AMERICAN SCIENTIST, *54*, 1, 1966

AMERICAN SCIENTIST

SPRING · MARCH 1966

THE LOGIC OF THE MIND*

By J. BRONOWSKI

1

I AM honored by your invitation to speak, under the auspices of two famous societies, to this meeting which marshals the whole diversity of American science. I am honored to speak in a series which has the admirable title *Science and the other Humanities*, and I am also and rather particularly grateful. I am grateful, of course, that you have asked me here, a newcomer to the American scene who is still at the delightful stage of being bewildered by the unexpected range and the energy of your scientific interests.

Yet when I said that I am particularly grateful, I had a second reason in mind. It is this. A man of my preoccupations, that is, a practicing scientist with a passion for literature, is often asked to speak about the relation of science to literature. But usually he is constrained, by the broad nature of his subject and of his audience, to speak in very general terms. I feel no such constraint tonight. There cannot be an audience anywhere in the world more deeply devoted to science, and to the play of the human mind in experiment and invention, in logic and imagination, than this great gathering. That gives me the chance to speak more searchingly and, as it were, more professionally about my subject than usual; and I take that chance gratefully.

I gave a series of lectures early this year at the American Museum of Natural History in New York on science and literature as modes of knowledge, which were published last month as a book under the title *The Identity of Man*. There are a few places here and there in the book (I think I count four) where I should have liked to speak more fully and more circumstantially, had I been speaking to a professional audience.

* The Phi Beta Kappa-Sigma Xi Lecture, delivered to the American Association for the Advancement of Science at Berkeley, California, December 29, 1965.

I shall single out one of these places, which sketches what may be called the machinery of the mind, and make it the occasion for a larger analysis tonight. I hope to develop the others on other occasions.

The mind is an elusive entity, whose workings are not wholly confined within the brain. But because I am looking at the logical processes of the mind, it is fair that I concentrate in the first place on the brain as the organ in which these processes must be mechanized. The subject of how the mind works as a mechanism, what machinery we can imagine to operate within the brain, has its intrinsic interest in any case. We know that the brain is made of the same stuff as the rest of nature, and its atoms must therefore obey the same natural laws as other atoms do. In that sense, then, it is tempting and even reasonable to say that the brain must be some kind of a machine. But alas, to use the word *machine* in this catch-all sense misses the crux of the question. The real question about the human mind lies deeper: it asks, Is the brain a machine with a formal procedure of any kind that we can now conceive? Let me quote a pertinent passage from *The Identity of Man*.

> A machine is not merely a whirring train of gears or a humming set of electric circuits. These happy, busy strings of hardware are only the middle step, the visible link, in a procedure which has three steps, and to which the other two are as integral as this is. The machine is the procedure, and the whole procedure, all three steps of it. The first step is the instruction or input, which is the modern form of the button that starts the machine: and which must itself be precise and mechanical, an unequivocal set of holes or marks on a tape that directs the machine into one branch of its network of possible paths. Then comes the physical machinery which obediently carries out the instructions and turns them into actions. And the third step is the result or output, which is equally decisive and definite: in a computer, it is another set of holes or marks on a tape.
>
> It is of cardinal importance here, and essential to my description, that the output from a machine must be exact and unambiguous as the input is. For a modern machine, like a man, is asked in part to regulate itself, and for this purpose it must be able to feed its output back into itself as a new instruction. Its output must therefore be as sharp, within the tolerance of the machine, as capable of symbolic expression, as well defined and as single-minded as its input.

2

Our field of inquiry is the grey region between the input to the brain and the output from it; that is, between the information that the senses send to the brain, and the instructions or other decisions that issue from it. In this grey region, the brain manipulates the input and draws conclusions from it. During this process, the brain presumably uses some symbolism which translates and codifies its conceptions of the outside world. We do not know what this symbolic language is, but if it is indeed to be mechanical (on any system that we understand) its units must consist of configurations of atoms, and of changes in these configurations which are displayed as electrical signals. If then the brain reasons like a logical machine, these signs or units which it employs in its reasoning must constitute a formal language or series of languages which follow

precise rules, just like the language of symbols in which we write out logical and mathematical arguments. The brain cannot be a machine in any sense that we understand unless its language is as strict and as artificial (in the logical sense) as any of our own marks on a magnetic tape.

The symbols with which the brain works, its language (or successive languages), are physical, chemical and electrical. But this makes them no different from marks on a paper or on a tape. Provided that they are exact, and are always translated exactly in the same way, they constitute a formal logical language. What is to be said about them, then, comes not from physics and chemistry and biology, but from symbolic logic. This is why I, a mathematician, presume to talk about it to physicists and chemists and specialists in biology.

3

We know a good deal now about symbolic languages and the logical procedures that they can express which was not known when I took the Mathematical Tripos at Cambridge in 1930. There were two of us who offered what was called Mathematical Philosophy that summer: Max Black and I. The man who had lectured to us had been the prodigious, prodigal Frank Ramsey. But he had died early that year, a month before his twenty-seventh birthday, and I imagine (though I cannot be sure) that we were examined in his place by his friend Richard Braithwaite. Whoever the examiner was, he blandly asked us on one of our papers to discuss the *Entscheidungsproblem.*

The *Entscheidungsproblem,* the problem of decision, was a startling question which David Hilbert had posed: whether it is self-evident—whether indeed it can be shown—that all mathematical assertions which make sense can necessarily be proved to be either true or false. The question had gone unanswered for a long time, and neither Max Black nor I was likely to settle it at short notice in an afternoon. I no longer remember what general arguments I produced in the examination room for and against the disputed possibility. For history caught up with us and our examiner in a spectacular way, ironically within a year.

Most professional scientists now know what happened. In 1931 a young Austrian mathematician, Kurt Gödel, proved two remarkable and remarkably unwelcome theorems. The first theorem says that any logical system which is not excessively simple (that is, which at least includes ordinary arithmetic) can express true assertions which nevertheless cannot be deduced from its axioms. And the second theorem says that the axioms in such a system, with or without additional truths, cannot be shown in advance to be free from hidden contradictions. In short, a logical system which has any richness can never be complete, yet cannot be guaranteed to be consistent.

That was in 1931. In the next few years, other unpleasant theorems

were established. A. M. Turing in England and Alonzo Church in America showed that no mechanical procedure can be devised which could test every assertion in a logical system and in a finite number of steps demonstrate it to be either true or false. This is Hilbert's *Entscheidungsproblem* in its direct form. In a sense, Gödel's result is deeper than this; and Alfred Tarski in Poland proved an even deeper limitation of logic. Tarski showed that there can be no precise language which is universal; every formal language which is at least as rich as arithmetic contains meaningful sentences that cannot be asserted to be either true or false.

In order to leave no room for doubt, let me linger on the essential content of these extraordinary and far-reaching theorems. They are theorems in mathematical logic, and in one sense the mathematics cannot be removed from them. That is to say, any logical system to which they apply must include the arithmetic of whole numbers as a basic and distinguishable part. But with this proviso, to which I shall return, they apply to any system of thought which attempts to set up a basis of fundamental axioms and then to match the world by making deductions from them in an exact language—the language of physics, for example, or the chemical language inside the brain.

Such a system of axioms has always been thought to be the ideal model for which all science strives. Indeed, it could be said that theoretical science is the attempt to uncover an ultimate and comprehensive set of axioms (including mathematical rules) from which all the phenomena of the world could be shown to follow by deductive steps. But the results that I have quoted, and specifically the theorems of Gödel and of Tarski, make it evident that this ideal is hopeless. For they show that every axiomatic system of any mathematical richness is subject to severe limitations, whose incidence cannot be foreseen and yet which cannot be circumvented. In the first place, not all sensible assertions in the language of the system can be deduced (or disproved) from the axioms: no set of axioms can be complete. And in the second place, an axiomatic system can never be guaranteed to be consistent: any day, some flagrant and irreconcilable contradiction may turn up in it. An axiomatic system cannot be made to generate a description of the world which matches it fully, point for point; at some points there will be holes which cannot be filled in by deduction, and at other points two opposite deductions may turn up.

4

The implications of these results for any theory of knowledge have long been stressed, for example by Rudolf Carnap and by Karl Popper. But in addition I am stressing here, as I have done before (in *The Common Sense of Science* in 1951), their implications for empirical science. For I believe that any exact science must include in its system the axioms of arithmetic, in the form of procedures which require us to dis-

tinguish all the whole numbers. For example, if we seek to reduce all the sciences to physics, then we shall need the theory of groups and the statistics of assemblies of particles; and both these operations are subject to Gödel's theorems. In the same way, the statistical limitation on the recurrence of physical systems which Henri Poincaré first demonstrated in ergodic theory are, in my view, another expression of Turing's and Church's theorem that it is impossible to decide for every instance whether it is a consequence of the axioms. And finally, Tarski's theorem demonstrates, I think conclusively, that there cannot be a universal description of nature in a single, closed, consistent language.

I hold, therefore, that the logical theorems reach decisively into the systemization of empirical science. It follows in my view that the unwritten aim that the physical sciences have set themselves since Isaac Newton's time cannot be attained. The laws of nature cannot be formulated as an axiomatic, deductive, formal and unambiguous system which is also complete. And if at any stage in scientific discovery the laws of nature did seem to make a complete system, then we should have to conclude that we had not got them right. Nature cannot be represented in the form of what logicians now call a Turing machine—that is, a logical machine operating on a basic set of axioms by making formal deductions from them in an exact language. There is no perfect description conceivable, even in the abstract, in the form of an axiomatic and deductive system.

Of course, we suppose nevertheless that nature does obey a set of laws of her own which are precise, complete and consistent. But if this is so, then their inner formulation must be of some kind quite different from any that we know; and at present, we have no idea how to conceive it. Any description in our present formalisms must be incomplete, not because of the obduracy of nature, but because of the limitation of language as we use it. And this limitation lies not in the human fallibility of language, but on the contrary in its logical insufficiency.

This is a cardinal point: it is the language that we use in describing nature that imposes (by its arrangement of definitions and axioms) both the form and the limitations of the laws that we find. For example, it may be held that if we can remove the arithmetic from physics, we may yet get an axiomatic system which is complete and consistent. I do not share this view, but it is arguable; yet it does not seem to me to bear in fact on our present formulation of the laws of nature. On present evidence, we must conclude (in my view) that the human mind is constrained to conceive physical laws in arithmetical language: the whole numbers are literally an integral part of its conceptual apparatus. If this is so, then the mind cannot extricate the laws of nature from its own language; and we are not at all, as Leibnitz and others have thought, in a "pre-established harmony" with the language of nature.

5

Every scientific system as we understand that phrase now is incomplete: simply as a logical machine, it cannot cover all the phenomena of nature. It therefore follows, not merely in practice but in principle, that the system must be enlarged from time to time by the addition of new axioms, which cannot however be foreseen nor proved to be free from contradictions. How does the outstanding scientist come to propose such a decisive axiom, while less imaginative minds go on tinkering with the old system? How did Gregor Mendel leap to conceive the statistical axioms of genetics? What moved Albert Einstein to make the constancy of the speed of light not a consequence but an axiom in the construction of relativity?

An obvious answer is that the great mind, like the small, experiments with different alternatives, works out their consequences for some distance, and thereupon guesses (much like a chessplayer) that one move will generate richer possibilities than the rest. But this answer only shifts the question from one foot to the other. It still remains to ask how the great mind comes to guess better than another, and to make leaps that turn out to lead further and deeper than yours or mine.

We do not know; and there is no logical way in which we can know, or can formalize the pregnant decision. The step by which a new axiom is added cannot itself be mechanized. It is a free play of the mind, an invention outside the logical processes. This is the central act of imagination in science, and it is in all respects like any similar act in literature. In this respect, science and literature are alike: in both of them, the mind decides to enrich the system as it stands by an addition which is made by an unmechanical act of free choice.

As for the invention that is added, the new relation in science or the imaginative shift of vision in literature, its birth is always the same. It begins in the multiple meanings and overtones, the hidden ambiguities, which human language contains in spite of our best efforts to make it sharp. The language of thought consists for the most part of general words, and although such a word may be as matter of fact as *parallel* or as solid as *mass*, as down to earth as *table*, there is always about it a penumbra of uncertainty and ambivalence from which new relations may suddenly become apparent. *Parallel* may become the beginning for non-Euclidean geometries, and *mass* may become equivalent to energy, for the universal reason that even a *table* cannot be defined in terms which allow us to say with absolute decision of every object in the universe that it is either a table or not a table. Frank Ramsey, of whom I spoke earlier, proved that this is an indispensable factor in the development of any science; and in this important sense, he anticipated some of the implications of Gödel.

It is characteristic of human language that it is made up of past meta-

phors and analogies, and they are a fertile ground for the exploration of ambiguity and the discovery of hidden likenesses. Here begin the unexpected links and conjunctions which literature (and all art) constantly produces; and the inventive ideas of science begin here too.

As to how these ambivalences are developed in science and in literature, that is the theme of *The Identity of Man*, and I can only summarize it here. In science, the aim is to disentangle each ambiguity, and to force nature to decide between the alternatives by a critical experiment. In this way, we progress in science (as it were) by turning the information from nature through the logical machine of the brain into an effective tape instruction. In literature, the ambiguities are not resolved, and the brain works or plays with the information without ever turning it into a machine instruction. But in both, the new invention is taken by the same kind of step, and at the moment when the step is taken, we are in no logical system: we have left one system and are about to enter and form the other, and are in a no man's land outside logic.

6

The first half of my theme, which I have now completed, has consisted of theorems in mathematical logic and their application first to the language of science and then, incidentally, to literature. What I have shown there, the surprising demand that they imply for a kindred imagination in both, is unsettling, of course, and awkward, because this is not at all how we wanted the grand panorama of knowledge to look. But there it is, we must come to terms with it; and so far, I have simply displayed what the terms are, as a matter of fact.

Now I turn to the second part of my theme, to discuss a sharply different aspect of the same problem. I shall still be concerned with these maverick theorems in logic, but with something else about them: not so much with their existence and implications as with their origin. For there is a common source from which all these theorems spring, and it is uncommonly interesting and revealing.

Specifically the two theorems of Gödel, the theorems of Turing and Church, and Tarski's theorem say different things. Each of them establishes some limitation on a logical system, either on its completeness or its consistency, and these limitations are not quite the same. Yet they do form a common family of limitations, and this is because they all arise from a common difficulty in all symbolic language. The difficulty is that the language can be used to describe not only parts of the world, but also parts of the language itself.

Many logical problems grow from this common root, namely that the range of reference of any reasonably rich system necessarily includes reference to itself. This creates an endless regress, an infinite hall of mirrors of self-reflection. And the regress comes sharply to a focus in all the para-

doxes of logic, which are cousins of one sort or another to the classical contradiction that the Greeks knew: what they called the Cretan paradox. This is the contradiction implied by the statement of Epimenides the Cretan that all Cretans are liars.

There are many modern forms of this and its related paradoxes. One form is Bertrand Russell's definition of the class of all classes that are not members of themselves. Another is the paradox of Jules Richard, which (roughly) gives this a numerical dress: Gödel constructed his theorems on this pattern. Perhaps the punning, linguistic quality of these contradictions, their oddly literary playfulness, is best displayed by a remark in the same vein by Groucho Marx, who said that he would not think of belonging to a club that was willing to have him for a member. Yet these are not trivial matters: they face us whenever we contrast rules and exceptions, tolerance and intolerance, and all the human issues which join and divide us in argument at the same time.

The mathematical paradoxes, and the devices derived from them that Gödel and others exploited for their theorems, all have the same feature: they depend on the use of concepts whose range of reference includes the concept itself. In short, the model for them all is the Cretan paradox, the simple sentence, "What I am now saying is not true." This is obviously a self-contradiction: if the assertion is true, then by its own evidence it is not true; and if the assertion is false, then that tells us that what is being said must be true.

Bertrand Russell tried (with Alfred North Whitehead in the *Principia Mathematica*) to untie the knot in this kind of paradox, and to put an end to the infinite regress of assertions about assertions, by constructing a theory of types. This was intended to prevent us from using the same language to discuss our language as we use to discuss the things that the language names. A hierarchy of types was created, starting with simple sentences about things, going on to sentences about sentences about things, then to sentences about sentences which are themselves about sentences about things, and so on. No one could look on this infinite construction with anything but a suspicious eye, and so it turned out; the theory of types is an unhappy artifice. If as human beings we want to use human language, then we must accept that part of its richness is in its capacity to refer to itself.

I stress, in what I have just said, the word *human*. Animals use language to signal to one another, and what they have to say essentially refers to states of affairs (factual or emotional) and to nothing else. Such a language has no problems of self-reference: it is intended to pass information from one animal to another, directly and unequivocally as an instruction. In this sense, René Descartes was right to say that animals are machines and human beings are not. Human language is richer precisely because we think about ourselves. We cannot eliminate self-

reference from human language without thereby turning it from a genuine language of information into a machine language of instructions.

In particular, all philosophy and epistemology operates by its nature within the field where the difficulties lie, the field of self-reference. I mean by self-reference the construction of sentences, in thought or in speech, whose range of application includes that very kind of sentence. On this definition, "I am hungry" contains no self-reference, but "I am troubled" does. All thinking about thinking implies self-reference: the first statement of principle in the philosophy of Descartes, *Cogito ergo sum*, refers to itself. It is this very cogitation, or the class of cogitations that includes it, which gives the speaker the right to assert that he is cogitating. Philosophy is not possible without the regress of cogitation about cogitation. Whatever could be thought by machines, philosophy certainly could not. Indeed, on my view of human language, philosophy could not even be thought about by animals.

7

It is clear enough that statements in philosophy are, by their nature, often dogged by self-reference, and that philosophy as a discipline is therefore limited even more severely than science by the logical gaps that the theorems of Gödel and Tarski have laid bare. In mathematics and science, it is a surprise to find oneself bounded by these theorems; it is not at all obvious, and indeed is unexpected, to learn that mathematical and scientific statements cannot be wholly cleared of self-references (or of some equivalent recursive regress). But it is evident from the outset that philosophy is full of self-references, and therefore that, if the breakdown in the machinery of logic has its origin in self-reference, then philosophy is surely subject to it. Indeed it is clear that, while mathematics and science are subject to it only from time to time, when a new step has to be taken, philosophy is subject to it severely and constantly—because self-reference is built into its very method.

In the same way, we can see at once why pyschology and psychoanalysis, regarded as sciences, are most severely subject to the theorems of logical limitation. There was a time when no clear boundary was drawn between philosophy and psychology; Thomas Hobbes, John Locke and David Hume all wrote philosophy much of which was a study of the mind, and was, for its age, a form of psychology. Now that psychology has entered into less conscious fields of the mind, the logical problems that are created by self-reference are very patent. Many natural scientists complain that psychology, and other studies of human thought and behavior, lack the rigor of a true science. This is usually excused on the ground that such human studies are young, and have not yet developed the proper formal apparatus by which information can be turned into exact prediction. But I suggest that the logical theorems now show us

that this excuse is mistaken. There is an essential difficulty in casting these disciplines into an axiomatic system; they are limited, more severely and more constantly than the natural sciences, by the self-reference that underlies them everywhere. And it cannot be got out of the system by the occasional addition of a new axiom, as in the natural sciences. The axiomatic method as such may be unworkable in these studies, and whatever machinery is discovered for them in the future will (I think) not be of this traditional kind.

This is illustrated by Karl Popper's account of how he became disillusioned with the psycho-analytic explanations of Sigmund Freud and Alfred Adler. In natural science, remarked Popper, a theory is expected to make a prediction, and one prediction only, about the outcome of an experiment; and it is discarded if this forecast is not fulfilled in the experiment. But the theories of psycho-analysis are not of this kind at all; as Popper found, they are constantly explaining that my neighbor on the right is polite because he has an inferiority complex, and my neighbor on the left is rude because he has an inferiority complex. If therefore I turn the concept of the inferiority complex around, I get the unhelpful prediction that it may cause my neighbors either to be polite or to be rude. This is not what we expect of a scientific theory. And indeed it is not: all arguments derived from Freud's invention of the unconscious have this paradoxical content, precisely because their use of self-reference creates paradoxes. The Cretan who said that all Cretans are liars was talking a classical form of the language in which the psycho-analyst frames such concepts as the unconscious and its inferiority complex. If he had lived not in Crete but in Vienna, he would have said that all Viennese have inferiority complexes.

8

Beyond these borderline fields stands, full face, the particular interest to which I am drawn: the art of literature. A work of literature is in the first place a description or a story: William Wordsworth's poem *The Daffodils* is a description, and the *Oedipus Rex* of Sophocles is a story. Neither a description nor a story need contain any overt self-references. For example, the description of my interests and the story of my career which are set out in the programme of this conference are neutral accounts which do not demand that you involve yourself in them by referring any part of them to yourself. Unhappily, when that has been said about this description and this story, nothing at all has been said either about *The Daffodils* or about the *Oedipus Rex*. Yes, it is possible to have descriptions and stories whose content does not draw us into them, and in which our minds do not reflect on themselves. But as these programme entries show, such accounts simply do not have the power of Wordsworth and Sophocles. Neither, I am afraid, will they have their immortality.

From these simple examples it is at once clear that literature is literature only when it demands and commands our personal involvement. It is insistent because it insists that the bland descriptions of flowers and the remote Police Gazette records of incest and suicide concern us. They become part of us and we of them, they draw us to the human race and the human condition, and they make us one with Wordsworth on his couch and Jocasta in her bed and the plague-racked men of Thebes all over the world.

What is true of literature is true of every art. The work of art is a constructed thing, and is so even when it happens to have been found in nature in the form into which we now read a human meaning. It has been made in essence, its meaning has been created, by a human being: it expresses his vision of the relation between man and nature, and it invites us not to like it or to dislike it but to be drawn into it. The work of art compels us (when it is compelling) to look at the world with it, and to look through it into the mind of its maker. We cannot dissociate the work from its origin, which is to be a made thing—a thing made by a man which expresses how the man sees himself in the world. It interests us only as it engages us, and asks us to see ourselves in the same world also. Although what is expressed in the work is another man's self, the reference is to ourselves because the reference is universally to the human self.

Let me be explicit in my meaning here. I am not merely remarking that there is self-reference in the moral reflections of the Greek chorus, or in the reflections,

"In vacant or in pensive mood"

that fill Wordsworth's inward eye in solitude. These are only of the same kind as the reflections of Descartes in philosophy and of an analyst interpreting dreams. But the self-references of literature, and of art in general, go deeper than these formal thoughts. My argument is that literature is composed essentially of self-reference, and takes its life from the dual tension between watching our own minds from the inside and watching someone else's from the outside. And this is one of the classical paradoxes in the theory of knowledge, how and when we know that others do indeed feel as we do, which Ludwig Wittgenstein for example discussed in the *Blue Book* long before I discussed it in *The Identity of Man*.

The force and meaning of literature is to present the lives of others to us in such a way that we recognize ourselves in them, and live them from the outside and from the inside together. We do not understand Wordsworth unless our heart also turns over at the golden host, and the tragedy of Oedipus differs from the gunplay in the Sunday paper only if we recognize ourselves in the characters. We have to see that Oedipus is us, capable of killing a stranger at the crossroads and blundering into a

labyrinth of horror. We have to see that Jocasta is us, longing for the lost youth who so transparently is a part of herself in both senses: the son who is also the symbol of her own youth, that she longs to recapture and sense again in her leaping womb. And when we recognize that in Jocasta and in ourselves, it is more tender, more heart-breaking, more deeply human than the explanations of psycho-analysis. Of course Freud was right about the Oedipus complex; but Sophocles wakes deeper echoes than Freud, because he brings home to us the longing of Jocasta for herself—the self that she was and the self that she gave birth to—in the same hushed breath with the familiar and family jealousy of Oedipus.

Literature and art live by, they come alive in, the sense of our own self stretching into the actions and disasters of someone else's self, and thereby mapping the human self as a whole. This is how I put this part of my theme in *The Identity of Man:*

> I hold that each man has a self, and enlarges his self by his experiences. That is, he learns from experience: from the experiences of others as well as his own, and from their inner experiences as well as their outer. But he can learn from their inner experience only by entering it, and that is not done merely by reading a written record of it. We must have the gift to identify ourselves with other men, to relive their experience and to feel its conflicts as our own. And the conflicts are the essence of the experience. We gain knowledge of ourselves by identifying ourselves with others, but that is not enough—that only gives us the fantasies of sex and the parodies of power; the absurd strutting daydreams of Secret Agent 007 and *Butterfield 8*. We must enter others in order to share their conflicts, and they must be shown to have grave conflicts, in order that we shall feel in their lives what we know in our own: the human dilemma. The knowledge of self cannot be formalized because it cannot be closed, even provisionally; it is perpetually open, because the dilemma is perpetually unresolved.

9

Let me recapitulate the steps in my argument. I treated my theme in two parts: the first was concerned in the main with science, and the second with literature. In both parts, I was at pains to show that the brain as a machine is certainly not the kind of machine that we understand now. It is not a logical machine, because no logical machine can reach out of the difficulties and paradoxes created by self-reference. The logic of the mind differs from formal logic in its ability to overcome and indeed to exploit the ambivalences of self-reference, so that they become the instruments of imagination.

In the first half of my theme, I explained the limitations (they derive from self-reference) which circumscribe any axiomatic and deductive system of a reasonable richness, in mathematics and (I hold) in the natural sciences. The logical theorems which I quoted and explained show that this must be so, and they also show how these logical gaps have to be filled, and new theorems incorporated as added axioms in a system, at each step. The decision to take new matter into our systems, in science or in literature, has no analog in any logical machine. It is an imagina-

tive step of a kind that we do not understand but that we can watch in the work of a great scientist or a great writer; and it is alike in science and in literature.

The second part of my theme goes further. Here I pointed out that human language, when it is specifically human, and is concerned with reflection and judgment about our own lives, is necessarily full of self-references. This is clear in philosophy and in psychology. But it reaches deeper in literature, because the essence of literature (and of all art) lies in the identification of ourselves with other human beings whose actions we are watching and judging as if they were our own. Here the self-reference is so integral that we cannot construct any of the provisional systems with which mathematics and science make do for a time, and which they then mend when the need arises.

In literature, there is no provisional description which can take the place of the work itself. We cannot replace it, as in science, by an axiomatic system which will do until it turns out to fall short and has to be enlarged. The references in literature by the writer to himself and others, and by the reader to himself in what he reads, penetrate the work through and through; and there is no way of getting round Gödel's theorems and Tarski's theorem and the others by any step-by-step procedure. In this respect, science and literature are different.

Neither science nor literature ever gives a complete account of nature or of life. In both of them, the progress from the present account to the next account is made by the exploration of the ambiguities in the language that we use at this moment. In science, these ambiguities are resolved for the time being, and a system without ambiguity is built up provisionally, until it is shown to fall short. This is why the results of science at any given moment can be presented on an axiomatic and deductive machine, although nature as a whole can never be so presented because no such machine can be complete. Whatever kind of machine nature is, it is different from this.

But in literature, the ambiguities cannot be resolved even for the time being, and no provisional system of axioms can be set up to describe the human situation as the writer and the reader seek to see it together. Here the brain cannot act as a logical machine even for the time being: by which I mean, that it cannot take in the information, sort out its ambiguities, and turn it into unambiguous instructions. That is not what a work of art does to us, and we cannot derive such instructions from it. I will quote at the last the passage from *The Identity of Man* of which, as I promised at the beginning, this essay is a detailed exposition. It states for the machinery of the brain the same limitations that I have exhibited tonight in its account of the machinery of nature.

> I am asserting that there is a mode of knowledge which cannot be spelled out formally to direct a machine. It may be asked, Any machine? If this is a question in the present, then the answer is Yes. For example, we know (from the work of Kurt Gödel and A. M. Turing) that no machine that uses strict logic can examine its own instructions and prove them consistent. But if it is a question about the measureless future, then it cannot be answered. A machine is not a natural object; it is a human artifact which mimics and exploits our own understanding of nature; and we cannot foresee how radically we may come to change that understanding. We cannot foresee and we cannot conceive all possible machines—if indeed the word *all* has a meaning in this sentence. All that we can say, and all that I can assert, is that we cannot now conceive any kind of law or machine which could formalize the total modes of human knowledge.

There is however one respect in which my exposition now goes radically beyond this passage, not merely in detail but in substance. That is in tracing the common quality of imagination in science and in literature to the logic of self-reference; and in showing that, within this common quality, the difference of mode between science and literature reflects the different extent to which self-reference enters their languages.

BIBLIOGRAPHY

Braithwaite, R. B., *Scientific Explanation*, (Cambridge 1953).

Bronowski, J., *The Identity of Man*, (New York 1965).

Carnap, Rudolf, *The Logical Syntax of Language*, (New York 1937).

Church, Alonzo, "A note on the Entscheidungsproblem," *Journal of Symbolic Logic*, Vol. 1 (1936), pp. 40–41, and 101–102.

Gödel, Kurt, "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I.," *Monatshefte für Mathematik und Physik*, Vol. 38 (1931), pp. 173–198.

Hilbert, David, and P. Bernays, *Grundlagen der Mathematik*, (Berlin 1934–39).

Kleene, S. C., "Recursive predicates and quantifiers," *Transactions of the American Mathematical Society*, Vol. 53 (1943), pp. 41–73.

Myhill, John, "Some Philosophical Implications of Mathematical Logic," *Review of Metaphysics*, Vol. 6 (1952), pp. 165–198.

Nagel, Ernest, and James R. Newman, *Gödel's Proof*, (New York 1958).

Poincaré, Henri, *Calcul des Probabilités*, (Paris 1912).

Popper, Karl R., *Conjectures and Refutations*, (London 1961).

Ramsey, F. P., *The Foundations of Mathematics*, (London 1931).

Richard, Jules, "Les principes des mathématiques et le problème des ensembles," *Revue générale des science pures et appliquées*, Vol. 16 (1905), pp. 541–543.

Riesz, Frédéric, "Sur la théorie ergodique," *Commentarii Mathematici Helvetici*, Vol. 17 (1944–45), pp. 221–239.

Tarski, Alfred, *Logic, Semantics, Metamathematics*, (Oxford 1956).

Turing, A. M., "On Computable Numbers with an Application to the Entscheidungsproblem," *Proceedings of the London Mathematical Society*, Series 2, Vol. 42 (1936), pp. 230–265, and Vol. 43 (1937), pp. 544–546.

Whitehead, A. N., and Bertrand Russell, *Principia Mathematica*, (Cambridge 1910–13).

Wittgenstein, Ludwig, *The Blue and Brown Books*, (Oxford 1958).