Inductive inference, coding, perception, and language

H B Barlow
Psychological Laboratory, Cambridge, Cambridge CB2 3EB, England
Received 31 July 1974

Abstract. The sensory system of animals, the words and language that connect the thoughts of one individual with those of another, and man-made communication systems, are all greatly influenced by the way information is coded. It is suggested that the role of inductive reasoning is to improve the efficiency of linguistic communication by changing the code, and its effectiveness in performing this function explains its survival value—the reason why this habit of mind is biologically successful and persists. The principles of coding are best understood in a limited, well-defined situation, but inductive reasoning should be viewed in a broad context, which must include the physiological and psychological mechanisms that form perceptions from physical stimuli and attach words to perceptions.

Introduction
Consider the following three situations. First, a bird approaching the branch of a tree whose brain must control its muscular movements in such a way that it executes a safe landing. Second, the owner of a newly-commissioned radio telescope who must base a satisfactory number of publications on the results obtained with it. Third, the spare-parts manager of a remote automobile agency who, for reasons we need not divulge, is in constant danger of overspending the sum allowed in his budget for telegraphing orders to the head office.

The situations appear quite different, but the feature they have in common is the necessity of converting a voluminous amount of raw data into a comparatively small but highly appropriate form. The bird's problem involves the physiological mechanisms of sensation, perception, and motor control. The radio-astronomer provides us with a model of scientific inference, in which inductive reasoning is the major component. The parts manager has a simple practical problem: he would have to learn to code part numbers in order to use his telegraphic budget most efficiently.

These three situations might be taken as paradigms of perception, induction, and coding, but it is in the latter, relatively simple, well-defined case that the principles involved can be most clearly understood. My aim here is to show that there is quite a close analogy between these principles and the rules of inductive inference. Furthermore, once the analogy is seen, induction appears as the linguistic part of a process which extends back to the nervous mechanisms that form perceptions from sensory stimuli, in other words, to the situation confronting the bird.

Perception and coding
It may be asked why I, as a physiologist, have allowed myself to be led into the apparently remote territory of induction. It is because the problem of sorting and arranging sensory messages is a central one in sensory physiology, because ideas from communication theory seem to me to contribute some insight into their nature, and because this insight also seems to extend to the problem of induction. Helmholtz talked of 'unconscious induction', meaning the process whereby our simplest conscious perceptions already contain elements dependent upon a process similar in principle to inductive inference. This was connected with his firm adherence to an empiricist, rather than nativist or intuitionist, view of the factors behind sensation.
Nowadays the extreme empiricist view is no longer tenable; recent advances in sensory and developmental physiology have shown that some sensory neurons only respond to a highly specific pattern of excitation, and the specific connections that endow such neurons with this selective capacity are apparently established under genetic control. On the other hand in the cerebral cortex they are also subject to environmental modification, and their selectivity seems in some cases to be adapted to the pattern of excitation they have received. I have recently given my views on the current situation (Barlow, 1972) and the experimental advances now being made will certainly necessitate further reviews in the near future. Meanwhile, there can be no reasonable doubt that perception is to some extent dependent upon the past history of sensory stimulation; our inherited brains somehow weld together current instance and previous experience to give us all that we know about the outside world, and the way in which this is done must be related to induction. It is true that processes occurring at rather early levels in the nervous system are the ones that concern present-day physiologists, whereas the philosopher’s analytic tool only allows him to discuss the fully conscious parts of induction; thus the philosophical part of the problem deals with linguistic usage and that is why words and language must be brought in, but it is surely worthwhile looking at the end product in the light of one’s knowledge of the raw materials and process of manufacture.

It is surprising how little has been written along the lines of this essay since related ideas were suggested by Helmholtz (1877; 1896), Mach (1886), and Pearson (1892). Possibly their materialistic explanations of mental propensities did not suit the age in which they were advanced; possibly Russell, Whitehead, Wittgenstein, and the Vienna school provided distinct alternative notions to focus upon; or possibly developed ideas on the logic of scientific inference (Popper, 1934; 1972), statistical inference (Fisher, 1935) and probability (Reichenbach, 1938; 1949) were necessary to clarify the notion. Kenneth Craik (1943), who was already thinking along cybernetic lines, came close to this theme, and Attneave (1954) and I (1959; 1961a; 1961b; 1969) have written upon it more recently. Certainly the analogy is in the cybernetic tradition, for Norbert Wiener (1948) was never timid in the pursuit of cybernetic ideas beyond the borders of his own field of special knowledge into the domains of physiology, sociology, and psychology. In such forays one rarely escapes all the hazards of the unfamiliar territory, and my knowledge of current thinking on the philosophical problems of induction is certainly incomplete. From Swinburne’s (1974) valuable compendium I judge that the views expressed here fit well with those of Wesley Salmon (1964), who in turn builds on Reichenbach’s probabilistic formulation, and suggested solution, to the problem of induction. I think the arguments advanced here explain very clearly the advantage Salmon seeks for inductive reasoning. Our sensory and perceptual mechanisms have to encode as much as possible of the vast amount of information impinging on our senses, and induction is the culturally visible, linguistic, continuation of this process. Anything that improves the efficiency of these operations has survival value, and thus we may expect our genes to endow us with an intuitive aptitude for inductive reason, and we may hope that our culture will nurture this aptitude. Such, at least, is the insight that I think the cybernetics of coding gives about induction, perception, and language.

The analogy
It is suggested that inductive inference is like modifying a code so that an ensemble of messages can be transmitted down a communication channel more efficiently. The reader must refer elsewhere for a more detailed discussion of what can and cannot be achieved by coding (e.g. Shannon and Weaver, 1949; Woodward, 1953) but the three simple facts that are essential for the present discussion are as follows.
First, a mass of data such as the sensory signals of the alighting bird, the output of the detectors in a radio telescope, or the list of stock numbers required by the automobile-parts manager can be represented completely in quite different forms from those in which they were originally presented or received. Second, these alternatives may differ greatly in economy of representation, even if they are exactly equivalent in the sense that the original can be reconstructed completely and unambiguously from the alternative representation. And third, the degree of economy in any representation is inversely related to the amount of regularity it contains; any regular or predictable feature of a representation reduces its economy, and it is worth noting that the regularity need only be approximate and statistical (A is about twice as probable as B) rather than exact (two A's occur for every B). The kind of alternatives that are contemplated are the representation of the output from the radio telescope in the temporal frequency domain rather than as a sequence of values at successive moments in time, or the conversion of the impulse frequencies in the alighting bird’s optic nerve fibres into estimated positions and future positions of the branch it is approaching. It is obvious in these examples that there are important considerations besides those of economy. For instance, how accurately do the impulse frequencies represent the positions? Is the representation as impulse frequencies conveniently arranged for making the corrective muscular movement required in the current situation? The analogy is not claimed to extend to such details, important as they may be; it is thought to apply to the overall goals of induction and perception, and for this we need to look at the economy of alternative representations. This will be clearer and simpler in the final example.

A typical stock number which the parts manager has to telegraph might be a sequence of a dozen or so letters, numbers, and punctuation marks: for instance ‘B18: 73B/10246, C’. One can immediately see that this is uneconomical, for there are nearly 100 keyboard characters so that 15 of them represents a selection from some $10^{30}$ alternatives, which certainly exceeds the number of stock items by many orders of magnitude. A very simple step that might be taken would be to note the 26 most frequently ordered parts and assign to each a single letter; to begin with, less frequently ordered parts could continue to be represented as before. Of course there would be the interesting problem of persuading the head office to accept the code, and the most frequently ordered parts might suddenly change when, for instance, local fishermen discover the ready removability of wheel nuts and how excellently they perform as weights for their nets. But these are side issues that should not obscure the main point: there were regularities in employment of the original stock numbers because some were never used at all, while others recurred frequently. Once a regularity has been detected, a more economical code can be developed and the pressure on the telegraphic budget eased.

The 'regularities' referred to above are forms of redundancy in the informational sense, and can be expressed quantitatively. But the technical term, though correct, is liable to be misunderstood. Also it may be a bit misleading to use it in situations where all the factors required to express redundancy quantitatively are unknown, or even unknowable. The term 'regularity' is perhaps preferable for qualitative discussions.

Coding can achieve economy of representation, but is this what induction does? It is surely inconceivable that an inference could be lengthier than the data upon which it is based, and the radio astronomer writing his papers surely knows well that one aspect, at least, of what he is doing is to tell other scientists as briefly as possible what they will find if they repeat his observations. It also seems a fair account of the inductive inference employed in everyday conversation. We say “The donkey trod on Elwin's toe”, and do not tell our friends all the detailed
observations that lead us to this conclusion, unless they are of a particularly entertaining nature. Let us therefore compare the means of achieving economy in coding and induction.

Use of statistics
The evidence required to devise a code consists of counts of the frequency of occurrence of messages or sections of messages. These counts enable one to estimate the quantity of information (\(-p \log p\)) associated with the receipt of each message, and if the specifications of the channel are known one can also find the cost of transmitting the messages through it. If too much is being paid for the information received, then a new code can be devised better suited to the statistical properties of the ensemble of messages being handled. Of course, codes are also used to protect messages from loss or distortion by noise, but this aspect will not be considered here and a noiseless system is assumed; there is then no loss of information and the messages can, if required, be recoded into the original form after receipt.

Little argument is needed about induction also depending upon the determination of statistical regularities from counts of observations. The assertion ‘tenderness in the right lower abdominal quadrant together with a moderate elevation of temperature signifies appendicitis’ depends upon observing that these features of an illness are frequently associated with certain subsequent pathological findings. It is true that statements of this type are often made on the basis of an individual’s remembered experiences, without any detailed statistical documentation, and it must be admitted that such documentation is unimportant if the association is sufficiently obvious without it. However, the essential feature of a statistical statement is present, even if no figures are compiled, for one is asserting that this type of pain occurs with this pathology more often than it would by chance: a null hypothesis has been suggested and found wanting.

Consider now a very simple example of coding. Suppose that, as a result of counting letter frequencies and digram frequencies, it has been found that in written English "q"s are very nearly always followed by "u"s. To exploit this the rule “where "u" follows "q", delete it” is inserted in the codebook. An invariant association has been found, and this enables one member to stand for both. The messages are shortened, but no information has been lost, because to form the original text again all that need be done is the insertion of a "u" after every "q". Note that if the association is not invariant one can say “if "u" does not follow "q", insert it”, and then restore the original by a deletion rule. One still shortens the messages on average because "qu" is common and reduces to "q", whereas "qX" is only rarely lengthened to "quX".

The inductive inference drawn from the association of tenderness, fever, and a certain pathological report is that appendicitis causes the pain and temperature. We no longer describe at length instances where the three occur together: we simply label them ‘appendicitis’. There appears to be a real similarity between allowing the pathologist’s report to describe all three observations, and allowing the "q" to stand for the pair "qu".

More sophisticated examples
Consider next the assertion that the equation \( s = \frac{1}{2}gr^2 \) represents the relation between position and time when an object falls freely. It is true that regression lines, homogeneity tests, and confidence limits may have to be used to see if the paired observations fit the equation, but these statistical tests cannot be regarded as crucially important for the discovery, because they were quite unknown in Galileo’s day. Hence it may be suggested that the analogy breaks down, because induction is characteristically a matter of discovering subtle or deep relations without statistical tests, whereas coding requires no more than the compiling of elementary statistics.
Before discussing the alleged simplicity of coding, one should recall that induction has in the past been regarded as a simple matter. According to Cohen and Nagel (1934), Bacon thought he had reduced the process to a mechanical technique that left little to the "sharpness and strength of men's wits", and J. S. Mill tried to elaborate this and to make it more precise, while sticking to the notion that it was a process which could be put in motion by anybody once it had been properly described. However, they were only able to make this sound at all plausible by taking examples of inductive inference in which the hypotheses formed and tested were of a very simple and trivial nature.

The formation of hypotheses has a close counterpart in the coding situation, and our examples made coding appear too simple in the same sort of way as Bacon and Mill made induction appear too simple. This was deliberate, in order to illustrate the principle, but codes are not necessarily simple, and greater economies may be achieved by using more complex ones. Furthermore, as soon as one starts to contemplate different codes the 'elementary' statistics one thinks of compiling change. For instance the parts manager might suspect that a group of items were always ordered together, and could be represented by a single letter in a code, but this would require a more complicated search of his records than a simple compilation of the frequency of ordering each separate part. Again one can think of statistics that would be useless, but are as elementary as letter digram frequencies: for instance the distributions of the separations of successive occurrences of the same letter could be determined, but the properties discovered would have little value because they would not be simply related to the cost of transmission in the types of code normally under consideration. Thus it will be seen that 'compiling statistics' is far too vague a term to describe what is required in devising or improving a code. One needs ideas as to the nature of the redundant feature and the code that would exploit it; only then can the messages be classified and the entries counted in a useful way.

There is never any guarantee that a correct scientific concept will be proposed and tested, nor can one ever be sure that the best code will be found. Suppose, for instance, that the messages consist solely of lines from the sonnets of Shakespeare: this might perhaps be noticed by a telegraphic clerk, but it would not be easy to catch this fact in a routine statistical test. If it was discovered, however, each line—a message of some 50 alphabetic characters and punctuations—would require only a four digit number for unambiguous selection.

These considerations show that code selection is not necessarily a simpler or more mechanical operation than inductive inference. Messages can be reduced to a more compact form because they are not random; similarly the regularities found in observations or experimental results provide grist for the inductive mill. In both cases the mechanical application of statistical procedures will test for nonrandomness, but something else is needed to advance knowledge or achieve economy, and this is the imaginative formulation of testable theories and concepts about the underlying cause of the phenomena. Forming hypotheses is quite different from testing them, but this step occurs in both coding and induction, and is considered next.

**Formulating null hypotheses and devising codes**

Null hypotheses are to the evolution of knowledge what individuals are to the evolution of species: hypotheses and individuals can prove they are the fittest only by surviving while others perish. There is astonishing unanimity amongst statisticians (Fisher, 1935; Neyman, 1957) as to the key importance of null hypotheses in inductive inference or behaviour, and in developing the analogy our interest shifts to this aspect.
When making observations one always has something much more specific in mind than the very general idea of 'performing induction': one is thinking, "Do these facts fit this hypothesis?" As pointed out by Bronowski (1953), the code that is in use or under consideration is a good analogy for the null hypothesis that is being subjected to test. First consider the way in which it guides one in the classification of messages preliminary to counting them. No two messages are strictly identical, for if they differ in no other way they certainly occur at different instants in time. The code guides one in deciding upon the scheme of classification to be used; as suggested already, it may show that it is preferable to count words rather than letters, or that the separation of pairs of letters is not a helpful classificatory principle, or that certain parts are invariably ordered together. In the same way a null hypothesis, if it is fully stated, includes assumptions about the parent population and the way the sample has been drawn from it, and, as with the code, these assumptions are a guide to the classification of the observations. For instance it is very commonly assumed that the observations are samples from a population that does not change in time, which means that events that differ only in their time of occurrence may be placed in the same class. Without assumptions of this type, no two observations could be considered alike, and hence counts other than one or zero could not be obtained.

The analogy continues to hold when one looks at the conclusions that can be drawn as a result of statistical tests. As Popper showed, a hypothesis can be shown to be unacceptable, but it can never be proved correct, for one can never prove there are no other hypotheses that fit the facts equally well. Likewise, it is doubtful if one can ever prove that a particular code is optimum. The most favourable outcome for a code is failure to show that it is not optimum; for a hypothesis, escaping disproof.

**Special features of induction**

So far we have pointed out that codes and induction allow information to be conveyed more compactly, that they both depend upon statistical tests, and that code and hypothesis perform similar functions in enabling these tests to be done. In these ways the analogy holds up well, but the fact that inductions are incomplete representations of data, and the fact that they enable predictions to be made, may appear to distinguish them from the usual type of code selection; these points will now be considered.

**Complete and incomplete communication**

When selecting a redundancy reducing code one normally has to accept all the messages presented for transmission. A moment's consideration of inductive behaviour shows that it is very different: a scientific paper does not contain a description of every experiment performed, nor do inductive generalisations normally enable the specific observations upon which they are based to be reconstructed completely. In other words induction involves selective retention and rejection of information, whereas coding does not. At first this appears a genuine and important flaw in the analogy, so let us look carefully at the way that the selection is performed in induction.

Some of the grounds for omission fit the analogy well. We do not say that the man we met in the street had two legs, since it would be redundant to do so: we suppress the information because it is taken for granted that men have two legs, in much the same way as a code might suppress the u's that follow q's. Again, things that have been said once do not need repeating, so we do not include all our 'typical' experimental results in a scientific paper. The difficulty does not arise over the omissions that make what we communicate less redundant—this type obviously fits the analogy perfectly. But we also omit results that we attribute to chance variation
and, it must be admitted, we sometimes omit results that we cannot account for at all. The matter of distinguishing the orderly from the unaccountable—picking out the signal from the noise—is certainly a legitimate and important part of induction; it would be disappointing if the analogy could not be modified to include a counterpart.

Let us suppose that the channel down which we are asked to transmit messages is of limited capacity. Even after devising an optimal code we may find that the recoded messages still require greater capacity than is available, and the only possible thing to do is to reject some of the messages presented to us. Now ask this question: "Is there any criterion for selecting messages that will minimize the number that have to be rejected, and maximize the proportion that can be successfully transmitted?"

First suppose that by some random or arbitrary means we reduce the number of messages to a level that can be successfully handled; on looking only at those accepted, some statistical regularities are likely to be found, a code can be devised, and there will then be channel space to spare. As a direct consequence of coding, more of the available messages can be accepted, but this advantage can be put to better use than appears at first sight. Knowledge of the statistical regularities in the messages originally accepted was used to devise the new code, and these messages were drawn from a larger population of available messages. If the same regularities hold for this larger population we can use our new knowledge to select positively in favour of messages containing much redundancy, and therefore capable of being much compressed, while rejecting those messages which appear random because we have found no orderliness, and hence cannot compress them. Knowledge of redundancy, as well as enabling one to avoid wasteful utilization of the communication channel, also enables one to select messages that can be effectively compressed by the codes that have been developed. This is clarified by considering the spare-parts manager; when his budget runs low he will only order the 26 commonest parts with their single letter codes, since this will enable him to order more and satisfy more customers. Thus knowledge of regularity gives a double advantage when trying to transmit as high a proportion of the available messages as possible.

It is now clear how the selection criteria that troubled us in inductive behaviour find their counterpart in coding. If too many messages are being received to pass down the channel even after optimum coding, then one must select the redundant ones in order to pass on as many as possible. It is remarkable to find that redundancy is the clue to selection as well as the factor allowing compression; the redundancy of 'signals' is the means of distinguishing them from 'noise' even when one has never been told what signals to expect, nor the characteristics of the disturbing noise. This is the core of the problem of detecting signals from extraterrestrial aliens, and it is also a problem that occurs in many pattern-recognition situations, or in listening to a new piece of music for the first time. By incorporating the selective element of induction the analogy has been significantly changed, but has also been strengthened.

**Compression and prediction**

The second difficulty concerns the importance of prediction in induction. Ernst Mach (1886) and Karl Pearson (1892) held that the merit of a concept or hypothesis lay solely in its ability to achieve 'Economy of Thought'. This idea is closely related to the analogy discussed here, but many people have felt intuitively that it underrates the value of a scientific law, mainly because it does not explain its predictive value. Once the association of abdominal pain, moderate fever, and appendicitis has been recognized, we can use the first two to predict what the pathologist will report, and to ensure that he collects his specimen from the surgeon's table not the autopsy room. Similarly, $s = \frac{1}{2}gt^2$ is not only a compact expression of much observational material: it is also a
tool enabling us to make true statements that we cannot make from the observations alone. The formula expresses what can be generalized, as well as telling us what the generalization is. It picks out from the mass of observational details those aspects which are independent of particular circumstances, thereby saying which aspect will continue to hold in the future. It will be said that it is the ability to predict which is important, and that compactness—the sole aim of coding—is a trivial aspect of induction.

For a partial answer to this, notice that coding can, sometimes, lead to prediction. Thus if one deletes the $u$'s that invariably follow $q$'s, one can reinsert the $u$ immediately after a $q$ is received, even before the $u$ has been transmitted; likewise the parts manager who spots that two items are ordered together can predict the need for one when he receives a request for the other. The regularities that are made use of in coding may be of this type, where the early features establish the existence of the pattern and thus enable a later feature to be predicted, but it does not necessarily happen this way. There are situations in which codes that make use of sequential relations involve additional delays; for instance this would be the case if the redundant $u$ preceded the $q$ instead of following it, and Shannon's well known theorem proving the possibility of optimum coding only holds where delays are allowed. Clearly these delays are liable to convert a prediction into a useless post-diction.

I think these difficulties with the Mach-Pearson doctrine on Economy of Thought can be resolved by making two minor amendments. The first springs from the incompleteness of communication in the inductive situation that was discussed above. For a concept to be a good one it is not enough that it represents a range of physical events with maximum economy: it must also represent the greatest possible range of physical events. This is really implied, though it is not emphasized, in their original statements, for in considering economy you obviously have to take into account what you get for your money as well as what you pay. The second amendment is to admit the force of the objection about prediction. Not all forms of redundancy reduction are equally valuable even if they lead to the same economy. In the simplest case, it is more valuable to find the high conditional probability of a consequent than of an antecedent event.

The analogy has given a glimpse of insight into two special features of induction. It is like finding a redundancy-reducing code, but codes with predictive power and those that do not involve long delays are preferred. There is also the important point that, whereas coding normally aims at complete transmission at reduced cost, the object of induction is to capture and communicate as much as possible, realizing that it is impossible to catch all.

The biological role of induction in changing language.

So far the coding analogy has been applied to 'induction' as this is usually defined and written about, namely as a procedure that can be adequately described and discussed in words such as causality, simple enumeration, or arguing from the particular to the general. The procedures can be more formally defined, symbolized, and manipulated in formulations of the basis of probability, but this is still a matter of marks on paper which are passed around and inspected by others interested in the problem. I cannot of course avoid the same restriction in what I write here, but the coding analogy suggests that we should view these marks on paper as communication symbols comparable in some respects with the representation of a stock number by the electrical signals in a telegraph wire, or the position of the branch of a tree by optic nerve impulse frequencies. If the analogy is appropriate, then the implication is as follows: as a result of his lesson in coding, the parts manager changes the relation between the telegraphed signals and the stock numbers; our lesson therefore tells us that induction changes the relation between the symbols of our discourse and the
deeper levels of understanding that words, language, and marks on paper arise from and address. The next task is to discuss these changes, but first something must be said about perceptions, which are so intimately connected with words and language.

*Perceptions and coding*

From the physiological point of view the main question about perception is how the scene perceived is represented by activity among the myriads of nerve cells in the parts of the brain that handle sensory messages. Is it, as I have argued (Barlow, 1972), by a small number of active cells, each representing a specific and quite detailed aspect of the sensory scene, and each selected from a vast population of other potentially active cells, every one with its own equally specific 'meaning'? Or is information represented as complex combinations of activity among groups of cells, more like the representation of an instruction or number in a computer? The physiological question is not necessarily important here, but questions of the organization and selection of the sensory information represented in perception certainly are, for words are linked to the physical stimuli which they describe and represent by way of perception. For this reason it is important to refer to, though I shall not repeat, the argument that perceptions are related to the physical stimuli by a code which reduces redundancy (Attneave, 1954; Barlow, 1959; 1969; 1972), and which also selects for transmission those regular features it is able to code nonredundantly (Barlow, 1974). The coding analogy, in other words, applies to the formation of perceptions from sensory stimuli as well as to induction, as I have argued here. What this means for the forthcoming discussion is that perceptions are not like the uncoded stock numbers that the parts manager was faced with; it is as if his predecessors had already been at work, and all he can do is to improve the code to meet current conditions or incorporate new knowledge. With this in mind, let us see if the analogy gives any insight into words and language.

*Words as nonrandom associations*

In the hackneyed example of coding, the *u*’s that follow *q*’s were deleted. In doing this the single entity ‘*q*’ was made to stand for the double entity ‘*qu*’, and this was done because *q* and *u* were associated in a nonrandom way. The essential step was to recognise this nonrandom association, and to use a single symbol to represent it. Now nonrandom associations that present themselves to our senses very often correspond to simple names and words. Four legs and fur become ‘cat’; claws and pain are the precursors of ‘fear’: breast and milk, hunger, satiation and comfort, become ‘mother’ with all the abstract and concrete overtones. Verbs represent frequently occurring actions and relations, as when mother feeds baby, or the cat frightens it. It may be trivial to point out how words thus crystallize recurrent features and events, but it is an easily overlooked property that makes us view languages differently. Clearly they are long-evolved instruments incorporating the inductions of many ages; they are already well-developed codes for communicating between individuals.

Now the analogy says that induction corresponds to the choice of an improved code: is it true, then, to say that induction means changing the language?

*Growth of language*

It is a remarkable fact that language can both change, growing to incorporate new knowledge, and yet retain virtually intact the meaning of an earlier literature. It is worth casting a glance at this problem from the viewpoint of the analogy, for the redundancy of language acquires a new significance when viewed from this angle.

Three types of growth can be distinguished, the first a continuous process to which we all constantly contribute, the others more violent and disruptive processes
which make spasmodic additions to the language. The mere choice and utilization of a word in a new or unusual context changes its meaning, for words are gregarious creatures and in our minds each one is hung about with habitual associates. Changing the membership of these verbal tribes, or forming new associations between them, changes the language. This normal and respectable process of growth has taboos to be heeded, and where new knowledge is accumulating fast these are broken and less reputable customs develop. One is the use of jargon; here a word with its own established position in the linguistic community forms a liaison by appearing in a new context, but instead of altering the membership and links of its original tribe, it breaks away and forms an isolated community. Another process is the creation of neologisms, entirely new words, or bastards of an irregular union.

Both these processes are liable to excite violent opposition, but to the popular mind nothing proves so clearly that a 'breakthrough' has occurred as the creation of a new scientific term; this popular reaction is justified and provides fuel for the conflict between those anxious to coin and distribute their own currency of words, and the linguistic conservatives whose respect for established usage and old associations leads them to detest all innovation. Jargon and neologism are only exaggerated or misshapen examples of necessary growth, and in the face of advancing knowledge the conservatives have to yield, but it is right to argue the issues, for they are extremely important: words have more power than any currency and longer lives than any individuals. We tend to think of them as simple and arbitrary labels to known entities, but a word's mere existence tells us that something is worth naming, thus drawing attention to a mass of associations observed in the past. Thereafter its animal capacity to form links, further associations, and intermarriages with other words provides a fertile receptacle for new knowledge.

In summary, then, the question "Does induction change the language?" must be answered in the affirmative: language grows by the creation of new words and the new usage of old words, and this growth occurs in response to new knowledge. But I think the analogy has cast new light on one aspect of language. The earlier cybernetic discussions pointed out its high redundancy, implying that this was wasteful and uneconomic, or at best served only to protect it from distortion by random errors or noise. The analogy shows that this redundancy—this nonrandom pattern of associations between words—is its very life, for it holds the wisdom of the tongue, and without it growth of knowledge could only occur by the accretion of new words.

Conclusions
It is difficult to summarise and conclude because of the range of preconceptions with which different people approach the problems considered here. Some will probably have found that I have evaded or missed out all the aspects they think are important and interesting, but I hope others will have come across relationships they had previously missed, and will see the unity of goal in the topics discussed. What I shall do is show how working through the analogy cast light, for me, on issues that are quite deep philosophically, or of interest for other reasons, but I fear that such answers as I try to give will not be nearly thorough enough to satisfy professionals.

In the first place, one sees a continuity between the formation of sensory messages from physical stimuli, the rearrangement of these messages to make our sensations and perceptions, the formation of concepts about these perceptions, the attachment of words to concepts and percepts, and the modifications to the meanings of words by inductive inference. These are all stages of what might be called inductive coding—the economical representation of as much as possible about the sensory messages we receive. Of course the universality of inductive inference in
nature has been widely recognized. To illustrate the fallibility of induction Russell (1912) used the example of the chicken that learned to run out every day to be fed by its owner; ultimately the lesson it had learned merely enabled the owner to catch it for the pot and wring its neck. But the example has an additional significance if words are themselves formed by processes similar in principle to those that control the chicken's actions. No one would liken Russell’s reasoning powers to those of a chicken, but is it not interesting to conclude that the main tools of his reasoning, namely words and language, are the product of something rather like chicken’s reason, and hence as fallible? The pitfalls this creates in the use of language do not seem to me, even now, to be fully recognized.

I think there is another problem where the analogy affords some insight. In all fields of science there tends to be conflict between practically and theoretically oriented individuals. The practical man resents the attempt to explain away his painstakingly extracted facts, and is especially upset when a theory is successful and thus devalues continued labour in his particular field. The theoretician, on the other hand, scorns the desire to collect every minute empirical detail, and regards all facts as trivial except those bearing critically upon his theory. In the light of the analogy I think one can find a reasonable middle attitude. Newton’s tag “Natura enim simplex est” seems no longer a mere declaration of the theoreticians’ faith or prejudice: it can be regarded as the modest claim that even a little simplification is worth striving for when faced by chaos, because without simplification chaos is indescribable. Thus on the one hand the empiric may more readily accept the need for theory if he translates the tag “Simplicity must be found in Nature if we are to describe it”, and on the other hand the analogy may indicate to the theoretician the survival value that lies behind his restless pursuit of simple truths in complex situations.

The analogy shows very clearly the usefulness of partial truths, for they allow much of the economy attainable with generalisations that have no exceptions. “All swans are white” is not true, but it is nonetheless useful to assume that a swan without a qualifying adjective is a white swan, for the same reason that it is economical to delete a $u$ after a $q$; the exceptional cases can be handled at somewhat greater length without eliminating the economy, as when we proposed inserting a $u$ after a $q$ in the exceptional cases when it was absent.

This line of thought leads to the definition of inductive truths as fictions of the imagination that facilitate description. This is not belittling if one appreciates the scope of what there is to describe, and there is surely no need to believe that inductive truths have the same absolute character as those obtained and used in deductive reasoning.

Since Kant, some have worried about the a priori element in our perception and reasoning. They say that we see only what we have been programmed to see, and can organise perception and knowledge only according to inflexibly programmed rules. The analogy suggests a goal for our perceptual codes and for the inductive generalisations by which we organise perceptions. If the goal is known, it becomes possible to check whether it is being achieved, and the possibility of this internal check is crucial. Of course genetically determined mechanisms may—in fact must—limit the choice of ways in which we can perceive and think, but the inflexible element is much less disturbing if the choice can be checked for its goal achievement.

To conclude, the analogy tells us that the selective advantage of inductive reasoning, and of the perceptual mechanisms that precede it, is the more complete and accurate representation of the world around us both in our individual minds, and in the corporate knowledge stored in language and libraries. Viewed in this light induction is bound to seem, not a habit of mind learned early by most and practised expertly by scientists and a few others, but the inevitable consequence of evolution under conditions where
knowledge of the environment is advantageous. If he who knows more, grows more and sows more, then the genetic mechanism will ensure that techniques of acquiring knowledge evolve, and social mechanisms are likely to ensure that they are taught.

References
Popper, K. R., 1934, Logik der Forschung (Springer Verlag, Vienna).