which could still rule out the possibility of attaining human-like levels of mental capability by artificial means. In the first place, the mass of computational activity performed each second by the living brain, and/or the mass of information available to the brain for use during these computations, might be so large as to make electronic duplication of the brain's activity implausible. Moreover, even were this not the case, the algorithms which regulate the computational activity of the brain might be so marvelously subtle as to frustrate their rediscovery by artificial intelligence researchers for a very long time. Our next task is to examine these possibilities.

The human brain consists of approximately 10^{11} neurons, though this estimate is uncertain to within a factor of 10. Neurons typically (though not invariably) communicate by transmitting discrete electrical spikes (action potentials) to a population of follower neurons. As far as is known, the precise amplitude and shape of such a spike, and the precise time of its arrival within an interval of 2 milliseconds or so, are physical details which the nervous system is not able to exploit. This allows one to model each spike

as a single information-carrying 'bit' which can be present or absent in a neuron's output stream. We can therefore regard a neuron as producing output information at rate of approximately 100 bits/second. This leads to an estimate of 10^{13} bits/second, give or take a factor of 100, for the internal 'bandwidth' of the brain.

The computational activity of individual neurons involves a considerable variety of mechanisms, still very imperfectly understood. Nevertheless, a considerable mass of experimental evidence supports the following general picture. Information is transmitted by a neuron to its follower neurons at inter-neuron junctions called *synapses*. A single neuron can have as many as 10,000 such synaptic inputs, though in some cases many fewer and in other cases as many as 100,000 inputs are known to converge on single neurons. Thus the total number of synapses in the brain can be estimated as 10^{15}, though this estimate is uncertain by a factor of roughly 100. Input signals transmitted to a neuron (generally chemically) across a synapse trigger a wide variety of reactions. A common effect, and one which seems certain to be of particular importance for the fastest computations performed by the brain, is modulation of the ionic conductivity of the affected neuron's membrane, which either raises the voltage of a portion of its interior (excitation) or lowers this voltage (inhibition). The affected neuron then combines the voltage changes generated by such synaptic effects (after attenuation in space and time in a manner determined by the detailed chemistry and geometry of the neuron and its synapses) and, if the resulting combined (e.g. summed) voltage exceeds a reaction threshold, the neuron generates an output spike, which is then transmitted to all its output synapses.

Other forms of synaptic input are known to have slower but longer-lasting biochemical effects than the ionic effects which probably support the bulk of the brain's information-transmitting activity. Stimulation of certain synapses can, for example, trigger enzymatic activities within a neuron which modify its biosynthetic activities in significant ways, e.g. by increasing or decreasing its susceptibility to subsequent fast excitatory or inhibitory stimuli acting ionically. Depending on the chemical effects involved, such synaptic modification of faster synaptic responses can exert an effect either for relatively short periods (e.g. 50 milliseconds), or for periods of several seconds, minutes, or days, perhaps even permanently. Other synaptically-triggered enzymatic reactions can [illegible] sequenced biochemical changes which, for example, enhance a neuron's subsequent electrical response for several tens of milliseconds but then inhibit its responses for a longer period, leading to complex patterned alternations of behavior. The varied single-neuron behaviors which can be engendered by the wide spectrum of enzymatic actions that have been demonstrated experimentally have been explored in simple animals such as *Aplysia*, some of whose neurons are known to have highly individualized patterns of continuing, periodic, or burst activity.

a wide range of

wildly unfair

history

memory

of

rough

times

memory of

wiring

memory

of art

performance

rotating memories

of
memories

the

fall of raw

power

the

rediscovery

of

innate
capabilities

temporal

neuronal

among

connections
'immigrant' and 'native'

origins

spatial patterns

geometrical

positions

at point *x*

the absence of
decay

delay

the blurring

the
spreading

the evolution of internal state

patterns of

decay

in abstract form

arbitrary

random

delicate

image-like

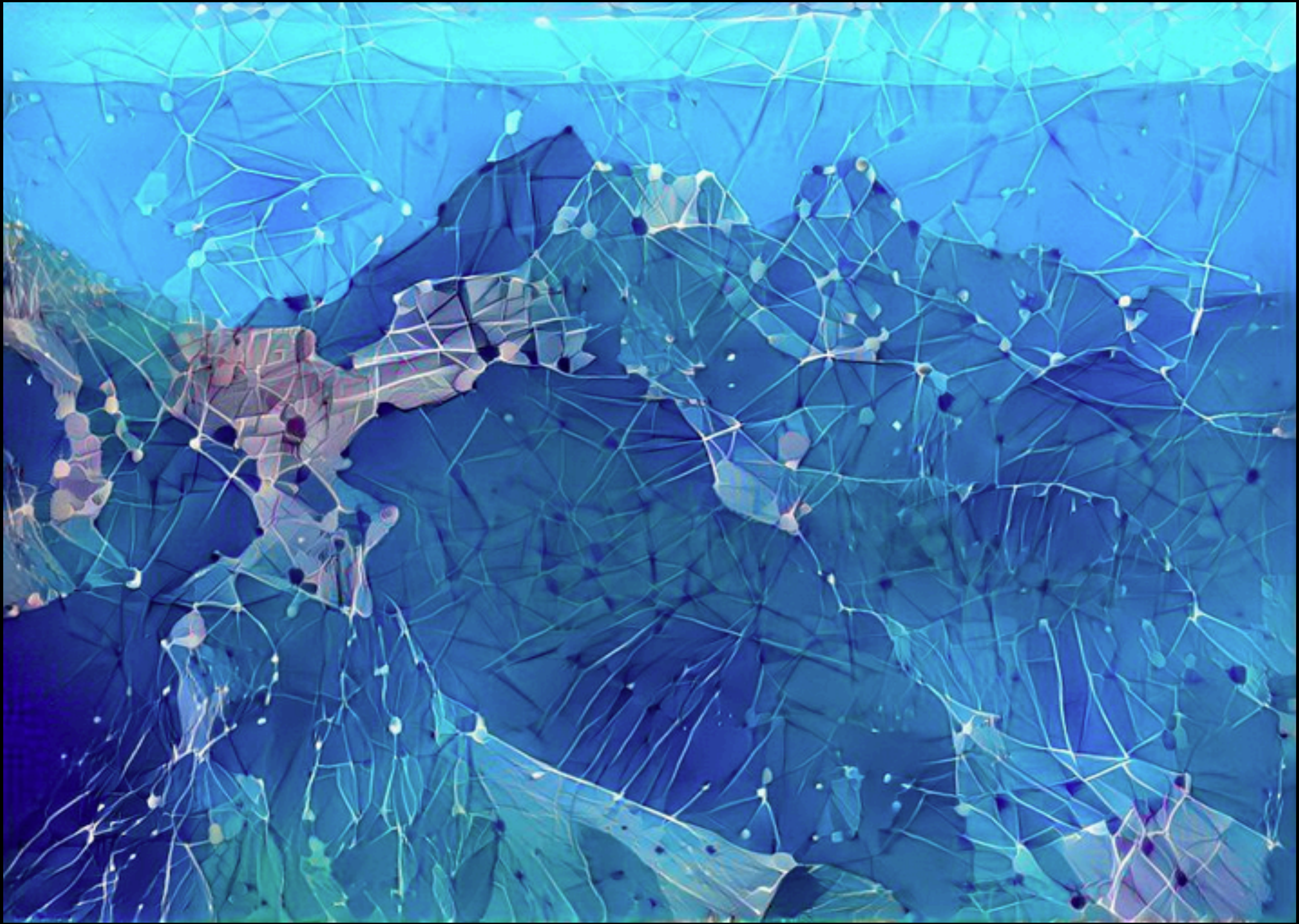
complex

evolution

by sudden leaps

artificial

over natural



sensory functions such as vision, tactile sensation, and hearing than to represent the brain's ability to deal with more discrete or symbolic material, i.e. to reason. The most remarkable, and perhaps fundamental, part of this is the brain's ability to organize information presented in relatively disordered form into internally organized structures on which sophisticated, coherent courses of symbolic and of real-world action can be based. It is the present lack of this ability that makes it necessary to *program* computers rather than simply to *teach* them; teaching would be vastly more convenient and which would bring the era of artificial intelligence very close if it became possible. To clarify this basic distinction, note that the ability of computers to accept, retain, and utilize *fully structured material* is already enormously superhuman, e.g. a computer can acquire and proceed to use the very complex set of rules for compiling a programming language in just a few seconds; nothing in the biological world other than the transmission of a full set of genes during conception matches this enormous rate of information transfer. On the other hand, although a computer can easily acquire and retain the whole text of the *Encyclopedia Brittanica* (even by reading its pages successively) computers are at present incapable of making any active use of the information which these volumes contain, since this text falls far short of the degree of rigorous order and standardization which present computers require. If this basic obstacle could be overcome, computers could immediately proceed to ingest the information contained in all the world's libraries and use this information with superhuman effectiveness. For this reason, a basic goal of artificial intelligence research has been the discovery of principles of self-organization robust enough to apply to a wide variety of information sources. Any such organizing principle would have to allow coherent structures capable of directly guiding some form of computer action to be generated automatically from relatively disorganized, fragmented input.

The present state of artificial intelligence research is most fundamentally characterized by the fact that no such robust principle of self organization is as yet known, even though many possibilities have been tried. Indeed, high hopes for the success of one or another apparently promising general principle of this type have characterized successive periods of research in the history of the subject. A typical attempt of this kind, particularly intriguing because of the great generality and potential power of the mathematical tools which it proposes to employ, has been the attempt to use formalisms drawn from symbolic logic as the basis for a self-organization capability. Mathematical axioms and theorems are mutually consistent fragments of information which can be accumulated separately and indefinitely; mathematical proofs based on these axioms and theorems are highly structured wholes which arise from these fragments according to the simple, well-understood principles of formal logic. If they could be generated automatically, these proofs, or various proof-like structures easily derivable from them, could be used almost immediately to produce many other

symbolic structures, including computer programs. Here a door to the most ambitious goals of artificial intelligence seems to swing open. Unfortunately, this prospect, like all others that have been explored to date, has proved to be blocked by fundamental considerations of computational efficiency, which we will now review.

The modern quantitative theory of computational infeasibility deriving from the work of Godel and Church allows one to prove rigorously that enormous computational costs will always make it impossible for programmed systems to answer certain general classes of questions in all cases. The original Church-Godel result is qualitative rather than quantitative, and can be summed up in a short *unsolvability* statement: there can exist no computer program P which is capable of examining every other program Q and determining correctly, in finite time, whether Q will run forever or halt eventually. Since many other combinatorial problems can easily be proved equivalent in difficulty to this basic unsolvable problem, they are just as unsolvable. Recent more quantitative work along these same lines has shown that there exist significant classes of mathematical problems which, although algorithmically solvable in the sense that one can write programs capable of solving each of the problems in such a class, are nevertheless *intractable*, since most of the problems in each of these classes carry minimal computational costs which rise with enormous rapidity as the program classes are progressively generalized in directions which eventually carry them over into the Church-Godel zone of complete unsolvability. As this happens, seemingly small loosenings of the constraints defining a particular class of problems always increase the cost of dealing with the generalized class enormously.

Problems in computational logic, whose efficient solution would provide very general and powerful tools for development of artificial intelligence, illustrate these general remarks. Any mathematical statement can be written in a convenient yet perfectly rigorous way using the simple notations of predicate logic. For example, the predicate statement

$$(\textit{FOR ALL } x, y, z, u, v, w) \tag{2}$$

$$(\textit{REAL}\ (x)\ \&\ \textit{REAL}\ (y)\ \&\ \textit{REAL}\ (z)\ \&\ \textit{REAL}\ (u)\ \&\ \textit{REAL}\ (v)\ \&\ \textit{REAL}\ (w)$$

$$\textit{implies}$$

$$((x+u)^2 + (y+v)^2 + (z+w)^2)^{1/2} \le (x^2 + y^2 + z^2)^{1/2} + (u^2 + v^2 + w^2)^{1/2})$$

captures the geometric fact that a broken line in three dimensional space is always at least as long as a straight line connecting the same endpoints. (In the preceding formula, clauses of the form *REAL* (x) express the fact that the variable x designates a real number.) Because of their great generality, predicate formalisms like that seen in the preceding formula provide very interesting testing grounds for artificial intelligence research. Any method which allowed the truth or falsity of large classes of formalized statements of this kind to be decided automatically and efficiently would also allow one to perform many other operations, including the automatic composition of many kinds of computer programs, the planning of grasping positions and motions for robot arms, and many other geometric and spatial analyses. However, a considerable body of rigorous theoretical analysis now rules out this possibility. Specifically, it has been shown that algorithms for deciding the truth of entirely general predicate statements cannot exist, nor can there exist algorithms capable of performing any entirely general process of formal reasoning, construction, or problem solving equivalent in difficulty to the task of classifying entirely general predicate statements as true or false. Indeed, the existence of such algorithms is directly ruled out by the basic Church-Godel theorem referenced above. On the other hand, algorithms capable of deciding narrower but still quite interesting subclasses of predicate statements do exist. For example, a famous theorem of Tarski asserts the existence of an algorithm capable of deciding any statement concerning real numbers which can be written using only the four elementary arithmetic operations of (addition, subtraction, multiplication, and division), comparisons between real numbers (e.g. clauses of the form 'x is greater than y'), the elementary Boolean connectives (and, or, implies, not), and the standard predicate quantifiers (*FOR ALL x, FOR SOME x*). However, the task which this algorithm accomplishes lies close enough to the Church-Godel zone of unsolvability that even apparently slight generalizations of this problem prove to be algorithmically unsolvable. For example, the same decision problem for the class of statements having exactly the same structure, but in which variables designate whole numbers (integers) rather than arbitrary real numbers (which for technical reasons are somewhat easier to deal with), is unsolvable.

Moreover, since the Tarski decision problem for real arithmetic is nearly unsolvable, any algorithm capable of deciding the truth/falsity of any statement of the form described must require enormous, and indeed prohibitive, computational resources in the worst case. Specifically, a theorem of Ferrante and Rackoff, proved in 1975 shows that the running time even of the fastest possible algorithm capable of deciding the truth or falsity of every statement s of Tarski form must rise exponentially with the length of s, for some (though not for all) such statements s. Thus in unfavorable cases the minimum running time of such algorithms will be probably in excess of billions of years, making their existence a matter of theoretical interest rather than of practical significance. Theorems of this same sort apply to many other classes of mathematical statements having decision problems of roughly the same degree of inherent difficulty as the Tarski class, and imply even higher degrees of computational difficulty for more general statement classes. For example, although the full class of statements of Tarski form becomes undecidable if applied to integers rather

than real numbers, the subclass of statements involving only arithmetic addition, subtraction, and comparison operations (but no multiplications or divisions) remains decidable even if applied to integers. However, here again we lie close enough to the zone of absolute unsolvability for computational costs to rise prohibitively high. More specifically, a theorem of Fisher and Rabin (1974) shows that these costs must be just as large as the Tarski case costs described above.

These general statements of computational infeasibility play the same role in computer science generally and artificial intelligence particularly that the first and second laws of thermodynamics play in physics and engineering, i.e. they set limits to what it is reasonable to attempt. While they do not at all rule out the possibility of artificial intelligence, they do suggest that it cannot be attained by programming any unitary mechanism of complete generality from which all that is needed will follow by simple specialization. Instead, it may be necessary to develop a relatively large number of artificial systems which mimic particular types of reasoning and mental functions in cases specialized enough to admit of particularly efficient treatment, and by systems whose 'coverage', while broad enough to be very useful, is less comprehensive than is assumed by naive mathematical statements of the problems they address. The individual functions thereby produced would then have to be integrated into a software structure capable of a very advanced level of function, which hopefully would also assist substantially in its own further development. Painfully detailed manual development of very many separate subcomponents of a highly complex total system capable of exhibiting a high level of intelligent function will only be avoided if some relatively uniform principle allowing computers to learn in human-like fashion is somehow developed. At present we have no real inkling of how this might be done, though the preceding model of neural function suggests that it ought somehow to be possible. It is equally unknown whether this present incapacity is a consequence of grossly insufficient computing power, as some of the estimates made earlier in this article seem to suggest, or simply reflects the fact that we have not yet found those simple yet efficient mechanical learning techniques whose discovery will enable much more rapid advance.

5. Limitations of the Present State of Knowledge in Artificial Intelligence

Since principles of self-organization allowing generation of broadly useful symbolic structures from more disorganized and fragmentary input would be crucial to the progress of artificial intelligence, work aiming at the discovery of such principles has been much emphasized. Signs of progress in this direction have always generated particular excitement. Unfortunately, all such efforts to date have run aground on the computational cost difficulties outlined in the preceding section. This fundamental fact constrains the immediate perspectives of the field severely. Of course, the many intriguing techniques developed during twenty years of artificial intelligence research do not lack application; indeed, their applications can be expected to grow steadily in scope and number. However, in the absence of any unifying principle of self-organization, these applications must be seen as adaptations of diverse ideas, rather than as systematic accomplishments of a still mythical 'A.I. technology'. We are still at the point at which the success of such applications depends far more on clever special algorithms and code reflecting particular application content than on use of the still impoverished general-purpose **tools of art**ificial intelligence. Moreover, since specialization is still generally vital to success, it is hard to characterize the extent to which success in any one application should be read as representing advance of the artificial intelligence field as a whole; to the degree that an application comes to **depend on** special techniques, special data structures, and special algorithmic approaches, we can no longer rightly regard it as evidence for the viability of a general approach distinguishable from artful programming in general. Nevertheless, some of the more specialized research efforts inspired by general artificial-intelligence notions have succeeded modestly in **mimicking** limited but interesting aspects of **mental** capabilities such as vision and natural language understanding.

To clarify this assessment, the present status of work along various significant lines will be summarized in this section. It is useful to arrange this work under three main headings: sensory functions, motor control, and reasoning. More detailed articles on the various areas reviewed should also be consulted.

5.1. Sensory Functions

These include analysis of **images** (computer vision), analysis of natural language made available in written form, and of continuous speech.

5.1.1. Analysis of Images

In spite of a great deal of work on the first steps of image processing (e.g. deblurring, edge detection) we are still far from being able to duplicate **the eye's** remarkable **ability to** detect objects in the presence of large amounts of visual disguise. Nevertheless our ability to identify objects within scenes is steadily improving, particularly for scenes containing only objects whose geometry and coloration is known in advance. Even if large parts of the objects present are **obscure**d, such scenes can be handled more easily than entirely general images (e.g. images of outdoor scenes containing shrubbery.) This reflects the fact that the problem of identifying known bodies and determining their orientation (the 'model based' vision problem) is entirely objective; in contrast, the problem of imposing useful perceptual groupings on entirely general scenes is at least partly psychological, i.e. to solve this second problem we need to match the functions of the human visual system well enough for introspection to serve as an accurate guide to the way in

which a robot **vision** system will react to a scene.

Among the many methods which are becoming available for handling the easier 'model based' vision problem are: direct matching **of curves** having fixed geometric position on known object surfaces; use of projective invariants of object silhouettes; probing techniques applicable for objects known to be presented in one of a finite number of allowed positions (e.g. objects lying on a table-top or conveyor belt) or on which one or more characteristic features can be reliably located; geometric reasoning using features (such as **corners**, straight corners, straight edges, **circles**) which can be detected directly or by statistical (e.g. Hough transform) methods.

Another promising object recognition technique is computation of invariants of local shape (rotational invariants) for the **edges** of two-dimensional figures and for the 'ridges' (curves along which at least one of a surface's extrinsic curvatures is large) of 3-dimensional objects. Any sharp **color** or reflectivity boundaries present on the surfaces of (painted or otherwise marked) 3-dimensional objects can also be used. To the extent that it is possible to define invariants stable against the disturbing effects of observation like **noise**, changes in **illumination** level, viewing **angle**, specularity, etc., this technique can support recognition even of heavily obscured objects and allows use of hashing techniques which greatly reduce the cost of identifying objects selected from large vocabularies of potential candidates. Beyond this, sophisticated use of color and **texture** cues available on object surfaces may prove possible. Here, however, we come to the point at which the human (or mammalian) visual system displays a sophistication that researchers seem far from being able to match, even after several decades of determined effort. In some remarkable way, the eye is able to integrate the evidential weight of **fragmentary clues** and to make use not only of dotted and dashed lines but of computationally elusive texture boundaries, vague differences of shading, and curves which are very badly broken up by obscuring objects (e.g. foliage) and complex **shadow patterns**. All this can be done in a manner resistant to the confusing effects of very large changes in illumination pattern, intense specularities, image blurring, and the myriad other effects all too painfully familiar to the vision researcher. Finally, all this is possible for scenes containing large numbers of objects, some unfamiliar, seen in a great variety of apparent sizes, from sharp and severely distorting angles, and **in the absence of** binocular information.

At the present time we have little **understanding** of how all of this is accomplished, and at what computational cost.

However, it is clear that image processing tends to be very expensive computationally (e.g. initial analysis of an image often requires examination of between 250,000 and 1,000,000 separate image pixels), so that substantially faster processors than are now available may prove to assist the development of this very challenging subject. These processors may include

The ability to hear

birdsongs

the possibilities

of

lying somewhere

smooth

having a

voice

transition to a

space

one

belongs

the presence of primitive

symbols

sounds signal

sequence

ambiguous

disambiguation of

language

pure statistical

frequency

follows

phonemes

the)
filtering
of

words

attention on
power
the inherent difficulty of

interpretation

a certain optimism concerning progress

the accessibility of

natural
irregularities

an
overgrown jungle
of

ambiguities

far from

appearing

natural
art intelligence
imbeds

inconsistent

relationships

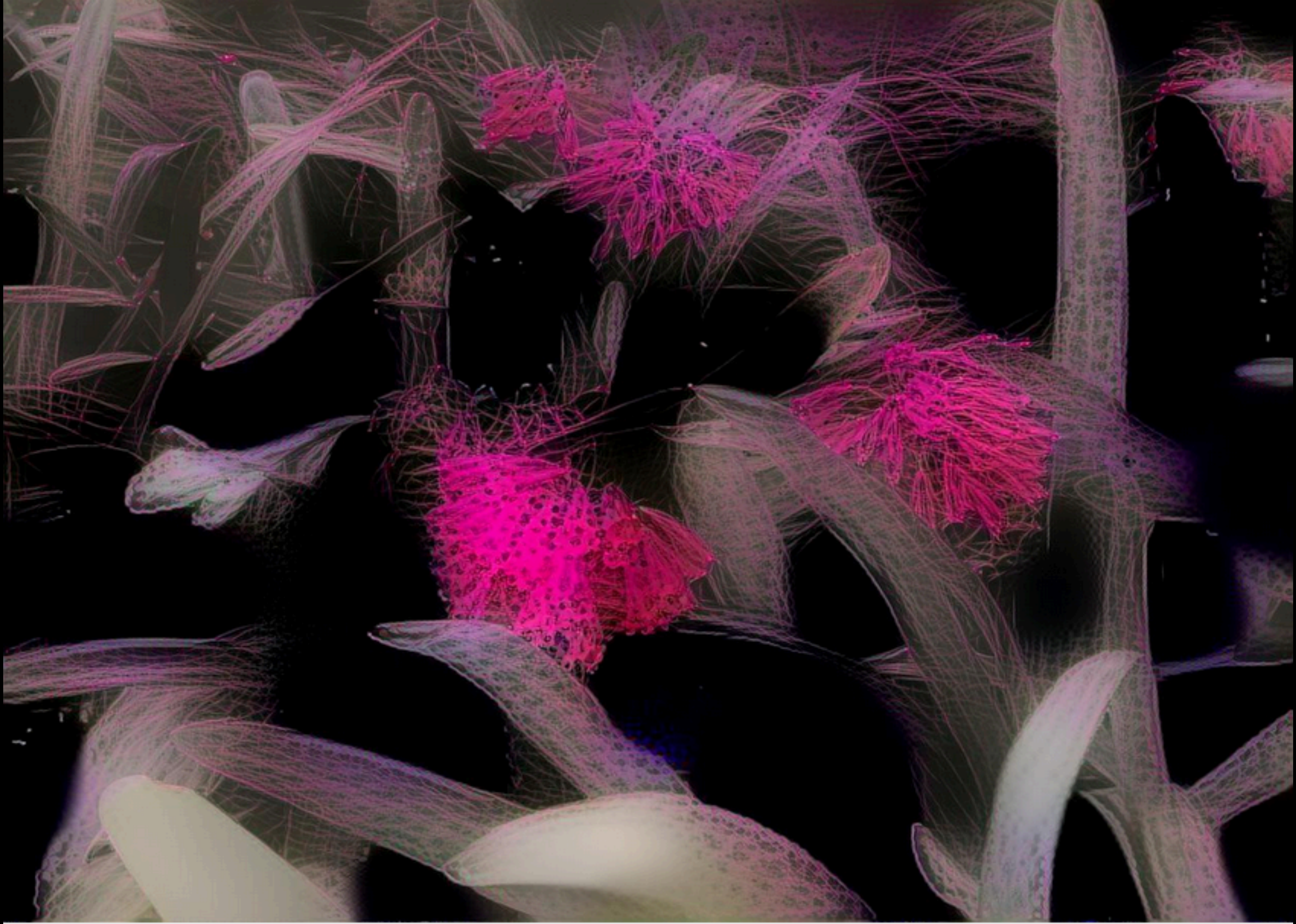
challenging

arguments

as fairy stories

common
sense

embarrassment



text streams.

At present we have little idea of how to treat most of these issues, which collectively reach to the heart of the artificial intelligence enterprise. For example, no 'probabilistic' or 'fuzzy' formalism beyond the well-defined but rigid semantic area mapped out by propositional and predicate logic has as yet demonstrated advantages sufficient to win it general acceptance. Moreover, the basic problem of what primitives a semantic formalism should use is surrounded by deep and ill-fathomed questions. One possibility is to somehow simplify the capture of information concerning the very many concepts appearing in natural language discourse by re-expressing them in terms of some much smaller family of simpler primitives whose properties can then be expressed by a significantly smaller set of rules. (This simplification would in effect require finding some way of extending the analytic reductionism characteristic of theoretical science to the entire range of phenomena which natural discourse addresses.) Any expectation that this can succeed easily is discouraged by consideration of the slow pace with which science has previously advanced into entirely new fields, and on the enormous computations sometimes required to apply general scientific laws to particular concrete cases. The opposite approach is to somehow build a semantic formalism which handles the very many terms appearing in natural language as unanalyzed primitives which it relates to each other by comprehensive sets of axiom-like formulae. Belief that this approach can succeed easily or rapidly is discouraged by the formidable difficulties of steering proofs in predicate calculus systems that try to deal with more than a dozen or so carefully crafted axioms.

Measured against these deeply rooted problems, existing techniques for dealing with natural language semantics appear sketchy indeed. *Semantic network* systems attempt to organize the enormous variety of objects and predicates appearing in ordinary discourse by representing them as nodes in graphs whose edges represent various logical relationships which are felt to be particularly fundamental to common elementary inferences. For example, such edges may connect nouns A and B whenever A is a 'kind of' B (e.g. when A is 'man' and 'B' is 'mammal') or when A is a 'part of' B (e.g. when A is 'arm' and B is 'man'.) A second aim of schemes of this sort is to accelerate simple semantic deductions by making the information they require directly available through short chains of pointers and by grouping related information needed for the commonest types of deduction under appropriate headings. The feasibility of attempts of this kind could only be demonstrated by exhibiting at least one readily extensible system able to cover some extensive domain of practical knowledge robustly, something which no one has yet done successfully.

Roger Shank's 1977 'conceptual dependency' scheme represents an attempt to reduce the myriad elements appearing in ordinary discourse to a

much smaller set of semantic subcategories. It is not inconceivable that such an attempt should yield some useful degree of systematization, even though a pessimist might might view it as a futile effort to enlarge the applicability of scientific modeling by casual invention of a classification scheme. The categories proposed by Shank include 'acts' (essentially verbs, which it is proposed to further subdivide as variants of purported primitive acts such as 'propel', 'ingest', 'expel', 'speak', etc.), 'picture producers' (essentially nouns), 'times', 'locations', etc. A related aim here is to classify all the inferences which attach to entities of these proposed semantic categories.

Marvin Minsky's 'frames' and the associated 'scripts' proposed by Shank define a more general (but accordingly more empty) framework for organizing common sense knowledge in a stereotyped form. Minsky proposes to classify all the logical entities (e.g. nouns) that can appear in a semantic network system into (a possibly large number of) fixed categories. With each such category, a Minsky 'frame' associates a fixed-format record layout listing all the attributes which an item of the given category might have, together with all the values or categories of values which each particular attribute can assume. For example, the frame for entities of category 'restaurant' might have a 'type' field with possible values 'cafeteria', 'full-service', 'full-service-with-hostess', etc., a 'food-style' field with possible values including 'Fish-and-chips', 'Mexican', 'Chinese', 'Thai', 'Seafood', and so forth. Categories can be defined to be specializations of more encompassing categories, whose attributes they inherit; certain of the attributes of a category can be optional.

Shank proposes to include records of another fundamental kind called 'scripts' in semantic systems. These are to be used to describe categories of activity (rather than of objects, as with 'frames'). Basically they list sequences of subactivities, which can in principle be conditional on specified conditions. 'Frames' and 'Scripts' are tied together by the fact that a script can specify the kinds of objects expected to appear in the activities it describes (by including pointers to the corresponding frames), while the frames describing an entity type can reference scripts describing the activities typically associated with these entities.

Taken *per se*, this mechanism is little more than a way of organizing some aspects of the data with which full-fledged semantic inference systems will have to deal, and does not answer the questions of how such an inference system is to be created any more than the inclusion of vaguely similar record types in programming languages such as Pascal and PL/1 answers the question of how to write complex compilers or symbolic manipulation systems using these languages. However, it can also be read as suggesting a semantic interpretation scheme having something of a 'higher level syntax' flavor. Specifically, Shank's 'scripts' can be viewed as higher level grammars defining a language of semantically plausible sentence sequences (whose rudimentary elements are clauses or other sentence fragments, already pre-parsed in some more standard syntactic sense). This 'grammar' of scripts would allow much 'nulling of' script elements, but then by using such a grammar to 'parse' a text and immediately 'unparsing' the result, with element nulling forbidden, one can hope to make explicit certain simple but very useful classes of normally implicit inferred elements. (Since grammars which allow large amounts of nulling tend to interpret given texts in highly ambiguous fashion, application of a scheme of the sort described may depend upon a rule which prefers the 'shortest' or 'simplest' semantic script-parse of a text to all others. Such a rule would amount to requiring that only those implicit elements necessary to a text's semantic interpretation could rightfully be inferred. Alternatively, the scripts driving the semantic interpretation process could associate probabilities with each elementary interpretation step, and some rule defining 'most probable' interpretations could be used.) A 'grammar of scripts' used in this way will necessarily be context dependent, since semantic connections would have to be maintained between elements (e.g. explicit or implicit nouns or pronouns) recognized at one point of a text and matching occurrences elsewhere. Hence 'parsing' according to such a grammar might come to resemble the very inefficient processes of computational logic much more than the relatively efficient processes of ordinary syntactic analysis.

It would however be easier to take such rationalizing suggestions seriously if straightforward formalisms had been proposed for use in this area and if some initial analysis of their computational cost were available. Unfortunately, however, the literature contains little but preliminary and often confusing heuristic suggestions and computational schemes set out without much justification, no one of which seems to have gained any general degree of acceptance.

This brief review of the difficulties which confront attempts to automate natural language understanding underscores the wisdom of Turing's 1950 suggestion that ability to conduct natural-seeming conversations should be regarded as a touchstone of progress in artificial intelligence. In spite of much work, even a computer able to read simple stories (e.g. ordinary children's stories or newspaper articles) and to answer simple questions about their content still lies far beyond us. Existing semantic analysis systems are fragile laboratory constructions which can deal only with narrowly restricted subject domains. The mechanisms thus far suggested as bases for more comprehensive semantic systems are all quite primitive. Since the problems with which they must deal seem to encompass almost the whole subject matter of artificial intelligence, only slow progress can be predicted.

5.2. Motor Control, Modeling of Spatial Environments, Motion Planning

Our review of these topics will illustrate the point that areas of artificial intelligence to which classical scientific and algorithmic techniques apply can be expected to progress more rapidly than areas which deal with deeper problems for which only less focused approaches are available. Many of the capabilities reviewed in this section are being explored in connection with industrial robotics. Since many of the problems encountered are technical rather than fundamental, it is reasonable to expect steady progress, at a rate largely determined by the resources brought to bear. However, it should be noted that work in this area creates very challenging problems of software systems integration, involves a complex mix of technologies, and is quite expensive. Studies in other areas of artificial intelligence such as computer vision may raise similar practical problems as they advance toward maturity.

Research in motor control aims to devise robots capable of exerting sophisticated hybrid force and positional control over grasped objects and to construct robots which can walk, run, leap, and climb. Typical problems of manipulation are to tie a knot in rope, to thread a nut of imprecisely known shape and pitch onto a bolt, and to pick up a jumbled sheet of cloth and fold it neatly. Techniques adapted from concepts presently belonging to nonlinear control theory (which should be considerably enriched by contact with robotics) should make sophisticated manipulation of rigid objects possible during the next few years. To do this, much work on such classical topics as the frictional and elastic reactions of bodies in contact will be required. Dynamic robot control, such as is involved in walking or running, should also progress steadily over the next few years. However, this will require close study of the complex physical situations created as motor-actuated mechanisms having various geometries and dynamic behaviors enter into repetitive contact with supporting surfaces.

The problems of dealing with nonrigid objects, e.g. cloth, are much less understood, and we lack even a vocabulary for describing some of the basic operations involved. How, for example, is a robot to find the edges of a hanging sheet of cloth preparatory to folding it? Roboticists have not yet begun to grapple seriously with such problems, and it is not now understood whether these will permit of uniform attacks or require development of special analyses and approaches in a large number of different cases.

With a few experimental exceptions, today's robots do not maintain any systematic internal model of their environment; the environment is typically known to them only as a source of tactile or visual interrupts, all sense of external object identity being lost as soon as a grasped object is set down or passes out of sight. To develop any deeper understanding of the environment, robots will require far more sophisticated environment-modeling software than is now available. Although the basic principles require for this are largely available from classical physics and geometry, it the paths

of art
are bound to

geometry and

physics
bodies

may be moving

in

this circle of
tricks

finding a path between
two

states

is

symbolic

transformations

can be astronomical

without the precise path being known

familiar
features
of

frustrated

search

for

desired

art

More

to conceal than to reveal

belief in the
disappeared art

whose edges are
a

path connecting

a thousand nodes

simple

Attempts to generate
systematic
specialized
techniques

have

been

negated

collections of

collection

in special

sets

have devised

decision procedures
for pure

power

even after

discovery of

power

I
have to
become capable of producing

the power of manipulation

to commercialize ideas drawn
from art

the expert

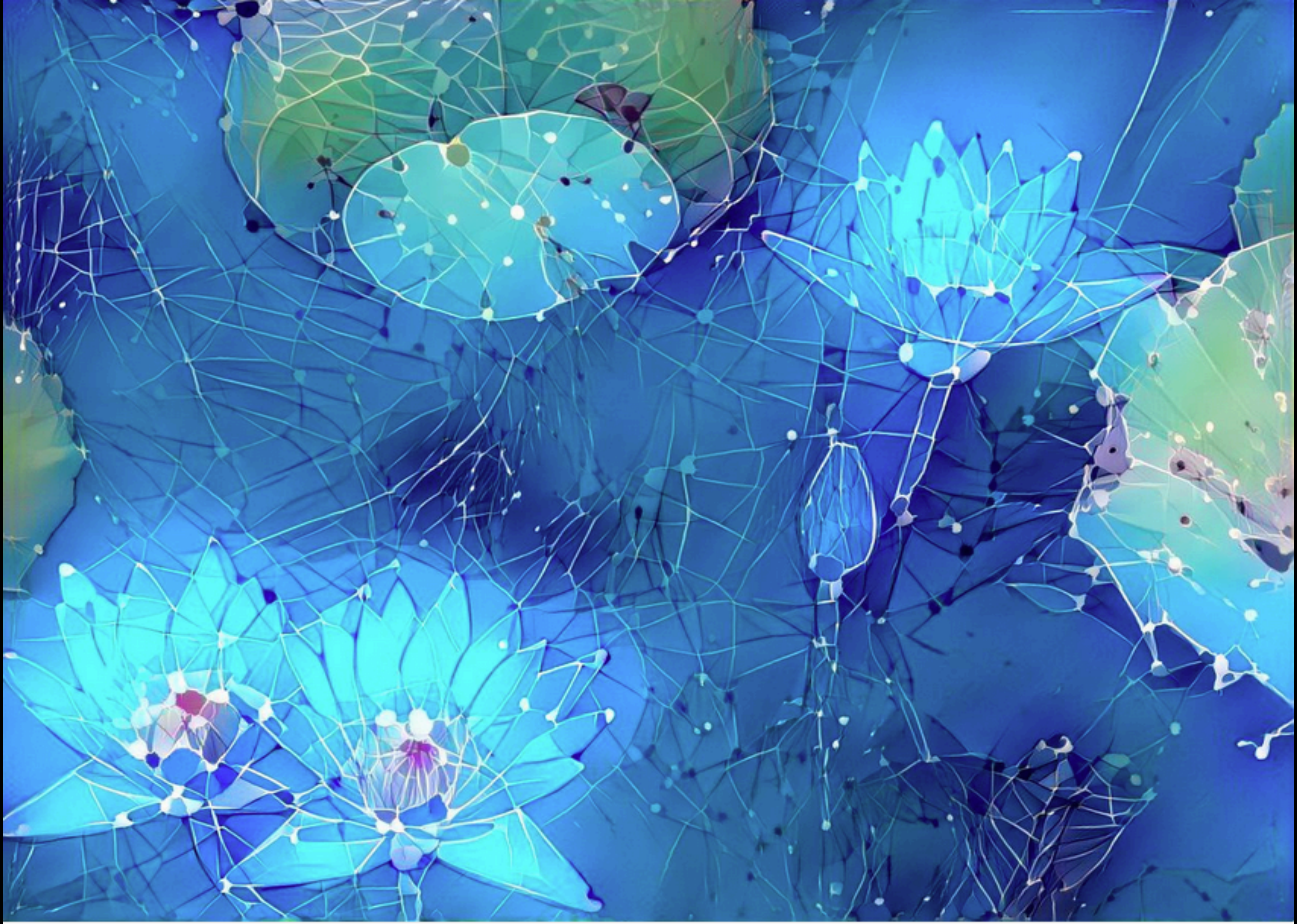
is
the agent of the
sophisticated

truth

an internal
engine
called upon to explain its

answers

'yes' or 'no'



(1) Expert systems can include attractive natural language and/or graphic interfaces.

(2) Instructions for carrying out any diagnostic procedures or tests required to answer queries of type (b) can be stored in such systems and made available when the system user is asked the corresponding questions. Specialized editors, databases, visual aids, and modeling systems relevant to a system's application domain can also be provided.

(3) Questions can be cleverly sequenced rather than simply being asked in fixed order. If evidence already supplied allows such a question to be answered either definitively or with high probability, or if it makes a question irrelevant to the type (a) final conclusions at which an expert system aims, the question can be suppressed.

(4) A system's user can be allowed to ask how particular final or intermediate conclusions were arrived at, in response to which the system can display its internal Boolean or probabilistic deduction steps, along with the built-in rules justifying these steps, in forms calculated to aid user comprehension.

(5) In some application areas, special deduction rules or other symbolic manipulations going beyond the merely propositional will be possible. For example, an expert system oriented toward chemical syntheses or analyses may be able to manipulate structural descriptions of molecules; an expert system dealing with electrocardiograms may be able to ingest raw cardiographic data and apply sophisticated spectral analysis or other pattern-matching procedures to it. The power of expert systems which include special techniques of this sort may rise substantially above the level attainable by primitive Boolean inference.

Overall, we can say that expert systems enhance their pragmatic applicability by narrowing the traditional goals of artificial intelligence research substantially, and by blurring the distinction between clever specialized programming and use of unifying principles of self-organization applicable across a wide variety of domains. This makes their significance for future development of deeper artificial intelligence technologies entirely debatable in spite of their hoped-for pragmatic utility.

5.3.4. Knowledge Representation

The phrase 'knowledge-based system' has become popular among scientists seeking to apply artificial intelligence research, and the associated dictum that 'finding appropriate representations of knowledge is one of the most basic problems of the artificial intelligence field' has often been propounded. Unfortunately, it is hard to identify any data structures created by the artificial intelligence research community that are other than superficial. Aside from clever internal implementations of such languages as LISP (which no one would consider 'knowledge representation' in any

specific sense), no structure more advanced than simple pointer networks seem to have been proposed. Of course such networks are quite familiar from many other applications as 'graphs' or simply 'mappings'. They involve nodes that are little different from the 'records' of standard data processing. This contrasts strongly with other branches of computer science, in which many quite ingenious data structures have been developed. In these fields, numerous successful examples have given the phrase 'data structure design' a mature technological meaning: any way of storing one or more abstract data entities in a manner which significantly accelerates the speed with which some well-defined battery of operations can be applied to these entities defines a significant data structure. Examples include B-trees, AVL-trees, Fibonacci heaps, compressed balanced trees, and many others. The underlying aim of artificial intelligence researchers in regard to 'knowledge representation' is of course the same as that of other computer scientists, namely to find data representations that can be used to accelerate the symbolic calculations that they would like to perform. However, progress toward this goal has stalled, since no acceptable formulation of the abstract structures to be implemented, or of the operations to be performed upon them, has yet become available. The one possible exception is use of 'semantic nets' for fast retrieval of items associated with other data items used as keys, a standard programming technique that artificial intelligence research actually has used in a manner no more sophisticated than is now common in database practice.

5.3.5. Learning

As stressed previously, one of the profoundest goals of artificial intelligence is to make computers capable of learning, i.e. capable of using disorganized information fragments to construct organized structures on which they can take action. Broad success with this one point would be almost equivalent to full realization of the subject's aspirations. Unfortunately, almost nothing has yet been accomplished toward this bold goal. The disappointments encountered are typified by the variety of schemes that have been tried for allowing a computer to acquire the grammar of simple formal languages by exposure to sets of grammatical strings belonging to such languages. Although various faintly encouraging theorems have been proved concerning the asymptotic convergence of various learning algorithms to a desired grammar given sufficiently large numbers of positive and negative sentence examples, the enormous number of candidate grammars that present themselves have frustrated all practical use of this scheme. (Related experiments include attempts to discover the simplest possible Boolean expression for a subset S of the set of all computer words of fixed length (whose bits can be thought of as representing true/false attributes of some class of objects or scenes). The input to such experiments are sets of positive and negative examples, or information concerning 'near misses' which can be given by stating the distance (measured in bits wrong) of each sample word from the nearest member of S. However, beyond various fragmentary heuristics, neither a practical approach to this problem nor any understanding of its inherent computational cost is available.

Other more trivial data-acquisition capabilities have been demonstrated and can be regarded as learning of a sort. For example, it is possible for a computer equipped with **an image** digitizer to acquire pictures of objects successively presented to it, then to calculate and store shape parameters for the boundaries of these objects, and subsequently to recognize the same objects when seen in other positions (at least, this **is** possible for favorable classes of objects). Perhaps this can be regarded as **a** rudimentary form of learning. Other techniques sometimes described as automatic learning involve use of data-derived statistics to adjust numerical parameters internal to a program. An even simpler **possibility** is to supply internal program constants progressively and interactively rather than all at program definition time. An example of this limited and artificial type of 'learning' would be **a** string analysis program, designed to be aware of the distinction of **single** characters (which it extracts internally from character data fed to it) into vowels and nonvowels, but not told initially which characters are which. Such a program can trivially emit an enquiry about each newly **encounter**ed character, following which the character can be inserted into one of two internally maintained sets, making subsequent enquiry unnecessary. The reader may or may not wish to regard this as true 'learning', since in much the same sense, one could view any menu-driven program **which elicits** and stores information concerning its user's preferences as a program which learns.

5.4. A Comment on Methodology

As might be expected of a young scientific discipline concerned with new, profound, and enormously attractive problems, the methodological level of research in artificial intelligence is often low. This contrasts with the situation in those other branches of computer science in which it has proved possible to define reasonably specific and feasible computational goals in a manner independent of the the techniques known at any given moment for trying to reach these goals. Where this has been possible, clear **challenges** have come before algorithm designers (who then often have found sophisticated and sometimes quite **unexpected** ways of computing important quantities) and computational complexity theorists (who seek to clarify the options open to the algorithm designer by proving theorems concerning the minimum computational cost of particular operations.) The systematic work flowing from this clarification of goals has substantially increased the maturity of other branches of computer science. Disappointingly, more primitive approaches have persisted in artificial intelligence research. Too many publications in this field simply describe the structure of some program believed by its authors to embody some function mimicking some aspect of

intelligence, but aside from this having no definition other than the particular procedures of which it consists. It is often impossible to determine just what such a program really computes, or whether it does so with **accept**able or catastrophic efficiency, or whether some other much more efficient technique might not have computed essentially the same thing. Still more **primitive** but nevertheless common publications consist of lightly or heavily edited **traces** of some program's internal activity, accompanied by author comments on felt similarities between this activity and the author's personal theory of mental function; a form of report which often leaves its reader without much **understand**ing of what the program described is really doing, or how, or with what **limitations**. The unsatisfactory nature of all this is frequently compounded by the rudimentary syntax of the LISP notations in which such programs are commonly expressed, which readily confounds trivialities with profundities. Until these signs of immaturity disappear it will be hard to regard the field as embodying much mature technology.

6. Artificial Intelligence and the Development of Programming Languages

As emphasized above, the most fundamental goal of artificial intelligence research is **discover**y of principles facilitating the integration of initially **fragmented material** into useful organized structures. This is also a fundamental aim of the programming language designer, who seeks languages that make it easy to use small independent code fragments to define complex processes. Such languages **eliminate** troublesome sources of programming error and can increase programming **speed** very considerably. For this reason, and because artificial intelligence researchers have regularly **grapple**d **with** unusually complex programming problems, their work has been a particularly fruitful source of advanced programming concepts.

A few of the most significant ideas of this kind are worth noting. The *LISP* language developed early in the history of artificial intelligence research introduced **power**ful means for defining entirely general and flexible data structures, and, since these also could be used to represent the particularly simple externals of the language, provided an environment in which other still more advanced programming languages could easily be implemented for **experiment**al use.

Rule-based programming aims to eliminate programmer concern **with** operation sequencing by allowing operations to be executed whenever corresponding enabling conditions are met, for which purpose statements having approximately the **form**

WHENEVER condition DO operation END

are provided. *Backtracking* simplifies the execution of complex explorations by allowing exploration to be routed along multiple parallel branches. The simplest way of providing this semantic facility is through a *choice operation* a process can create

new processes

terminate

combinations

backtrack

failed

procedures

trigger

a kind of

reality

manual

systems

disrupted by

artificial

design

to automate improvement

start from

intuition

generating programs

exploiting tricks

used by human

expressions

promising

programs
on
deep
knowledge

raise the same

superficial

problem

beyond our

Moral Limits

a new history
made
from
fear
terrifyingly fragile

imagine
predatory

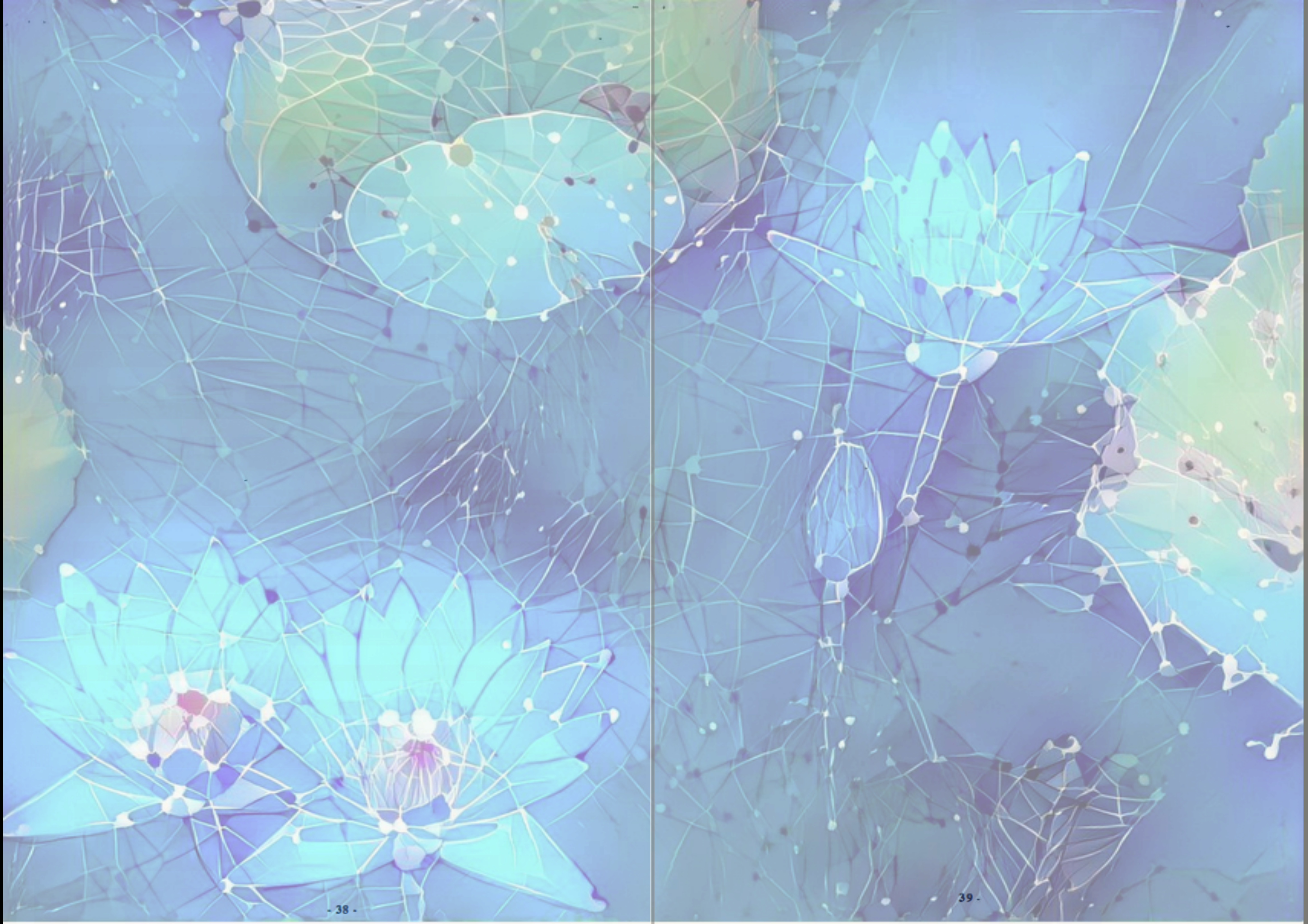
forces
entirely indifferent to

reason or

human spirit

intelligent
factories
set to
create

fiction



Credits

Original text
Schwartz, Jacob T (1986) The Limits of Artificial Intelligence. Technical Report, Computer Science Department, New York University. https://archive.org/details/limitsofartifici00schw

Source images
Photographs by J Bi Li Chan
Face image by Engin Akyurt from Pixabay

Style images
Photographs by J Bi Li Chan
Network image by Gerd Altmann from Pixabay
Circuit image by akitada31 from Pixabay

All images produced by J Bi Li Chan using Deep Dream Generator

Many thanks for valuable feedback from Hannah Jenkins and Mark McGuinness throughout the project. I am grateful for the helpful advice from Kathy Bowrey on IP issues.