Artificial Thought and Emergent Mind

Ivan M. Havel
Center for Theoretical Study
Charles University
Ovocny trh 3,116 36 Prague 1
Czech Republic

Abstract

We consider the question of whether or not a successful attempt to simulate human (rational) thought on a computer can contribute to our understanding of the mind, including perhaps consciousness. The now fashionable concept of "emergence" may turn out to be more appropriate, but still does not seem to provide a final answer.

1 Introduction

The field of Artificial Intelligence, and cognitive science in general, has had many great and celebrated acliievements over the years, but, at the same time, it has brought about quite a few controversies and heated debates about the adequacy of the computational metaphor and constructivistic methodology to our understanding of the real human mind.

It is not the aim of this paper to profess the author's support to one or another position in the dispute or to find one more flaw in any particular scholars argument. Rather, in hope of alleviating the tensions between various existing camps, I intend to propose certain conceptual distinctions which may help us find some common themes among *the* various claims made by scholars about the nature of thought and mind.

To make this long story short: I believe that a distinction between two types of mentalistic terms will serve the Ai community well, one less resistant to the adjective "artificial" than the other. I have chosen, perhaps irresponsibly, "thought" as the former term and "mind" as the latter term. My strategy is to include into the generic concept of thought, even those processes which can be externally and objectively described, and thus they are acceptable for an intentional constructive procedure.

In this respect it appears useful to consider various scales of magnitude, or levels of analysis, on which a given entity has a meaning. The distribution over scales seems to oiler a sharp distinction between complex entities in nature, like organisms or brains, on the one hand, and artifacts on the other hand (cf. also [Havel, 1993)).

Recently, with the cormectionist boom, in addition to "artificial", a new adjective is becoming fashionable: "emergent". There are certain reasons to grant it an intermediate status between "natural" and "artificial" but for some other reasons these three concepts are hard to compare. Anyway, it is too premature, I think, to argue for or against talking about emergence in connection with the mind. Thus the second part of the title of this talk suggests a challenge to, more than a project for, the AI community.

2 Making Things Think

In their textbook about AI, Rich and Knight 11991] define this field as "the study [of] how to make computers do things which, at the moment, people do better". A simple statement, each term of which, however, needs to be discussed further. Let me focus on the phrase "to make (computers) do". Obviously, it refers to two different activities, making do and doing. The nature of, and difference between, these two activities should be taken into account whenever one says that some entity (object or process) is. or may be, *artificial*

The concept of the artificial presumes, firstiy, that there is some *natural* entity which logically admits duplication (in our case it will be the process of thought, in the generic sense of the term); secondly, that a natural "person" (the *designer)* has a *prior intention* to construct a duplicate of the natural entity in question, and. thirdly, that there has to be an intentional *project,* i.e. a coherent and methodical series of intentional steps leading from the prior intention to the realization of the intention.

The distinction between prior intention and intention in project is similar to Searle's [1983] distinction between prior intention and intention in action. If I play randomly with pieces of cloth and wire and suddenly - lo and behold! - something that looks like a ttower appears in my hands, I should not, I believe, claim that I have made an *artificial* flower. Certainly I would not say this if, let us say, in playing with the cloth and wire a wind, instead of me, had created the flower. What was missing was the prior intention to make a flower. But such an intention alone is not a sufficient condition either: even if rny random playing with the cloth and wire, or my waiting for the wind to do it for me, were accompanied with my best conscious hope that a flower would sooner or later emerge, I would still hesitate to call the flower artificial. What would be missing then is the intention in project.

A project not only requires a prior intention, but also an explicit design specification, i.e., an external and objective description of all the relevant properties of the intended final product. External, because construction (unlike, for instance, learning) is an external activity; objective, because anyone else should be able, in principle, to reproduce the project in order to achieve the same outcome.

In terms of our definition of a project and the external and objective condition which a project demands, we can ask whether it makes sense to conceive of a project to create artificial thought, that is, to try to make something, perhaps computers, think. My claim is that any project of AI would require an external and objective description of the process of thought. This is a strong requirement, indeed, and if it were satisfied there would not remain, in our age of computers, too much more to do.

3 Thought and Mind

I have promised to propose a conception of thought that will lend validity to the project of AI, to the project of making computers think. Curiously, I achieve this goal by a simple trick: by proposing a sufficiently general concept of thought.

Tentatively, let us define *thought* as the act or process dealing with conceptual objects (concepts, thoughts) in an intelligible and meaningful way.

This requires some clarification. First, by a *concep tual object* we mean a representation of any entity, individuum. category, property or situation, either derived from direct experience in the real world or constructed in the course of past processes of thought. The seeming circularity (viz., reference to thought) is harmless; it actually allows for a gradual enrichment of the process.

Second, *intelligibility* means understandability by intelligent observers (for instance, by "us"). That is. the inherent overall logic of the process should be understandable, although it is not necessary to grasp the details, or the real causes of particular actions.

And third, the word *meaningful* should exclude randomly disordered, chaotic behavior on the one hand, and inert, stagnant or boringly repetitive behavior on the other hand.

These explanations betray the slight of hand of our simple trick: *the* definition does not mention causal microstructure. nor does it state an explicit requirement of a presence of an identifiable individual agent or "executor" of thought, nor does it require subjectivity or consciousness. Thus, for instance, conscious thought', artificial thought', and 'collective thought' may be three different (perhaps related) types of realization of thought. Let me illustrate the point by several examples.

(1) Assume, first. John S., sitting in the famous Chinese room ISearle, 1980] and manipulating Chinese characters. If he understands Chinese he can carry on a genuine Chinese conversation. Such a real conversation involves a real conscious thought.

(2) Now assume he does not understand Chinese. He can still perform a thought process, but it is much less sophisticated and has a different intention, namely, to manipulate the Chinese characters (for him meaningless squiggles) according to the prescribed instructions. Even this thought process is (normally) conscious but with consciousness playing a relatively insignificant role in the process itself (John S. could be easily substituted by a machine.)

3) In the previous situation, John S. and *the* Chinese room (the instructions included) perform *jointly* still another thought process. With the help of prescribed instructions for manipulating Chinese characters (symbols whose meaning is implicitly stored in tiie instructions) they behave as *if* they understood Chinese. I would classify this process as artificial and (most likely) not conscious. (Here again John S. could be substituted by a machine).

(4) Now a somewhat different case. Consider the set of thinkers who have worried themselves over the Chinese Room Problem, who talk, write, and debate about it. Together they realize a collective thought process, *not* artificial and *not* conscious (unless we believe in some sort of higher, collective consciousness) '.

Now, everything that was deliberately left aside or not mentioned in our "definition" of thought should be included in the concept of the *mind.* Relatively vague otherwise, our concept of mind should include, in particular, all important properties ascribed to it by philosophers: intentionality, rationality, free will, mental causation, subjectivity, and, above all, consciousness.

4 The Difference between Conscious and Artificial Thought

In recent years there has been a noteworthy shift of interest in cognitive sciences and analytical philosophy towards the issue of consciousness. Although in a recent monograph on neuropsychology ([Kolb and Whishaw, 1990)) the term "consciousness" does not

[1] I concede a certain degree of vagueness in using the adjectives conscious', artificial', collective', etc.. to describe a thought process: it is not clear, for example, whether the adjectives are to be understood as different *brands of process* or as different *brands of realization* of thought processes. Both interpretations are sound provided we do not take the latter alternative as meaning that *all* thought processes admit *all* brands of realization.

appear in the index, an increasing number of scientific and philosophical books are concerned with this topic (e.g. [Dennett. 1991. Marcel and Bisiach. 1988. McGinn. 1990. Searle 1992|). Whether an author grants a privileged ontological status to consciousness or whether he intends to eliminate reference to consciousness altogether, it is quite clear that AI cannot completely ignore this recalcitrant concept.

Consciousness is a part of our inner subjective experience and as such it is not, in principle, accessible to the outside observer. Therefore those functions of the mind, like conscious thought, that are intimately and inseparably connected to consciousness cannot be fully described and presented in an external and objective language without sacrificing some essential component. As we have argued above, without an explicit design specification there is no project, and without a project there is no sense in talking about something being artificial.

However, if we restrict ourselves to those thought processes which can be objectively described (recall our sufficiently general concept of thought), there should be no objection to anyone venturing into the project of realizing them artificially. In fact, when the description is sufficiently precise, complete and unambiguous, as it is in the case of formal algorithms, there are well-known standard tools for executing such a project (namely, programming systems and computers). There are limitations due to undecidability and complexity of certain tasks, but these limitations are not what we here consider.

I have argued that, in principle, conscious thought cannot be, without a substantial loss, converted into artificial thought (which claim may be taken as a rejection of a thesis analogous to the strong AI thesis ISearle. 1980|). What about the converse: is artificial thought convertible to conscious thought? In a certain sense and in certain cases the answer is. obviously, yes. The sense is the following: processes of thought which admit artificial instantiations admit, in principle, also instantiations with conscious control. Recall case (2) above (manipulating meaningless squiggles in accord with a prescribed set of instructions).

One may argue as follows. If there are instances of artificial thought processes realizable as conscious thought processes, does it not follow, contrary to the above claim, that at least some conscious processes, namely those just mentioned, can be realized artificially?

This argument is valid, but not philosophically interesting. Our claim does not deny the possibility of extracting certain parts or certain components of conscious mental activity and converting it into, say, a computer program, with the behavior or functional structure equivalent to the behavior or functional structure of that mental activity. There are two possibilities. Either the presence of consciousness is not essential for the mental activity in question (as in example (2) above) in which case the equivalence, being trivial, does not say anything at all about artificial realizability of consciousness or consciousness is essential for the mental activity (which may be the case of example (1)) in which case extracting the programmable component will yield an entirely different thought process.

The issue of inner (first-person) conscious mental life as opposed to an outer (third-person) point of view has been thoroughly discussed by several authors, most recently by Searle [19921. My aim in the rest of this article is to concentrate on another aspect of mental processes or on their (natural or artificial) realizations. This other aspect is related to the variety of scales of magnitude relevant, or essential, to these processes. For this purpose (and for those who prefer the visual metaphor) let me first introduce a new imaginary "dimension" corresponding to varying scales of space and time.

5 Scale Dimensions in Nature

Perhaps we can best start with an example: consider ordinary, geographical. 2-D maps of the same region but of different scales, superimposed one on top of another, with larger-scale maps on top of smaller-scale maps.

We can imagine scales (expressed, say. by real numbers) without limits both "downward", to the small, microscopic and submicroscopic scales, as well as "upward", to the large, astronomical scales and beyond. Assuming, moreover, a dense sequence of scales we obtain a continuum represented by a special coordinate axis. Let us call it the *scale axis* and the corresponding dimension the *scale dimension.*

Somewhere in the "middle" of the scale axis exists our habitat, the *scale-local* world of human magnitudes, our "scale-here". Unlike the ordinary spatial "here", the natural "scale-here" has the same position on the scale axis for all people (and perhaps for animals of about our size). Both spatial "here" and scale "here" are smeared and cannot be exactly localized to a point.

In the same way as we introduced the scale dimension for space, we can introduce the scale dimension for time representing various magnitudes or "speeds" of time. Again there exists a natural "scale-here" for time scales, corresponding to the rhythms of human life and thus, under normal circumstances, common to all people. In particular, our thought processes span the interval roughly between milliseconds and hours.[2]

6 Things, Events, and Artifacts: Their Distribution over Scales

The scale-local objects (entities of size/or duration basically accessible to humans) are a special case of

[2] In fact, the 100 millisecond scale has been proposed as the demarcation line between the classical symbolic-algorithmic paradigm in Artificial Intelligence (above 100 rns) and the subsymbolic-connectionistic approach (below 100 ms) IHofstadter, 1982]. (My opinion is that what plays really the main role is the interaction between distant levels; cf. Section 13).

objects that I shall call *scale-thin,* i.e. objects "occurring" in a limited range of spatial or temporal scales. We are used to organizing real-world entities according to tlieir relaUvely specific position on the spaUal and/or temporal scale axis and according to the mutual relation of their positions into several categories The most common of our organizing concepts are things and events.

By *things,* we Intuitively and typically mean those enUUes which are separable, with identifiable shape and size, and which persist over time. *Events,* on the other hand, typically have a relaUvely short duration and are composed of the interaction of several, perhaps many things (of various sizes). However, in the world of all scales there is no essential difference: things are just long-lasUng events and events arc just short-lived things (where "long" and "short" are relative with respect to our temporal scale perspective). Many other entities (vortices, clouds, flames, rivers, networks, sounds, bubbles, winds, ceremonies, meetings, wars) have an intermediate character.

It is customary, when describing concrete phenomena in the world, to treat tlieir spaUal structures (shape and internal composiUon) and their temporal structures (internal dynamics and behavior) separately. This separation, together with our scale-thin language and our scale-local perspective, helps conceptually, but, at the same time considerably narrows our perception of reality (in a way reminiscent of the Baconian "idola tribus" - idols of the tribe).

For illustraUve purposes, let us restrict ourselves to the case of spatial scale dimension. The reader is invited to make his own generalization to the temporal scale dimension.

Various, typical objects that we observe in nature can be categorized according to their "distribution" over scales. Let me explain what, here, distribution is. First, for each such object we identify various scales relevant to its spaUal features, like, for instance, sizes of its components and relative distances between interacting components. Obviously, the scales cannot be identified sharply so that a continuous function over the scale axis with salient peaks or elevations around certain scales would result. Let. us call this function the *relevance Junction* or the *{spatial) scale spectrum* of the object.

Now, according to the distribution of peaks in the scale spectrum we can, in a first approximation, idenUfy four basic (even if not sharply separated) categories of objects. First, there are *single bodies* (stars, stones, dust parUcles) with only one salient peak in , the scale spectrum (if we ignore the lower, molecular and atomic structure). Second, there are clusters (galaxies, clouds, heaps) with two or more sparse peaks Third, there are *complex systems* or *organisms,* typified by a large number of relaUvely dense but still distinguishable peaks spread over a certain interval of scales (I shall return to this category later). Fourth, there are *scale-homogeneous structures,* i.e. objects wiUi continuous spectra (in a certain interval) of relevant scales. These last structures are relaUvely rare in nature; examples are fractal shapes and also fluids near Uieir criUcal points.

Up to now, we have classified natural objects. To extend the concept of scale spectra to human artifacts (tools, engines, computers, houses, cities), we have to reinterpret the notion of (scale) relevance. Obviously, what matters here is much less the relevance (of various scales) for those objects themselves than the relevance for, or from the point of view of, those who conceive, construct and use them. There are, often only a few, relevant (in this sense) scales for such objects, occasionally separated by large gaps.

Consider, for instance, the computer as a physical object. The most important spaUal scale is the local scale of users (it is the scale on which the computer is designed and meaningful). Then there are several well-known relevant scales (of hardware architecture, processing units, logic circuits, semiconductors, down to the scale of quantum phenomena), each associated with a special engineering discipline and with a special design and specificaUon language.[3]

Whatever the scale spectrum in the designers' perspective is, there is always one and only one relevant scale (peak in the spectrum) for most artificial objects, including computers. It is the local scale "here" of us, the users, where the *meaning* of the object is located.

Basically the same holds for any natural or arUficial language and, in general, for any symbolic representation (provided humans can read the language and can interpret the symbolism). Symbols, symbolic patterns, and syntactic objects are suitable for conveying meaning only within a narrow range of scales, beyond which narrow range they are incomprehensible. We should bear this in mind when discussing differences between artificial and natural thought.

7 Levels and Their Hierarchies

In the case of complex objects, there is a close relationship between their distribuUon over scales and a hierarchy of their *structural, functional, or descripUonal levels.* In many situaUons in which we find it convenient to talk about various levels, we can also distinguish corresponding scales or ranges of scales, spaUal as well as temporal. Accordingly, Salthe [1991] uses the generic term *scalar hierarchy* whenever the levels are characterized by different scale, as opposed to the *specification hierarchy* with levels based on degree of specification or generality. While a scalar hierarchy may be based on the part-whole distinction. a specification hierarchy is typically based on the token-type disUncUon.

If a certain structure or object has disUnguishable salient peaks or elevaUons in its scale spectrum, it is natural for us to associate with such peaks and elevaUons appropriate levels of a scalar hierarchy. Moreover, because our languages are not suitable for large

[*] In actual computing, the temporal scale spectrum may be more important.

scale span, we develop specific descriptional languages for particular levels.

To illustrate this point, let us consider a cluster, for instance a cloud of dust. Such an object may be studied on the global level (for example, in terms of its overall shape and size and in terms of its global properties, like opacity) or on the level of its components (the dust particles, their individual properties, such as distribution, density, and interaction).

One may. of course, ask about the actual reality of levels as such. Do they exist independently of our analysis and description of objects and events? I believe that sometimes they do. at least partly, but that sometimes they are, again partiy, our mental constructs. The scale-thin world of our ordinary perceptions and thoughts makes it difficult for us to grasp more than a certain limited range of scales at once. For this reason, and other reasons, we tend to decompose objects of our concern into structural levels and events (and processes) into functional levels. Obvious differences of individual levels yield different descriptions, different languages and. eventually, different disciplines. If all is done properly, the decomposition may match something which approximates the real differentiation of nature.

There is one problem whicb may be crucial for our understanding of complex systems: whether and how can distinct (possibly distant) levels of a system directly interact. In the following section, I shall make a few comments on this problem.

8 Interaction across Levels

The scalar hierarchy is commonly treated in systems science with the tacit assumption that "constitutive dynamics at different scalar levels are largely screened off from each other (non-transitivity of effects across levels)" (ISalthe. 1991|, p.252), and with the resulting belief that "three contiguous levels should be sufficient to understand most of the behavior of any real system" (p.253). This assumption and belief are, in fact, included already in the term 'hierarchy' (in contrast to 'heterarchy'). According to system scientists, occasional influences from distant levels are generally considered as "perturbing fluctuations" that need not be included in a dynamical description of the system in question. I think this view is inherently connected with the explanatory role of causality in science.

Scientists base their understanding of the processes of nature mostly on *causal interaction;* it is the principal explanatory apparatus Yet the scale-thin conceptual field restricts our experience and intuition more or less to an *infra-level* "left-right" causation: we refer to an earlier event to explain a later event. Therefore, the ease of treating the *inter-level,* micro-macro causation as unproblematic, indeed the only acceptable, type of interaction between levels [Searle, 1992| is somewhat surprising.

Of course, there are cases in which it is quite legitimate to employ causal explanations between levels, usually from one level to an adjacent higher level. For instance: the properties of molecules cause the growth of a crystal to a specific global shape, the disorganized movement of molecules causes Brownian motion of larger particles, etc. But one has to be careful about generalizing this way of thinking to everything. For example, to say that "mental phenomena are caused by neurophysiological processes in the brain and are themselves features of the brain" [Searle, 1992] suggest too liberal an interpretation of the term "are caused", even if we agree, for the sake of understanding, on a strong assumption, namely, that there is a natural hierarchy of levels above the neurophysiological one with some higher level attributable to the mental phenomena.

One of the few theories in science that deal with inter-level interactions is Gibbs-Boltzmanns statistical physics (thermodynamics and the study of collective phenomena) It succeeds rather by eliminating the lower (microscopic) level from the macroscopic-laws through decomposition of the phase space to what is considered macroscopically relevant subsets and by introducing new concepts, such as entropy (which is, of course, a wonderful trick). As an indirect result we obtain, for example, an "explanation" of macroscopic asymmetry of (physical) time. Can we, however, really say that the time asymmetry is caused by behavior of particles?

There is a well-known technique of renormalization (cf.e.g.. [Wilson. 1979]) which deals with problems that have multiple scales of length. It is particularly suitable for phenomena near critical points and has applications in various branches of physics. But it is not a descriptive theory of nature and, therefore, has little ontological relevance.

Another relevant area is the study of deterministic chaos. Here people become more and more used to situations in which extremely tiny fluctuations are almost instantaneously amplified to a macroscopic scale. What seems to be a purely random event on one level appears to be deterministically lawful behavior on some lower level. This is (mathematically) a recurrent situation. We can, therefore, take the deterministic description as something permanently hidden behind the scale horizon, albeit always available formally for explanatory purposes.

Particularly interesting, and surprisingly much neglected, is the compelling question of the asymmetry of interactions with respect to the scale axis. Why is the arrow of putative causality, or of other natural influences, usually assumed to have a direction from lower levels to upper levels? Is it the heritage of the clock-work mechanistic conception of the nature or one of its inherent asymmetries? (It should be noted that not everybody excludes the idea of "downward" causation [Popper and Eccles, 1977]).

In this respect, we can find some inspiration in the notion of information (as an ontological category; cf. Bolim's concept of active information [Bohm, 1990]) and, even perhaps, in the notion of the mind (cf Section 10).

9 Organisms and Brains: Multilevel Interactional Structures

We have used (in Sec. 6) the term organism for natural objects with many relevant scales in a large interval; alternatively, we can talk about a relatively dense hierarchy of structural and/or functional levels. Moreover, in living organisms, tliere is a strong mutual interaction between particular levels. Both density and interaction are crucial features here. In the computer, for example, tliere are several prominent levels and tliere is (some) interaction; but there is no density. In the fractal tliere may be density but there is no (physical) interaction.

Let us state our key question: is it not the very existence of such a hierarchy of interacting levels that makes living organisms (and brains) diflbrent from computers, clouds, and fractals? Is it not perhaps just this property that makes it so difficult to submit their behavior to a mechanistic explanation?

1 doubt that the function, meaning, and actual being of living organisms can be associated with some particular level (or scale). Each of these aspects takes place equally well on the level of molecules as on the levels of cells, organs, individuals, social groups or ecosystems. It comes about at many dilferent scales of space as well as time.

Consider the following opinion of the physicist Barrow ([1991], p.97): "There exists a form of hierarchical structure in Nature which permits us to understand the way in which aggregates of matter behave without the need to know the ultimate microstructure of matter down to the tiniest dimensions." I am afraid that in order to accept this statement we would have first to restrict considerably the meaning of 'understand' and 'behave'. Otherwise, if we wanted to apply Barrow's statement, to so complex an organism as the brain we would run into problems of identifying the depth below which further levels cease to be significant.

In fact, there is a growing number of works suggesting a certain relevance to the brain activity of all levels down to the quantum scale. For instance Beck and Eccles [1992] propose a mechanism whereby the probability of exocytosis of synaptic transmitters is, by means of a quantum-mechanical tunneling process, influenced by mental events. Incidentally, if the lower level quantum effects happen to have a certain important role in the conscious mind, this may yield an argument for the unique scale location of the overall size of the brain.

In general our considerations do not favor reductionism in biology and psychology. Indeed, to understand life and the mind does not mean to reduce it to some basic components, but rather to appreciate various influences, bounds, and interactions between all structural and functional levels, close as well as remote with respect to the the scale dimension.

10 The Location of the Mind

Perhaps a similar expansion of the scope of view might help us to understand better the nature of the mind and consciousness, or at least to avoid certain persistent fallacies. One such fallacy is the belief, held by some cognitive scientists, that the mind is nothing but a collection of processes occuring on a certain privileged level above the level of neurophysiological processes in the brain. This fallacy is probably caused by at least two misconceptions.

The first rniconception may be the overjudgement of the computer metaphor. If someone maintains that the brain is a (sort of) computer equipped with programs and that the mind is a collection of such programs (or processes controlled by them), then he is immediately drawn to the language metaphor (that the programs may be "written" in some language). And, since languages happen to be scale-thin (cf. Sec. 6), it is natural to take mind to be scale-thin too, i.e. restricted to a specific level namely the same as the level of (human) communication.

The second misconception may be the confusion of the intentional content of mental states with those states themselves. This content, i.e. topics of our beliefs, objects of our perceptions, images of our fantasy, and goals of our plans, are primarily tilings of ordinary size - "things wliich a baby can handle and (preferably) put into his mouth" [Popper and Eccles, 19771. Mental states are believed to be neurophysiological states; because it is absurd to think that neurons can handle the same things as babies, it is taken for granted that the mental level is sufficiently above the neuronal level A similar attitude is held even by some of those philosophers of the mind who are bravely opposing eliminative and reductionistjc materialism. For instance, Searle [1992] claims that "conscious states are simply higher-level features of the brain" (p. 14).

But is it proper, in the context of mental phenomena, to talk about "levels" at all? Even if we did not restrict ourselves to the scalar hierarchy (cf. Sec. 7), it would be a mistake, I believe, to treat mentalistic terms as something that should be, by their use, carefully confined to a certain "level", "domain" or "subject area".

Hofstadter [1979] was one of the first authors who discussed the connection between mental phenomena and a hierarchy of levels. He rightly pointed to the importance of inter-level interaction on distance (including loops) but he wrongly embedded mental levels into the scalar hierarchy of functional levels of the brain and lie confused the latter with the specification hierarchy of indirect reference (p. 709). Moreover, his functional "holism" is not quite consistent with his own warning against the use of nonditterentiated language for dilferent levels of description, to which use he ascribes the many confusions in psychology.

It may well be the other way round. Perhaps all these confusions come from precisely such strict fragmentation of concepts into levels. If, for instance, consciousness is to be understood as a property of a body, it certainly should not concern certain some or another single level but the whole living organism "penetrating" through many mutually interacting and cooperating levels, including, perhaps, even the level of quantum physics.

11 The Emergence of Emergence

I suggested that we can limit the use of the term "artificial" to those cases in which there is a clear distinction between the project and *the* designer, and where an external design specification is possible and both a prior intention and an intention in project are present (Sec. 2). This applies very well to the traditional logical-symbolical-computational AI (nicknamed GOFAI - Good Old-Fashioned AI).

Recent advances in connectionist architectures suggest an alternative strategy which blurs the above distinction and in which there is no presumption of an a priori external description of the structure of the task [Havel, 1992]. One may construct a complex dynamical system in the form, typically, of a large number of mutually communicating units, and then patiently wait for an appearance of some complex emergent phenomena that might support certain processes of thought. Of course, this type of thought would be somewhat alien and incomprehensible to us, and it would not fall, in our terms, into the category of the natural nor of the artificial. Indeed, it would not be (entirely) natural because the requisite dynamical system (connectionistic or other) is artificial and it would not be (entirely) artificial because there is no intention in project (cf. Sec. 2).

This alternative strategy brings us to the idea of *emergent mind.* The concept of *emergence,* especially in the context of mental phenomena, deserves a separate study There are actually three meanings of this word, not always easy *to* distinguish. *The* traditional meaning in evolutionary theory (G. H. Lewes in mid-19th century, C. Lloyd Morgan in early 20th century) emphasizes the temporal aspect: the rise of a system that cannot be predicted or explained from antecedent conditions, e.g. the emergence of life, of man, etc.

In its second meaning, the word emergence has recently been increasingly used for phenomena that appear to be natural on a certain level of analysis, but somewhat resist (though not completely exclude) their reduction to an appropriate lower level. Typical examples of this are collective or mass properties: the liquidity of water (a molecule of H_2O is not liquid) or democracy in society (one person cannot form a democratic system). Often this type of emergence is only apparent when caused by the intractable complexity of the lower-level structure.

The third meaning is inspired by the second and is used often as an argument against reductionism. A property (specified by a certain theory T,) is said to be (properly) emergent if it has real instances, if it is cooccurent with some property recognized in a reducing theory T_2, but which cannot be reduced to any property definable in T_2 (cf. [Churchland, 1986], p. 324). *Property dualism* is characterized by the conviction that "even if the mind is the brain, the qualities of subjective experience are nevertheless emergent with respect to the brain and its properties" (ibid, p. 323).

Let us consider the thesis that thought occurs as an emergent phenomenon on some higher level of a hierarchical system, with low levels being purely mechanistic. This thesis would help materialistic monism avoid the concept of the soul as an unknown ghostly substance, which is regarded as flowing or flying in another world. However, if the motivation for tills avoidance is the mere resistance to accepting an unknown or unknowable entity, then not even the concept of (proper) emergence will help, at least until something more is known about the matter. For instance, as 1 pointed out earlier, there is little or no understanding of interaction between different levels.

On the other hand, the ernergentist thesis cannot be easily refuted and we can tentatively accept It for the sake of discussing the chances of connectionist AI.

12 Collective phenomena

Global behavior of a connectionist system can be viewed as a specific case of a much more general concept of a collective phenomenon. *Collective phenomena* typically occur in large collections of individual units, the behavior of each unit being partly dependent on the behavior of some other "neighboring" units. The remaining, autonomous part of this behavior may be based on a random and/or rational individual decision.

The connectionist model is an example of a one-level system (all units at the same level of description). In contrast, the concept of a *hierarchical collective systern* incorporates the idea of an iterated division of larger *tasks to* smaller subtasks. This idea *is* natural for the top-down AI strategy ([Minsky, 1985]), but at the same time it may support emergent collective phenomena.

According to the weight of the autonomous part of behavior of individual units, we can distinguish two opposite modes of global behavior or, using the language of statistical physics, two *phases:*

(1) the rigid, bureaucratic system of primitive obedient agents (low autonomy), and

(2) the chaotic, anarchic system where everyone does whatever he likes.

(Cf. [Dennett, 1991], Chapter 9). Various disciplines, from physics to the social sciences, oiler many examples of mixed or intermediate cases. For instance, we may have a system of initiative agents, competing for recognition, each with his own idea, while at the same time all are attentive to the ideas of their colleagues. In a rigid system, a good, new idea can occur only with great difficulty; in the chaotic system, it is always lost; but, in the intermediate system, it may propagate easily through large areas of the network. In physical systems we encounter similar situations near phase transitions [Little, 1990]. A great deal of attention has been recently paid to the intermediate case, called the edge of chaos, featuring interesting properties, among them a sort of evolutionary stability.

Collective systems with highly parallel activity are interesting alternatives to classical serial-computational models. Dennett [1991] uses the idea of multiplicity of competing agents (called homunculi) in his theory of consciousness. Collective systems have extremely large combinatorial complexity (the number of global states grows exponentially with the number of units). Such a complexity is not, however, a disadvantage. It yields redundancy and redundancy supports self-organization and self-improvement.

Another interesting collective phenomenon is the emergence of islands of *cooperative behavior or altru ism* in a large set of egoistic individuals [Axelrod. 1984].

13 Towards a Scale Holism

Let me conclude with a few more or less speculative remarks.

I cannot resist a suspecting that the term "emergence" is currently used mostly to conceal our ignorance of links between entities on a certain, relatively well-uderstood level, and entities observed or assumed on the next higher level (for reasons to be seen soon let us call it the *first-order emergence).* There is nothing wrong with such using the term provided it is just a first step in future reserach, at least in two directions.

One direction could be to study more thoroughly the nature of those links. We may ask, for instance, to what extent is the first-order emergence (from lower to upper level) reducible to phenomenona studied in nonlinear dynamical systems theory, like, let us say, greatly amplified fluctuations.

A second direction could be, I believe, of more critical importance, especially if we want to study such evasive entities as mental phenomena. I have argued above (in Sec. 9) that certain complex systems (organisms, brains) should not be understood as scale-thin objects, but, rattier, as structures penetrating through many scales and featuring interactions through many levels. This argument suggests good reasons for introducing a new type of emergence that I shall call the *second-order emergence.* An entity is second-order emergent if it is not associated with some particular level but arises from global interaction of many levels of some scale-extended complex system.

My thesis is that, if the mind admits of physicalistic or naturalistic understanding at all, then such understanding should be in terms of the second-order emergence rather than the first-order emergence. This applies also to its essential components, consciousness for example.[4]

One particular example might illustrate the point. One of the most important components of the mind is memory. This, originally mentalistic, concept is now more and more used as a feature- of a physical system - either the brain or the computer. In fact, the knowledge-memory distinction is an interesting variation on the mind-body distinction. If knowledge were emergent on a certain specific level of the scalar hierarchy, the next lower level would play the role of a substrate for syntax: it would house meaningless structures with combinatorial features allowing sufficient differentiation and assignment of atomic units of meaning. Not only this assignment but even the combinatorial features can be specified only by means of an external activity (of a designer or observer) which cannot be an outcome of spontaneous emergence. (In this respect cf. lSearle 1992|.)

Therefore a truly emergent knowledge has to have an extended presence over many levels or over many scales with smooth downward degradation of a local semantic content. This can take place in a rather rudimentary form already in distributed connectionistic systems, where very low-level objects (units and connections) are bearers of a minimal semantic content, undetectable but "ampliflable", typically through the cooperation of many objects [Havel, 1990].

Some researchers hope that by incorporating the connectionistic or cooperative approach into AI, one may achieve substantial progress in the general project of the artificial mind. Perhaps the introduction of these approaches can be considered as progress, not, however, towards anything artificial and probably not significantly towards what is essential about mind.

As I have tried to clarity, the artificial involves both intention and project, and the project involves both specification and solution. Now, overcoming *the* specification issue (lack of descriptive means) by constructing only a lower-level substrate (for instance an artificial neural net) and, then, by letting emergence work is a great idea. However, this strategy tends to eliminate the designer's intentional component. Moreover, such "hybrid" methodology works only in the case of first-order emergence. If it turns out that the second-order (scale-holistic) emergence is a necessary component of the mind, the hybrid methodology will fail, unless, of course, we discover the means for "creating" scale-extended objects - real organisms.

References

[Axelrod, 1984] R Axelrod. *The Evolution of Cooperation.* Harper and Collins 1984.

[Barrow, 19911 John D. Barrow. *Theories of Everything.* Oxford: Clarendon Press 1991.

[Beck and Eccles. 1992] Friedrich Beck and John C. Eccles. Quantum aspects of brain activity and the role of consciousness. *Proc. Natl Acad. Set USA* 89, pages 1 1357-1 1361. December 1992

[Bohm, 1990] David Bolim. A new theory of the relationship of mind and matter *Philosophical Psychology* 3:2. pages 271-286, 1990.

[4] I am not concerned here with the question of subjective nature of the mind.

IChurchland, 19861 P.S.Churchland. *Neurophilosophy: Toward a Unified Science of the Mind/Brain.* Cambridge, MA: The MIT Press 1986

[Dennett 19911 D.C.Dennet *Consciousness Explained.* Boston: Little, Brown and Co. 1991.

[Havel, 1990] I.M.Havel. Connectionism and Unsupervised Knowledge Representation. In: *Cybernetics and Systems'90* (RTrappl, ed.) pages 1001-1007. Singapore: World Scientiilc 1990

[Havel, 19921 I.M.Havel. Artificial Intelligence and Connectionism. In: *Advanced Topics in AI* (Mafik, Stepankova, Trappl, eds.) pages 25-41. Lecture Notes in Artificial Intelligence, Berlin: Springer-Verlag 1992

[Havel, 1993) I.M.Havel. Scale Dimensions in Nature. Research Report CTS-93-03. Center for Theoretical Study, April 1992

[Hofstadter, 1979| D.R Hofstadter: *Godel, Escher, Bach: An Eternal Golden Braid.* Brighton: Harvester Press 1979

[Hoistadter. 1982| DR.Hofstadter. Waking up from the Boolean dream, or. subcognition as computation. *Scientific American,* July 1982.

[Kolb and Whishaw, 1990] B.Kolb and I.y.Whishaw: *Fundamentals of Human Neuropsychology,* New York: W.H.Freeman 1990

[Little, 1990) W.A. Little: The evolution of non-Newtonian views of brain function. *Concepts in Neuroscience* 1:1, pages 149-164, 1990.

[Marcel and Bisiach, 1988) A. Marcel and E. Bisiach (Eds). *Consciousness in Contemporary Science.* Oxford: Oxford University Press 1988.

[McGinn, 1990) C.McGinn. *The Problem of Consciousness.* Oxford: Blackwell 1990.

IMinsky, 1985) M.Minsky: *The Society of Mind* New York: Simon & Schuster 1985

[Popper and Eccles, 1977] K.R Popper and J.C.Eccles. *The Self and Its Brain.* Berlin: Springer-Verlag, 1977

[Rich and Knight. 1991) Elaine Rich and Kevin Knight. *Artificial Intelligence.* 2nd ed. McGraw-Hill, New York, NY, 1991.

[Salthe, 1991) S.N.Salthe. Two forms of hierarchy theory in Western discourses. *International Journal on General Systems* 18:3, 251-264. 1991

[Searle. 1980) John R. Searle. Minds, brains, and programs. *Behavioral and Brain Sciences,* 3: 417-424, 1980.

[Searle, 1983) John R Searle. *Intentionality.* Cambridge: Campridge University Press 1983.

[Searle, 1992) John R. Searle. *The Rediscovery of the Mind.* Cambridge. MA: The MIT Press 1992.

[Wilson, 1979) Kenneth G. Wilson. Problems in physics with many scales of lenth. *Scientific American* 241:2. pages 140-157, August 1979.