

*Irreconcilable Similarities:
Man and Semantic Machines*
HERBERT HRACHOVEC

THE BIOLOGICAL structure of certain animals is very similar to that of the human body. And there are machines that can reproduce certain cognitive performances perfectly. It should be possible to combine these facts in building an *animal rationale*, artificially constructing man out of given components. That such a thought is repugnant to the humanistic tradition is hardly a reason for rejecting it out of hand. Instead of appealing to some mysterious essence of "being human," this tradition should be made to fight in order to make clear what it means by its emotionally charged pronouncements about man's incomparability to intelligent biomachines. My line of thought will in fact lead to the traditional conclusion, but it will demand controversial decisions at some crucial points. The issues raised by recent developments in cognitive psychology and artificial intelligence do not simply and inevitably lead to a humanistic dismissal of the subject, and I shall be trying to map out the fundamental turnings of the road that leads to a conviction that nothing can be compared to man.

This conviction sounds both commonplace and arrogant. In arguing it, decisions will have to be made that lead in the desired direction only if the desirability of the result is presupposed. That well-known difficulty of hermeneutical reflection is in fact appropriate to this subject matter. The "essence of man" is not a given which can be approached by arguments free of prior metaphysical bias. There is no hope of avoiding some very partial presuppositions in expounding my thesis, and any counterargument might find itself in a similar situation. But rather than indulge in mutual polemical deconstruction I choose a tentatively edifying procedure: there are some very good reasons to hold that recent research does not come near to showing that man can be considered as just an extraordinary complex artifact.

Biology will not be a factor in my argument. Since biology is a science exemplifying a particular form of cognitive capacity, its role within an investigation of what it is to be human is subordinate to the main problem, that of finding out if intelligent behavior (such as science) can be produced artificially. The hold on nature that would allow us genetically to construct a rational animal assumes a cognitive distance from nature that can be discussed without recourse to biological facts. A highly suggestive current account of intelligence regards it as something that in principle can be generated by "semantic machines." Treating certain systems (computers, brains) as mechanisms working towards potentially meaningful results by purely formal procedures has proved to be a fruitful research program. Think of a jigsaw puzzle. The shape of its pieces contains no information about the content of the representation that has to be retrieved. Finding out how the pieces fit together is a syntactic activity that can be performed according to formal principles. All those pieces just fit together in the end; but, remarkably enough, a picture of something has been assembled by this process. Evidently it is possible, by appropriate construction, to integrate formal procedures and the more complex relationships between signs and their interpretation. A puzzle illustrates semantic machines insofar as it leads to representation of reality in the absence of any prior semantic information. It is an appealing paradigm for someone looking for a link between the neurophysiological functioning of the brain and its involvement within the ongoing process of figuring out how the world is. If the puzzle of our nervous system fits together properly, we have thereby arrived at a representation of the world. The game has obviously been designed to harmonize the syntax of its pieces with the semantics of their resulting configuration. Like-wise evolution is thought to have led to a brain design that squares with those external impulses that are interpreted as impacts of "the world."

It is important to be clear about what is meant by semantics in this context. An array of formal tokens and types as well as manipulative rules is given, and meaning is assigned to it according to additional rules, resulting in an interpretation. Correct handling of the pieces of the puzzle leads to something that can be invested with

overall meaning in accordance with the fragmented meanings that (could) have been associated with the basic vocabulary (the pieces of the puzzle, in our example). For this account, what material the syntactic system is made of is completely irrelevant. The printout of a computer or the sign language of pantomime can be analyzed in exactly the same way: syntactic tokens are produced and related to something they signify. One of the most ingenious semantic machines, Richard Montague's "Universal Grammar," begins with the refusal to distinguish between artificial and natural languages. Formal semantics, of which Montague's work is an eminent example, simply has no way to mark out any meanings that have to be reserved for use in reconstructing human behavior. It does not matter for my purpose what the semantic universe is usually thought to consist of. But one important fact about its construction has to be stressed. Whatever candidates for meanings are considered (individuals, sets, structures, functions from possible worlds, truth values) they are assigned to formulas of some given object-language within a meta-language. In order to specify an interpretation, the meaningless expressions we are dealing with have to be quoted and given a semantic correlate, as only an already functioning language can do.

Obviously the metalanguage must include semantic operations. If it only were one more object-language, all that could be done would be to coordinate meaningless expressions from one language with meaningless expressions from another. Someone ignorant of the rules of pictorial representation could discern all the representational elements in a puzzle, but he would find them quite as senseless as the contours of the single pieces. But where can this pre-supposed meaning come from? We are approaching the first controversial decision. According to logico-linguistic usage, an isomorphism between two structures (one syntactic, one semantic) is all that is needed to give meaning to the expressions of a language. Whatever meanings are, just arrange them in systematic coordination with your syntax-to-be-interpreted. Look at the puzzle again. So long as the conventions of the syntax and of the metalanguage are observed, meanings can be ascribed to its formal constituents. This answer, however, evades the original question. It simply presupposes that a usable metalanguage is already available, upon whose semantic

content the expressions of the object-language can be mapped. Formal logic is of no further help here, since it has no hold on the distinction between formal constructions and the actual behavior of language users within which they are embedded. But a matter of principle has to be decided at this point. Daniel Dennett has suggested that starting from a very primitive level it may be possible to reconstruct all human semantical activity as basically consisting of an extremely complex formal system expanding the underlying syntax and progressively interacting with it. Somewhere along this reconstruction the crucial difference between meaningless signs and interpretative activity would have to vanish, giving way to various degrees of formal perspicuity. Such a hypothesis clashes with one of the intuitions on which the humanistic thesis is founded: every intelligible sense of "formal system" implies that such a system has to be *used* by someone in order to have any connection with meaning.

Formal systems by definition are devoid of meaning; they cannot simply acquire a semantic dimension by chance. Monkeys typing at random will never produce the last chapter of *Ulysses*. But isn't this to start from too exacting premises? Modest piecemeal reconstruction should lead to a gradual passage from mechanical functioning to configurations that can be endowed with sense. There is, however, a conceptual difficulty with this proposal. Isolated formal interpretations still remain on this side of sense; genuine interpretations involving the use of language can supply meanings, but in this case the metalanguage cannot be considered as a formal system. So our choice is between postulating the advent of a "formal system" that erases the distinctions on which our present concept of a formal system is built, or trying to work out the distinction implicit in the conceptual frame we actually use. While the first possibility is an exciting stimulus for science fiction, to take it as a starting point has the severe drawback that we literally do not know what we are talking about once we decide drastically to redefine the construction of our language. It is fun to experiment with the plasticity of common concepts, but the fun comes from the dizzying interplay between what we are able to comprehend and what only seems to be comprehensible. In short, in keeping apart science and science fiction,

one is obliged to elucidate the difference between formal operations and their application. As J. A. Fodor puts it:

The computer which prints out "Robin Roberts won 28" is not thereby referring to Robin Roberts. But, surely, when I think: Robin Roberts *won* 28, I *am* thinking about Robin Roberts and if not in virtue of having performed some formal operation on some representation, then presumably in virtue of something else.

This "something else" has to be the performance of nonformal activities. To this it could be replied that such activities might very well be formalizable, as Richard Montague's formal pragmatics and context-dependent interpretation in artificial intelligence show. But this only restates the underlying conflict: such behavior cannot entirely be formalized without either radically changing the meaning of "formalization" or else introducing the same problem one level higher up.

Faced with this choice, the humanistic tradition decides to work within the confines of the language we currently employ. (Its own verbal transgressions will become prominent later on.) It sticks to the commonplace assumption, vigorously defended, for example, by John Searle, that computers just do not understand the signs they are handling. One has to be aware of the thoroughgoing split within semantics this entails. Formal ascription of meaning has to be sharply distinguished from meaning conferred upon signs by using language. Assigning truth values to sentence-tokens and getting someone to assent to an utterance are very different activities. It is in fact extremely misleading to talk of "meaning" in both contexts. Formal semantic theories and theories built on the insight that semantics depends on actual verbal utterances are most likely incommensurable. The term is useful, nevertheless, in upholding the precariousness of semantics. It is either a discipline that is torn between formalization and use or one that submits to unification according to a single principle. In arguing the former view I shall find a close connection between the heterogeneous structure of semantics and the metaphysical constitution of man. Reducing semantics to formal

computation, as the latter view does, will turn out to be a modern version of explaining everything there is from a utopian point of view that aspires to absolute homogeneity.

Among the most famous semantic machines philosophy has ever constructed is Wittgenstein's *Tractatus*. It was supposed to yield a syntax incorporating every possible sense in advance of its actual employment. All well-formed sentences of its ideal language would have been guaranteed meaning by conforming to the unutterable laws of universal logical form. Wittgenstein found out that this program could not be realized. It runs into trouble right at the beginning, when an interpretation of the primitive expressions of the language has to be given. "Every interpretation together with what it interprets is suspended in the air; it cannot serve as its support. Interpretations alone do not determine meaning" (*Philosophical Investigations*). An entire network of presuppositions taken for granted underlies every successful interpretation. Its rules can be sorted out and treated formally, but then the problem returns with the attempt to specify the meaning of those higher-order formal expressions. Saul Kripke, in his recent book, *Wittgenstein on Rules and Private Language*, has convincingly outlined this difficulty for the formalist. Following a rule of language is an event entirely different from the functioning of a preprogrammed machine. Wittgenstein noticed the difference between actual language and calculi used for illustrative purposes when he pointed out that "in philosophy we often *compare* the use of words to games and calculi with fixed rules but we cannot say that, whoever uses the language, *has* to play such a game." No written command prevents one from using a signpost the wrong way around, and if there were one it would itself be subject to countless misunderstandings. Human social behavior is the basis for meaningful employment of signs.

In following these suggestions one finds oneself using language rather strangely. It is often said that for a rule to be understood and the meaning of an expression to be grasped, there has to be "some-thing in addition to the mere unraveling of a program." What might this mysterious capacity consist of? Simply positing it in order to

make up for a presumed lack in the formalistic account is surely not enough to allay the suspicion of philosophical obfuscation. Wittgenstein keeps stressing the crucial fact that every predetermined scheme has to be applied to contribute to the formation of meaning. "What is the criterion for the meaning of the formula? The way we permanently use it, how we were taught to use it." That does not preclude this use being insufficiently specified or changing over time. On the contrary: compared to the rigidity of programs, a priori understanding is distinguished by the instability of its designs. Misunderstanding has to remain possible for understanding to take place. Now such explanations might be taken to show that semantic activity does not spring from a secret inaccessible realm, but they do not clearly indicate how a slogan like "meaning is use" could be turned into a working hypothesis for a feasible semantic investigation competing with the proposals mentioned at the beginning. The later Wittgenstein was not interested in any attempt to methodically re-construct meaningful behavior along the lines he himself suggested. To see what his point of departure implies, one has to switch to hermeneutics, the systematic elucidation of nonformal processes of understanding. There is something puzzling in our usual confidence that we can comprehend the general meaning of a word even if we cannot predict the future circumstances of its employment: "It becomes strange when we are led to think that the future development has in some way to be present in the act of comprehension even though it is not. —Because we say that there is no doubt of our understanding this word when on the other hand its meaning is its use" (Wittgenstein). Hermeneutics deals with precisely this difficulty. It describes how human understanding is characterized by the peculiar relation between the intellectual abilities available at a given time, and a general, revisable structuring of the world embedding these abilities and being shaped by their contribution.

The hermeneutical process starts with fragments of utterances, texts, or behaviors that do not make sense by themselves. These are the formal elements under consideration now. But quite unlike the procedures previously discussed, they are not provided with a correlate called meaning. Instead, hermeneutics describes how they are tentatively integrated within the larger context of human expecta-

tions, predictions, and activities. Their interrelationship has already to be in place so as to direct the treatment of the isolated fragments. One has to know about puzzles to handle their pieces properly. Hubert Dreyfus has repeatedly made this Heideggerian point against attempts to formalize understanding completely. No bit of information can make sense unless it is related to the inexhaustible back-ground of being-in-the-world. To exhaust it would put an end to what we regard as human dealing with seemingly senseless symbols. This claim, it has to be admitted, faces a problem that is exactly the reverse of the formalist predicament: whereas the formalist can-not explain how sense arises out of pure syntax, the hermeneuticist must always presuppose sense. It is impossible to give an appropriate interpretation without appealing to circumstances already "soaked" with sense. And to make things worse, this presupposition leads into a kind of circle: the outcome of the interpretation of the formal elements depends on the choice of background. Heidegger's phrase, "every exposition leading to understanding must already have understood what it wants to explicate," suggests precisely what Wittgenstein found puzzling. In order to find the meaning of a phrase we have to work with prior knowledge about the way it fits in numerous situations — in fact with prior knowledge of its meaning. Needless to say, logicians are scandalized by this holistic methodological approach. What becomes of objectivity if meaning depends on such a self-stabilizing circuit? We have reached the first controversial point from a different angle: it has to be decided how the investigation of meaningful behavior shall be pursued.

There is little room for compromise between the piecemeal strategy of formal reconstruction and the holism of hermeneutical processes. The best that can be done is to be clear about the implications of the choice. It is very likely that powerful learning machines will be built on something like hermeneutical principles, absorbing information about the background as feedback to adjust their programs to the needs of the context. And on the other hand there is nothing to prevent formal semantics, strengthened by formalized pragmatics, from describing the logical structure of our discourse. But these possibilities should not be allowed to blur the fundamental distinction, that remains valid even if they are realized, between

machines that have to be used by someone and an entirely new kind of "machine" that is defined apart from any implied usage. I have already indicated that whatever those utopian devices may be called, they are not machines in any sense comprehensible to me. They could perhaps be taken as instruments of the final overthrow of the traditional conception of man. Very well, there is no guarantee that that conception will last. My point, however, is that in this event no concept accessible to us today would remain. One might by way of experiment suspend the language-game currently being played to make the most basic distinctions about man and his surrounding world, but doing so, one should not expect to be left with a conceptual scheme compatible with ours. Semantic machines in particular cannot be called perfect substitutes for man because it is entirely unclear what "man" would refer to if this proposition were accepted. It might happen that our intuitions were completely overthrown, but arguing on their basis about what would replace them is like trying to figure out the quality of heaven and hell. For centuries theologians kept busy with this task; maybe a similar role is played nowadays by science fiction accounts of artificial intelligence.

I have been assuming that we need to choose a theoretical framework (formal semantics, hermeneutics) appropriate to current pre-conceptions about the phenomenon of understanding. And I argued that understanding can be grasped as a human activity only if usage of language in addition to its formal signifying potential is considered. There is an obvious objection to this. Doesn't it amount to deciding upon what understanding is, independent of possible future evidence? The refusal to consider the effects an all-pervasive scientific revolution would have on the ordinary concept of man seems to arbitrarily privilege the language-game we now happen to have. The second major bifurcation is approaching. Holding on to an entirely empiricist perspective, there are few reasons against ingeniously experimenting to do away with the conceptual difference between man and machine. It is a healthy reminder of the unreliability of our everyday notional network that it can be thoroughly confused by appearances — animals, machines, and men looking

and acting indistinguishably from each other. Letting the matter rest here means accepting empirical investigation as ultimate result. This however takes a terribly narrow view of the capabilities and needs of intelligent beings. Questions about the legitimacy and direction of research programs frequently cannot be answered by piling up another load of empirical data. It is quite suspenseful to muse about what science will eventually discover about man, but another line of thought is at least as important, namely to investigate the conditions shaping our own empirical investigations. This cannot be done exclusively empirically, which suggests that something more than simple pragmatic scientism moves human inquiry along. Reflection on the frameworks used in science teaches something about the extension and limits of responsible research. If current basic intuitions are just one expendable set of empirical hypotheses, this is the end of the search for something distinctively human. Only by taking them more seriously can further results be obtained.

It might be helpful here to give a very pretentious version of the humanistic doctrine, pointing out its peculiarity and acknowledging resentment of its speculative character. Man differs with all the differences incorporated into the design of his theories. He has incomparable power to invent theories that shape the world according to man-made distinctions, but, because they are of his invention, such theories cannot serve fully to analyze this inventive activity. The hermeneutical insight into the inevitableness of the background is turned into a metaphysical principle by this reflection. Being-in-the-world is looked upon as an irretrievable presupposition to all the theoretical constructions man can devise in order to cope with the world's problems. I find this a tenable and convincing position, though frequently misused for sentimental preaching. By presenting the motives behind it while steering clear of its pretensions, I hope to minimize its scandal. In doing so I will abruptly turn to metaphysical considerations, disregarding the philosophy of science and epistemology, whose contributions to this subject would have to be described very carefully and could not spare us the question of principle anyway. Man as a peculiar configuration of differences with nature, and with himself insofar as he belongs to nature — is this an unfounded fantasy or a working proposal for explicating the

hidden implications of how we orient ourselves amidst the onrush of data?

Numerous objections can be made to the apparently unfounded arrogance of metaphysical thinking. As we just saw, its crucial move is to turn away from all formal and empirical exertions and open up a field of investigation, as it were, behind their backs. Meta-physics claims to discover something "deep" about the "human condition," but what are its own presuppositions? To be respectable it has to embody a particular theory itself, and then it is bound to exemplify the same limitations it seeks to demonstrate in other theories. An extremely dubious appeal to methodological superiority is sometimes made at this juncture: philosophy is considered the most fundamental enterprise of theoretical clarification. Yet it has been recognized, at least since Wittgenstein's *Tractatus*, that all such universal claims rapidly turn into paradox. Metaphysics as a theory about theories can hardly be reconciled with metaphysics as dealing with all the theories there are. Regarding it as a peculiar supertheory has been a common escape from this dilemma. This methodological impasse bears directly on our subject because only within the province of metaphysical questions can man be seen as some-thing unique. But if metaphysics leads to paradox or spurious superiority, there seems to be no reason to engage in such an enterprise. My strategy in answering this criticism will be twofold. First I will try to alleviate the misgivings one feels about the metaphysical project, and second, since its paradoxicality cannot be made to disappear altogether, I will turn around and face the fact that one has to like it. The phenomenon of human existence calls for a description incorporating conflicting elements without any obvious reconciliation of them into a harmonious totality.

Searching for the ultimate background from which all human endeavor can be derived is a borderline case of an otherwise legitimate curiosity. (Now it is the turn of the traditional argument to strain ordinary discourse.) It can often be enlightening to look beneath the surface of a given distinction. Every theory shows traces of its underlying real-life history, and Wittgenstein's remarks about the context-dependency of interpretations showed the way to develop this hermeneutical truth. Consequently it often is through in-

vestigation of what a certain theory presupposes that a clear picture of its functioning and relevance emerges. Only if some methodological hierarchy is conceded can such an undertaking make sense. It turns out, therefore, that in choosing the hermeneutical approach, we have already opted for different levels of intellectual involvement, some of them of necessity superimposed on others. In turning to metaphysics one only radicalizes this convenient scheme of thought. Theories are constructs pointing into two different directions — forward to the world they want to render intelligible and backward to their own context of origin. My initial point, that formal semantics cannot establish contact with meaning on its own, arrives at its full consequence now. A theory as such can only be said to reach out toward the world if its design serves the purpose of some-one who knows how to use it. If these considerations are assumed, there is no need for a spectacular jump into metaphysics, since they are already part of it. Metaphysics, understood as a straightforward general analysis of what it means to live as an intelligent being, is what this discussion has already initiated. No formal or empirical restriction was introduced; slipping into unconditional contemplation of the conditions of meaningful behavior seems to have been almost automatic. This tempting slip can of course be resisted, but only at the price of losing touch with the intuitions underlying our argument. (And metaphysics in some way is only systematic justification of some very basic common beliefs.) Theories are cognitive devices operating between an extratheoretical background and applications that show the world to conform with them. The theoretical status of this pronouncement, however, remains to be discussed. It sounds almost commonplace, and strangely enough, it also smells of contradiction. I characterized the humanistic claim in just this way at the beginning, and this is no coincidence. As it turns out, metaphysical method and the conviction that man is incomparable with anything else fit together perfectly.

Here is a short recapitulation of the trouble with metaphysics. It starts by dissociating itself from any other theories, searching out their most fundamental presuppositions, to which they themselves

are necessarily blind. This is the procedure indicated by the *meta-*, meaning beyond the corpus of science and the humanities. And where does metaphysics come from? Either it can be treated like an ordinary theory, in which case its transcendence collapses, or it has to be equipped with some very special set of concepts mysteriously surpassing all others. In trying to find the most general feature of being, philosophy loses its firm ground: reaching out beyond all established investigations, it is immediately thrown back into their confines, where it could not rest to begin with. Admittedly this pattern seems to call for therapy, which is exactly what many critics have proposed to offer. Nevertheless it is against those psychohygienic attempts to dissolve the tangle that a final step has to be taken. Its aim is to make intelligible the reasons for which attempts to articulate man's understanding of the world run into paradox in the end. Metaphysics is the result of not shrinking back from this conclusion even though it puts philosophy into an awkward position. Confronted with an unmediated rupture of intelligibility, it has to get a maximum of sense out of it. Not an enviable task, since all attempts to organize the clash between what can and what cannot be said are ultimately doomed to fail. But isn't that what human life basically consists of, struggling to build domains of meaning against the permanent men-ace of their unraveling? Such is the cross-connection simultaneously establishing a methodological-existential link and establishing the origin of metaphysics: it reaches back into man's struggle with chaos and lack of meaning. This is the traditional line. While upholding the claim, I want to strip it of its ideological embellishments by analyzing its structure and pointing out its essential fragility.

A widely used paradigm for the problem is the field of vision. The eye does not see itself unless a mirror is used. In some sense it can then be said to see "itself," but it cannot in principle see its seeing. These remarks accord with my previous reflections on the inevitable distinction between man and machine. If the process of seeing is not distinguished conceptually from what is seen, the language-game surrounding vision threatens to disintegrate. One of its characteristics is just such a peculiar combination of an ongoing activity and its circumscribed results. If this feature is eliminated (for example, neurophysiologically), nothing remains to distinguish vision from

any nonconscious biological process. Here the initial conflict flares up again, but for the sake of illustration only the structure of our intuition matters. The paradigm serves to illuminate the working of metaphysical method and the particularity of trying to cope with cosmic inarticulateness at the same time.

Theories lack access to their origins; in order to get a hold on them one has to use other theories, like mirrors reflecting the eye. This leaves us with the question of how any theory could possibly comprehend where theories come from. Just as wanting to see the process of seeing shakes the fundament of comprehensible talk about vision, this inquiry does not lead to the desired answer: as long as something is treated as a theory, it cannot rid itself of the problem of its background. One response to this predicament is to deny that anything important can be learned from it. Another is to turn it into an insight about man: whatever he can take hold of excludes the grip itself; it simply is not possible to fix the activity of fixing simultaneously with whatever is fixed. Or, to put it once more as a statement about our conceptual equipment, abolishing this difference is undercutting one of the elementary components of our cognitive apparatus. An undramatic interpretation of talk about human incomparability is suggested by this statement, which paraphrases the fact that we stop referring to the same thing when we relinquish this irreducible duality. And in doing so it affirms that this pattern does not fit anything but man.

Not surprisingly, we have found that the uniqueness of man can only be established within a framework that has a metaphysical bias built into it. Method and its subject matter have to be developed in tandem, as hermeneutical principles suggest. There is no essence of man outside all possible theories, waiting to offer itself when the right one comes along. Pushing certain lines of thought opens up perspectives whose range and direction can convince someone that the approach is useful. The pivotal point is deciding whether one is prepared to face the challenge of something that cannot satisfactorily be thought through. There is no possible continuum between science and the need to ask how science comes about at all, as there ultimately is no way to think successfully of man as an unproblematic mediation between nature and what is distant from nature. In

delineating such a harmony one is driven to replace the concept of man with what can be referred to as "God" or "utopian machines." In the development of such substitutions, the disturbing imbalance between formal closure and the opening up of new horizons is put to rest. Against such a quietism, the neurotic aspects of metaphysical striving can be seen as minor afflictions. Man is never safe from the impact of the incomprehensible, even though he can be quite successful in managing its consequences. This is acutely observed in Ridley Scott's film *Blade Runner* (based on a novel by Philip K. Dick), where a man-made "replicant" turns into something human by apprehending imminent death. Maintaining this general attitude can be regarded as the leading motive for the insistence that man be kept strictly separate from machines. His essence is to be able to stretch his understanding to this limit — being part of the world and, by asserting this, partially dropping out of it. A less crazy diagnosis is not feasible in this case. There is no doubt that one wrong step can annihilate this passing glory, but why should one blind oneself to it in advance? Metaphysics amounts to conceding that our condition might make sense ultimately even though we cannot say so without serious impediments. This concession leads to a stammering which yet contains a message: it is the wound that gives rise to wonder.