BRIAN BUTTERWORTH

Maxims for Studying Conversations

This paper is divided into three parts. The first outlines problems which the student of conversations faces; of these, some are specific to conversational material, others general to many kinds of scientific endeavour. The second part attempts to provide a philosophical framework within which these methodological problems can be considered. The third analyses a few existing studies of conversation, including some of my own, from a methodological viewpoint to see what lessons can be drawn from them. In the course of these analyses, I will formulate methodological prescriptions as maxims that can be considered as suggestions as to how to do better work on conversational material.

I would like to be able to say that the maxims I shall be propounding are those which have guided my own research: it would, however, be more honest to say that they are reconstructions from (the better portions of) my actual practice.

THE STUDY OF CONVERSATIONS

Let me start by defining the problem in a simple, though not necessarily uncontroversial, way. The problem confronting the student of conversations is to interpret the various behaviours exhibited by conversationalists. This may not, of course, be the student's primary purpose, nor indeed how the student formulates the problem to himself. Often, perhaps typically, conversations provide the only or the best naturalistic database which bears upon a particular aspect of human ability or performance. It thus becomes a test-bed for assessing theories in that domain. In my own case, my primary purpose is to find out about the mechanisms responsible for language production, and conversations are obviously the paradigm example of naturally-occurring speech: it is reasonable (though not necessarily correct) for me to suppose that whatever the mechanisms are, they are set up to do conversations first and foremost, and that other speaking behaviours adapt or make use of these conversationally-oriented mechanisms. Nevertheless, it would be an illegitimate inference to suppose that some particular behaviour has a simple, single source – a single underlying mechanism.

To take an example from my own work: a pause may be the function of the production mechanism delaying output on account of some time-consuming decision process, like choosing the right word; on the other hand, the speaker may have no such decision to make and is pausing to create some conversational or social, and essentially nonlinguistic, effect (see, for examples, Abercrombie 1968). The behaviour may have two sources, say, social and cognitive, which combine additively or interactively or in some other complex manner.

Thus even if one's purpose is to model some one aspect of conversational behaviour, the employment of conversational data to support that model requires taking into account models of other aspects which might also concern themselves with describing the same behaviour.

It should go without saying a conversation is an extremely intricate phenomenon in which cognitive and neuromuscular skills are put at the disposal of a range of personal and social purposes, and the whole embedded in interlocking systems of social and linguistic conventions. The pattern of behaviour conversationalists deploy, skilfully or inadvertently, shows comparable intricacy. Vocal output; the movement of limbs, head and trunk; the posture and position relative to other conversationalists; gaze; facial expression; the manipulation of objects in the environment: all these are located in time and space, and can be ordered with respect to each other or with respect to similar behaviours of other present conversationalists. Each aspect itself has an internal ordering of its parts: the order of words in vocal output, the sequence of hand movements, etc.

It is not surprising then that the range of analytical tools the student can bring to bear on the description and interpretation of conversations is enormous, and drawn from a number of different disciplines which communicate with each other haphazardly if at all.[1]

Each discipline tries to isolate aspects of behaviour whose interpretation generalises across conversational corpora, subject to constraints on generalisation which are determined empirically or *a priori*. An example of the latter type of constraint can be found in linguistics. Chomsky (1965) has formulated a distinction between 'competence' and 'performance', *a priori,* which enables him to disregard certain kinds of naturally-occurring linguistic phenomena; thus slips and errors in speech do not constitute part of the corpus which a grammar is obliged to describe, and hence syntactic rules do not generalise to corpora containing them.

Empirical constraints can be exemplified from the literature on pausing. Pauses may serve to give the speaker time to decide among potential continuations – the next word – (Goldman-Eisler 1958), and it is possible to identify which pauses in a given corpus serve that function (Butterworth 1976). However, it turns out that well-rehearsed speech, or reading, has a quite different pattern of pausing (Goldman-Eisler 1968, 1972) and makes quite different cognitive demands on the speaker. Thus the interpretation of pause functions does not generalise from spontaneous speech to rehearsed speech or reading.

The rehearsal constraint raises an important but neglected methodological issue, namely, whether the variables mentioned in the model are meant to be continuous or discrete. For example, a model could specify *degrees* of rehearsal as the independent variable and *degrees* of approximation to the spontaneous pattern in the distribution of pauses as the dependent variable. Any measuring procedure will likely give the appearance of continuous change in both variables. When Goldman-Eisler (1961) plotted the decrease in pause time from trial to trial (i.e., with increasing rehearsal), the resulting graph showed a negatively accelerating reduction in pause time. One interpretation is that rehearsal is a continuous factor which modulates each component of the process in a continuous manner. On the other hand, one could postulate, following Hughlings Jackson (1878), two separable classes of components: already-organised' components and 'now organising' components. Rehearsal would then consist of the discrete replacement of a 'now-organising' component with an 'already-organised' component from one trial to the next. Insofar as pause time reflects the decision processes involved in current organisational processes, pause time will decline in proportion to the increase of 'already-organised' components. There are thus two serious obstacles to deciding whether to propose a discrete or a continuous model: first, the appearance of continuity is not a guarantee that the underlying mechanisms work in a continuous manner; second, the data may be too gross to differentiate continuous change from small incremental or decremental steps. In the case of rehearsal and pause time, independent grounds and additional data would be needed to decide between a continuous model and a discrete one.[2]

One finds in linguistics models where an independent variable of unspecified form is held to be responsible for continuous distributional changes, but where each data point has a discrete value. An example of this is the 'variation theory' account of class or regional dialects. Labov (1966) has proposed that features of a particular dialect are not constant in the speech of an individual, but vary according to the social circumstances and/or the speaking task in which the speaker is engaged. A prestige feature of American English is the post-vocalic /r/. The proportion of occurrences of this feature depends on whether speech is casual conversation or reading a list of words. For any potential occurrence of this feature, there are just two discrete values: present or absent; and for any potential occurrence, variation theory cannot say whether the feature will be present. It can only say that in, for instance, casual speech there will be x% and in reading a list there will be y%, where $x < y$. The values of x and y for a given speaker will depend on both the task and his social class. Trudgill (1974) has made similar findings for class and task variables in a regional sample of British English (where, interestingly, it is the *absence* of the post-vocalic /r/ which has high prestige). He, like Labov, proposes a set of grammatical and phonological rules which contain a probability function determining the applicability – rather, the frequency of applicability of these rules; and the probability function is itself sensitive to task and class factors.

For many linguists working with discrete models (e.g., generative grammar), the introduction of a probability function modifying the operation of discrete rules is both awkward and unsatisfying, since for each potential occurrence, the theory is silent about whether the rule will be applied or not. The introduction of a probability function is unsatisfactory as an explanatory device: it just trivially adapts the theory to fit inexplicable data. Probabilities are not explanations, only summaries.[3]

Not all models using continuous independent variables are, of course, probabilistic in the sense of being indeterministic (see note 3). Some hypothesise a *continuous relation* between a single continuous independent variable and a single continuous dependent variable. Some require a threshold device, or a criterion value, so that although the variables are continuous their relationship is not. Some rely on the interaction of several underlying continuous variables in a continuous relationship with a single dependent variable. (We have yet to see a plausible or well-worked out linguistic or psychological theory which employs Catastrophe mathematics, where several continuous independent variables interact to give a non-continuous relationship with the dependent variables. But see Zeman 1976.)

Thus the discontinuity of the dependent variables cannot guarantee that the underlying mechanisms are not modulated in a continuous manner. Additionally to borrow an analogy from Goldman-Eisler (personal communication), separate entities like oil and vinegar can be combined to give a continuous result: vinaigrette. A discrete model of vinaigrette is necessary, which says that it is composed of two separate things, though some component of the independent variable (say, proportion of vinegar) must be continuous or minutely incremental, to explain changes in taste – a dependent variable.

So we are confronted by a range of problems associated with the form of the model and the parameters it invokes: there is the quite general problem of specifying the form of these parameters, and the obscurity of models where the form is not specified at all. At the same time, conversations are a kind of microcosm of the human condition and the student needs to appeal to a wide variety of disciplines to justify his interpretation of a particular piece of conversational behaviour.

A PHILOSOPHICAL FRAMEWORK FOR CONSIDERING METHODOLOGY

One way of trying to evade the difficulties mentioned in the last section is to abandon hope of a 'scientific' treatment of conversational phenomena. An investigator may refrain from rigorous quantification of the data, from the construction of formal models, from strict testing of theories against facts. Instead he will adopt a research strategy in which he contents himself with telling a 'story', arguing perhaps that a more 'scientific' version will follow in due course, and that in any case stories are necessary precursors of formal models and rigorous quantification. However, I shall argue that stories do not evade the difficulties, because the logic of 'stories' is just the same as the logic of science, and the method of stories is just, in essentials, the method of science but less responsibly carried out.

A second way of trying to evade the difficulties is to quantify quite rigorously etc. but restrict consideration to a single aspect of conversational behaviour, with a concomitant postulate of a single underlying mechanism to explain it. This strategy also fails, because the assumption of a single sufficient cause is usually, as a matter of fact, illegitimate. Popper's picture of the scientist generating hypotheses *in vacuo* and stringently trying to falsify them by crucial experiments is not rich enough to serve as a guide to investigators of conversational behaviour. Instead, I shall adopt an account of scientific endeavour proposed by the philosopher, Paul Feyerabend, though modified in certain ways.

Feyerabend's position must be understood against a background of debate in the philosophy of science where the main protagonists are (1) the Logical Empiricists (e.g., Carnap), (2) Popper, (3) Kuhn, and (4) Lakatos. The central dictum of the Logical Empiricists is that there must exist an observation language independent of any particular theory, and that statements in this observation language can be reached as consequences of two or more theories in the relevant domain. Rational choices between theories will be made just in case theories T1 and T2 have as consequences the observation statements O_i^1 and O_j^2. Theories are confirmed to the extent that their respective observational consequences turn out to be true. If T1 has more true O_i^1s and less false O_i^1s than T2 has true O_j^2s and false O_j^2s, it is to that extent better confirmed. For Popper it is necessary that some O_i^1 and some O_j^2 are logical contradictories; and thus a crucial experiment will falsify one O and hence rule the relevant T out of business. Lakatos proposes that not *any* falsification of a theory will rule it out of business, and has a complex, but more liberal, criterion for theory abandonment.

Kuhn (1962) takes an altogether less tolerant and more authoritarian view of science. In a mature science, he would argue, there is only one theory (or 'paradigm') at any one time. Theories go out of business after accumulating observational crises leading to a (non-rational) 'paradigm-shift'.

In his earlier papers, Feyerabend, like Popper and Lakatos, sees science as a 'critical activity' – that is, progress is achieved by rational criticism of existing positions. But in a number of important respects his account is crucially different from theirs, as well as from the Logical Empiricists and Kuhn. First, he argues, against the Logical Empiricists, that a theory-neutral language is impossible; moreover, interestingly competing theories[4] are *incommensurable:* the terms of one are not translatable into terms of the other. Therefore there cannot be a sentence in T1 which has a contradictory in T2. Thus one cannot decide between interestingly competing theories on the basis of crucial experiments. So Popper's rationalistic account of theory choice must be inadequate. Secondly, scientific criticism must involve not only matching the theory with the observations, but evaluating the theory against interestingly competing alternatives which might generate data beyond the scope of the to-be-criticised theory. He thus opposes Kuhn's idea that 'normal science' is governed by a single theory because in fact scientists typically operate with alternatives, but if they were to operate with just a single theory this would be equivalent to working with a closed metaphysical system incapable of radical revision.

Kuhn further maintains that 'revolutionary' – paradigm-shifting – science is characterised by a renewal of interest in foundational studies. Feyerabend notes that this is frequently true for the physical sicences, and indeed hinders them since alternative 'metaphysics' which could generate new data are ignored. However, biological sciences and even "certain parts of psychology are far ahead of contemporary physics in that they manage to make the discussion of fundamentals an essential part of even the most specific research" (1970:198-99 fn. 4).

The message of Feyerabend's philosophy of science – which I was able here only to caricature (the persuasive case studies can be found in (1970, 1975) – may be summed up in Mao Tse-Tung's dictum (alas never implemented in China in his lifetime): "Let a thousand flowers bloom. Let a hundred schools of thought contend." The study of conversation should thus provide a beautiful case for a Feyerabendian analysis. Conversation is approached from very many viewpoints, which differ in details, and more interestingly, in the very metaphysics the various practitioners bring to it.

Now one problem not sufficiently considered by Feyerabend is how competing schools of thought or theories contend. How is the score kept? Degree of confirmation – in the sense proposed by the Logical Empiricists – cannot be the answer because there is no neutral observation language. Simple disconfirmation, à la Popper, cannot be the answer for reasons cogently advanced by Lakatos (1970). Nor can Lakatos's own idea of progressive and regressive problem shifts; that is, does T1 account for a larger portion of the domain than T2? It may take time to formulate an alternative theory, and at some arbitrary time, the alternative may be insufficiently developed to provide a fair test. It would be like setting a boy to fight a man. Theories obey the law of uneven development. For this reason, the articulateness and specificity of the theories cannot be the test either.

This position is, of course, incompatible with the methodological strictures laid down by linguists. It rules out both 'discovery procedures' (of, say, Bloomfieldian linguistics) and also Chomsky's evaluative procedures: the second as being too restrictive, and the first as being *much* too restrictive.

However, Feyerabend has made out the case only that existing proposals for methodological prescription are unsatisfactory: we are not compelled thereby to take the position that all methodological prescription is illegitimate. At one point in his extraordinary book, *Against Method* (1975), he criticises Carnap for the way in which he abstracts from science in order to build his system of *applied* inductive logic. "But when abstracting from a particular feature of science we should make sure that science can exist without it, that an activity ... that lacks it, is (physically, historically, psychologically) *possible;* and we should take care to *restore* the omitted features when the abstract debate has come to an end. ... The physicist who has used geometry (which disregards weight) to calculate some properties of a physical object puts the weight back after he has finished his calculations. Not once does he assume that the world is full of weightless shapes." (184, fn. 7). Now I would want to argue that Feyerabend is guilty of exactly the same sin as Carnap is guilty of: he omits (and does not restore) two essential features.

First, science is a public (indeed, a co-operative) activity. This immediately entails that scientists must be able to communicate with one another, and therefore methods and procedures must be publicly accessible (i.e., to other scientists) and replicable. (Feyerabend presents a brilliant historical analysis of the communicative techniques employed by Galileo, in which the moral he wishes us to draw is that scientific communication can be, and in this case needed to be, a species of propaganda. But the fact that Galileo resorted to cheap tricks can not be taken as either typical or desirable.)

Second, whatever might be the problems about defining rational grounds for theory preference, whatever the problems about the incommensurability of theories and the impossibility of a neutral observation language, theories *are* judged in relation to facts. Ordinary folk – nonscientists – have principles for reasoning about matters of fact, which scientists themselves employ, albeit in a more rigorous and, perhaps, more scrupulous manner. For example, if it is claimed that Chelsea's failure to win a particular foobtball match was due to the inability of their forwards to convert scoring chances into goals, the claimant would have to revise his view in the light of facts that could be brought to his attention – like the fact that Chelsea lost 1-7 and only had four scoring chances. Of course, total abandonment may not be necessary. One line of defence would be: the opposition knew that Chelsea's forwards were hopeless in front of goal (could not convert scoring chances) and therefore were able to afford to send more men into attack. Nevertheless, some rethinking and recasting of the original hypothesis is forced. Similarly, if T1 predicts that $x = y$ and it turns out that $x \neq y$, this is a difficulty for the theory that has to be resolved (even though it might not be grounds for abandoning the theory). Facts are an indispensable aid to thinking about the problem at hand.

(One of Feyerabend's favourite authors, J. S. Mill, in *A System of Logic* (Book III, Chapter VIIIff.), neatly summarises the principles of reasoning about matters of fact: these are as applicable to lay as to scientific reasoning.)

This is a crucial reason why 'stories' are unsatisfactory: they rarely provide conditions for direct testing against facts. Ordinary folk and scientists are alike entitled to ask for reasonable empirical justification.

This student of conversations is in an ideal position to avoid operating with what Feyerabend terms "a closed metaphysical system", and to exploit instead the "pluralistic methodology" (1975:30) which becomes available due to the convergence of so many disciplines on conversational data. Feyerabend's preferred critical method, via consideration of radically different alternatives, will be employed below in an attempt to improve some existing accounts of conversational behaviour. This method in conjunction with my two objections to Feyerabend will lead to six methodological and meta-methodological prescriptions: the Maxims.

The first objection leads directly to the first maxim:

(1) MAKE YOUR METHODS PUBLIC

If your methods are not public, you are failing in a primary duty to the scientific community because you are not communicating vital information to your fellow-workers: they cannot reasonably hope to check out your ideas for themselves. (You are also failing in a duty to the wider community in exactly the same way as manufacturers of foodstuffs who do not list the ingredients in the product. The consumer, without the time or facilities to make his own analysis, is deprived of adequate protection, and although the list is not a guarantee it does give some additional protection. In science as in business, fraud is possible; thus for both, *caveat emptor.*)

By making methods public, I mean that the techniques and procedures employed in the research must be made as explicit as is necessary for another researcher to replicate the study (or, at least, know why he cannot replicate it). Incidentally, this requirement breaks down a distinction commonly found in the philosophy of science between the 'context of discovery' and the 'context of justification' (see, for example, Feigl 1970; but compare Feyerabend 1975, Chapter 14): you cannot make explicit in a communication about a method what was not explicit in the method itself.

STORIES AND THEORIES

I want to distinguish two kinds of account of conversational behaviour, 'stories' and 'theories' (though I intend this distinction to have general applicability). Stories differ from theories in a number of ways, and it is helpful to think of them at opposing ends of a spectrum.

Stories are less formal, less explicit, less cohesive, and less exhaustive (though they are frequently broader in scope) than theories. These differences are not meant to be value free: theories are better, but this is not to say that stories do not have a role in science. It can be argued that stories are necessary precursors of theories.

A good story picks out important features in the domain to be described, and shows how these features are to be interpreted. The criteria for picking out certain features are embodiments of the investigator's unique talents, experience, and insight, rather than explicit consequences of a fully-articulated theory, and are thus essentially private. To some extent, the investigator can share his ideas with others. He can persuade them to see in the materials what he sees, though he cannot make his criteria explicit. Thus one important way by which story-tellers recruit others to their point of view is through apprenticeship. There is no reliable way for the non-apprentice to read the story, examine the material and replicate the original results.

One crucial test of theories is not appropriate for the evaluation of stories: test by counterexample (empirical refutation). Stories do not exhaust the domain in the way theories do: they do not have to account for every detail. An account will be exhaustive if there are principles for excluding certain classes of data and considering the whole of the residue. One defensive measure that theories are often forced to take is shifting the boundaries of the excluded class. Generally, stories do not define the excluded class.[5]

A second important difference between stories and theories concerns the degree of cohesiveness and articulateness. If an empirical consequence of a theory turns out to be disconfirmed, the theory as a whole must be re-assessed, though the adaptive change required might well be small, even trivial. Statements in stories are not bound together in such a tight way. Some parts may be abandoned without affecting the rest of the structure. This is another reason why criticism of a story cannot be straightforwardly empirical – i.e., tested against facts. Rather, it must be evaluated by assessments of coherence and insight. In this way, the 'truth' of a story is evaluated in much the same way as a literary work, and its value is a function of similarly recondite and enigmatic properties.

There are many persuasive and plausible story-tellers whose work bears upon, and is based upon, conversational behaviour. The question of practice and principle that needs to be raised is whether another investigator can make use of the ideas in the story to analyse a new conversation. I shall take as exemplars two well-known, widely-respected, and influential story-tellers: Goffman (esp. 1959) and Scheflen (esp. 1964). They are chosen not only because they are skilful representatives of this kind of investigation, but also because their methods vary somewhat in explicitness. Obviously, I cannot hope to do justice to the extensive and varied work of these authors, so I will select, not unfairly I trust, some small samples which are directly relevant to my thesis.

(1) *Goffman* (1959, Chapter 2, section headed "Maintenance of Expressive Control". Pagination refers to the Penguin edition.) In this section Goffman mentions behavioural events which accidentally conflict with the impression the speaker is trying to convey.

> In our society, unmeant gestures occur in such a wide variety of performances and convey impressions that are in general so incompatible with the ones being fostered that these inopportune events have acquired collective symbolic status. Three rough groupings of these events may be mentioned. First, a performer may accidentally convey incapacity, impropriety, or disrespect by momentarily losing muscular control of himself. He may trip, stumble, fall; he may belch, yawn, make a slip of the tongue, scratch himself, or be flatulent; he may accidentally impinge upon the body of another participant. Secondly, the performer may act in such a way as to give the impression that he is too much or too little concerned with the interaction. He may stutter, forget his lines, appear nervous, or guilty, or self-conscious; he may give way to inappropriate outbursts of laughter, anger or other kinds of affect which momentarily incapacitate him as an interactant. (1959:60)

Although this has a ring of truth (or perhaps just a ring of confidence), one is entitled to ask not only how Goffman came by this statement, but how one might check it out. How are slips etc. distinguished from the intended part of the performance? Are we, for example, to take it that all slips of the tongue, falls, and so forth convey incapacity in any conversation? And if not, how is the analyst to use these generalisations in interpreting a novel conversation? Had this been a theory and not a story the class of excluded data would have been defined. There is a special problem with much of Goffman's work which is well illustrated here. The 'gestures' he lists are not all simple behaviours; some are interpretations. Being flatulent, scratching oneself and yawning are behaviours; appearing nervous, inappropriately laughing, appearing angry are interpretations of behaviours. The interpretation is already built into the behaviour which Goffman offers to interpret. So unless one sees with his eyes, it is impossible to give his ideas a decent chance with the data.

(2) *Scheflen* (1964) gives us a kinesic analysis of the postural accompaniments to unrehearsed therapeutic conversations. He illustrates some of the movements he is considering with pictures which show the typical relationship of certain movements with the speech they accompany. As most readers will know, postural shifts are held to mark structural changes in the interaction. To take just one example, Scheflen defines a structural postural unit he calls a 'point'. This can be marked by a head position, and terminated by a shift in head position. 'Shifts' mark the boundaries of these units. Each 'point' is held to correspond to a point – an idea, perhaps – of one or more sentences of the discourse (i.e. check that, say, head shifts correspond to boundaries of discourse points).

If you wished to replicate or extend Scheflen's study, you would need to know at least the following things, none of which are mentioned in the paper: (i) what sized movement is to count as a 'shift'? (ii) are all such movements to be interpreted as 'shifts'? (iii) how close a temporal correspondence needs there be between shifts and the boundaries of discourse points? (iv) how can discourse points be determined? (v) what percentage correspondence should be expected between postural and discourse points?

Now it is likely that Scheflen, consciously or not, has managed to filter out irrelevant head movements (i.e., tacitly defining the class of excluded data) and knows a discourse point when he hears one: he is a highly experienced, shrewd observer and interpreter of human behaviour. He is thus doubly guilty of dereliction of duty to the scientific community: he is not communicating enough information for other investigators to continue his line of research, and by keeping crucial criteria tacit he is depriving the rest of us of valuable fruits of his unique experience.

What is more, since we are not given adequate grounds for recognising counterexamples we will be a bit short on facts which might force us to rethink the story. Stories of this sort can become impregnable authorities, or nothing. Either way they will fail to provide the creative focus which is vital to scientific progress. It is striking that although Goffman and Scheflen are widely cited, I have found very little detailed critical appraisal of them. One must either simply cite them or ignore them.

Although, as I remarked above, conversations are exceedingly complex phenomena, it is possible and desirable to formulate explicit theories and use explicit investigative procedures. One example of an account which approaches these desiderata (whatever its shortcomings in other respects) also deals with nonverbal aspects of conversational behaviour, in particular with hand and arm movements accompanying speech. Geoffrey Beattie and I (Butterworth and Beattie 1977) had been puzzled by finding an asynchrony between the onset of a gesture and the onset of the word or phrase with which it appeared to be 'iconically' or representationally associated. (The instance which first brought this phenomenon to my attention occurred when a speaker raised his right hand and two seconds later said "If you want to raise any problems ... ".)

We examined a number of videotapes of unrehearsed, usually dyadic, conversations and noticed that this phenomenon was quite common. Indeed, we discovered no instances in our corpus where gesture onset followed word onset. These data led us to formulate a theory which tied gesturing to the speech production mechanism, and which would explain the asynchrony. Since the meaning of the word(s) and the 'meaning' of the gesture were associated, the gesture represented the meaning or part of the meaning of the word; the speaker must have known at least part of the meaning of the word he was going to say before he said it by a duration equal to or greater than the asynchrony. We hypothesised that a gesture is controlled by that stage in the transduction from thought to speech when the speaker knows the meaning of the next few words he will utter, but has not yet found the words themselves, i.e., their phonological form.

This hypothesis is fairly specific: it asserts that gestures are *not* controlled by that stage where only the general outline of the next chunk is known, nor by the stage where the lexical item itself has been located. There is some evidence that speakers plan ahead one 'Idea' at a time, where an Idea comprises from one to about ten clauses. When the speaker is formulating the plan of an Idea, his speech is relatively hesitant; once the plan has been worked out in some detail, speech becomes relatively fluent, interrupted only by shortish pauses for making the final lexical choices and for marking some major syntactic boundaries (Butterworth 1975, 1976).

The asynchrony occurs because the time to make the choice of words is longer than the time to make the choice of gestures, and it is longer because the ensemble of words from which the speaker has to choose is larger than the ensemble of gestures; and we know from Hick's Law (Hick 1952) that choice time is a function of ensemble size.

In order to get this theory off the ground we had to construct exclusion principles which defined the classes of data we were and were not considering. (i) We confined ourselves to hand/arm movements. (ii) We needed to exclude movements which might not be speech-related, and so all movements which were not timed with the stress pattern of the speech were ignored (e.g., scratching, self-touching, rubbing, etc.). This was a conservative criterion: we could be reasonably sure that movements timed with speech had something to do with the speech-production mechanisms, but we could not be sure that the others were unconnected. This is always the problem with exclusion rules (cp. the ongoing debate about the autonomy of syntax). (iii) We had to exclude movements which did not bear a meaning congruence with the words. Effectively, this class was defined as those movements for which we could not *see* a meaning congruence. It turned out that the excluded class had a number of other properties which we had not anticipated and which were not shared by the class we were considering. For example, the excluded movements were repeated frequently in the conversation, they were usually movements in one plane, they distributed differently with respect to the pausal pattern, and they seemed to have a clear emphatic function.

Given these exclusions, the theory makes quite specific predictions about when gestures should occur. They should occur predominantly *after* the initial formulation of the Idea has been carried out. In unrehearsed speech, there is a regular alternation of hesitant planning phases and fluent execution phases. As I said, the planning phase is where the Idea is formulated, and the execution phase is where the plan is turned into words. So gestures should predominate in fluent phases. Moreover, they should onset in pauses in the fluent phases, that is, where lexical choices are being made. Thus it turned out. (Quantitative data can be found in Butterworth and Beattie 1977.)

Incidentally, the 'emphatic' speech-related movements did not show either pattern: they tended not to start in pauses, and were not more frequent in execution phases.

The advantages of trying to interpret the behaviour with a theory rather than a story can now be identified more fully, with reference to the above example.

1. The methods are explicit for the most part, and

2. where the method does use tacit principles – e.g., in the association of meaning between word and gesture – it it clear where this happens. It is also possible that the tacit principles will come to be replaced by explicit ones.

3. The classes of excluded data are defined.

4. The numerical presentation of data which *ipso facto* indicates that there are counterinstances to the theory, implies that the causes of the behaviour are not exhausted by the theory.

5. Though not expounded here, the form of the variables under consideration can be outlined – in this case, a basically discrete system with a continuous parameter (time) on the lexical selection process.

6. The whole study is straightforwardly replicable.

Thus maxim (2) is

(2) THEORIES ARE BETTER THAN STORIES

There are other *theories* of gesture which employ reasonably explicit methods and quantitative data. I wish to draw attention to one that suffers from a quite different kind of methodological fault.

In a number of papers on turn-taking, Duncan (1972, 1973) has claimed that gesturing is a signal or cue intended to convey to the listener that the speaker has not finished his turn; conversely, gesture-termination is a cue that the turn is now over and the listener may take the floor. Duncan presents correlational data which he interprets as revealing that gestures are cues – generally in conjunction with other cues – since turn-changing is more likely following gesture-termination.

Our theory is not only not exhaustive, but we find confirmation of Duncan's data: gesture-termination significantly often precedes a turn-change. There may thus be converging sources of gesture onset and termination. But can Duncan show that turn-signalling is an *additional* source and not a second-order effect of lexical selection? That is to say, can we interpret Duncan's data so that it becomes a consequence of our theory? Alternatively, can Duncan explain our data so that it becomes a consequence of the turn-signalling theory? We think he cannot, since he offers no resources to account for the asynchrony or for the difference in distribution between gestures proper and emphatic movements.

Our critical appraisal of Duncan's theory rests not on direct refutation, but on bringing in additional data which his theory does not deal with, and additional conceptual tools. Consider what is happening in the conversation. The speaker plans and then executes the plan of an Idea. Gestures, being linked to lexical selection, begin at some point *before* the completion of the Idea, and are likely to terminate close to the end of the Idea, soon after the last lexical items requiring some decision time. Suppose simply that speakers do not wish to be interrupted in the middle of an Idea, but are more willing to exchange turns when they have finished an Idea. If this supposition is correct, the correlational data falls out as a second-order effect without the need to posit a distinct signalling function for gestures.

Of course, skilled conversationalists may become sensitive to the temporal patterning and linguistic organisation of Ideas, and will thus come to time their contributions to the conversation so as to coincide with Idea boundaries – those points where floor-claiming will be more successful and more polite. Conversations are not just the orderly exchanging of turns, but include, as usually the primary purpose, the linguistic expression of thoughts. Hence maxim (3), which is really a special case of Feyerabend's plea for a pluralistic methodology.

(3) REMEMBER THAT CONVERSATIONALISTS TALK

A pluralistic methodology means doing justice to the complexity of the phenomena under investigation. Failure to do this is, alas, a common failing of studies of the nonverbal aspects of conversational behaviour. Argyle and his colleagues (past and present) feature among the guilty parties, notably in their account of gaze. Again we find gaze treated as a cue to turn-taking (though, Kendon (1967) for example has a rather more complicated view) (see Argyle 1972, and 1975, Chapters 8 and 12). Again we find oversimple interpretations of correlational data which shows gaze at the listener (abbreviated to L-gaze) near the end of the turn and gaze away at the beginning of a turn. A correlation or contingency analysis will doubtless 'demonstrate' a significant relationship between the presence of the cue and turn-changing.

Remember that turn changing[7] typically occurs at the end of an Idea. When the new speaker begins, he starts at the beginning of a new Idea and, unless he has planned this Idea while listening, he will start off with a hesitant phase. In this phase, our theory asserts, the speaker has more cognitive work to do. When the speaker is looking at the listener he may actually be seeing him, and, for example, monitoring his reactions to what has been said. This will demand *additional* cognitive work.

There is considerable evidence that when a subject has to carry out two cognitively demanding tasks at the same time, performance on one or both suffers (in comparison with performance on each task carried out separately). Psychologists who have maintained that the cognitive system is a single, limited capacity channel (like Broadbent 1958) find this; so do those who assert that the cognitive system is more labile, and can be organised into parallel channels for each task. Macleod (1977), for example, has shown an overall decrement in performance even though the two tasks do not interfere with each other: he argues that capacity is required to keep the tasks organised into separate channels. Interference only occurs either when there is cross-talk or when the two tasks share the same processing subsystem.

The skilled speaker may try to avoid situations where performance decrement is likely: he will try to avoid taking on a monitoring task when his speech requires maximum capacity i.e., during a planning phase. Beattie (1978) has shown that this is indeed the case. This result by itself would explain the correlation between turn changing and gaze, and make the monitoring theory at least as plausible as the signalling theory But we can go further. On those occasions when speakers do L-gaze during planning phases we would expect performance decrements: inefficient monitoring (which I cannot at the moment measure) and/or planning failures. And it turns out that planning phases where L-gaze occurs are accompanied by five times as many false starts as those without L-gaze (Beattie 1978).[8]

Consider the contrasting predictions of the signalling and monitoring theories for conversations where the opportunity to monitor is denied, as on the telephone. Again we are confronted by at least partially incommensurable theories, in that they focus on different aspects of the data. However, it is possible to derive contradictory predictions.

When visible signals are absent, says Argyle (1975:163), "as on the telephone, interaction is found[9] to be more difficult", and "auditory cues will replace (visible cues) for the purposes of feedback and floor-apportionment" (1972).

The monitoring theory as outlined here, really has no prediction to make about turn-taking latencies, but it offers no grounds for supposing that they will be different on the telephone; it does predict that there will be no substitution of auditory for visible signals, since signalling is not a theory resource. It also makes clear predictions about aspects of the data not covered by the signalling theory. Overall hesitancy (the proportion of silence within turns) will be reduced since some of the pause time required for interpreting listener's reactions will not be needed. In particular, pauses which have a communicative function, for marking major syntactical boundaries, should undergo the greatest reduction on the telelphone since some of these pauses seem not to be used for cognitive work (Butterworth 1976), and if the speaker has adapted to using them for monitoring work they should be disposable. Insofar as gaze is a cue to grammatical boundaries, as has been claimed by Kendon (1967), the signalling theory would predict the opposite: grammatical pauses should become more frequent and longer to substitute for the absence of the cue.

Monitoring theory draws attention to another aspect of performance, namely the quality of speech output – in terms of both content and style. It is possible that speakers will try to compensate for the additional cognitive load in the more taxing condition, by trading off quality for fluency. We say that the face-to-face condition is the more taxing; Argyle would presumably say that the telephone condition is the more taxing. In fact, without controlling for quality any inferences are illegitimate.

Now we all know that most people converse perfectly efficiently on the telephone; and a detailed examination of the relevant parameters bears this out (Butterworth, Hine, and Brady 1977). It also throws light on the comparison between the signalling and the monitoring theories. We found:

(i) The quality of speech is unimpaired on the telelphone.

(ii) There is no increase in the use of behaviours classed as 'vocal substitutes' by Cook and Lalljee (1972).

(iii) There is no increase in the number of grammatical boundaries marked by a pause. In fact, there are fewer and they are half-a-second shorter on average!

(iv) The overall proportion of pause time is lower on the telephone.

(v) The latency of the next turn is unaffected by the condition.

It is, therefore, as well to remember when investigating conversations that human beings are subject to limitations on, among other things, how much information they can handle at any one time. This is another example of the need to take into account ideas and facts drawn from neighbouring disciplines, in this case the psychology of human performance. The methodological lesson here can be summarised in a maxim:

(4) REMEMBER THAT CONVERSATIONALISTS ARE HUMAN

Notice that the monitoring theory made additional data relevant, encouraging a broader conception of conversational performance and providing more opportunities, thereby, for facts to force us to rethink our theories.

SOME FINAL THOUGHTS

The student of conversations is caught between two depressing options. He can try to be rigorous and scientific, and is in danger of construing his task too narrowly. Or he can go for a broad sweep, and therewith end up talking to himself and his immediate associates. There is yet another danger which I have not mentioned so far: trying to give a complete but theoretically neutral description of a conversation and hope later to put all the pieces together. As practitioners will know this is extraordinarily time consuming. One attempt, Pittenger, Hockett, and Doheny (1960), took a whole book to describe the first five minutes of one conversation. Moreover, a neutral description is not only impossible, but any attempt to do it will conceal from the investigator the hypotheses he is in fact bringing to bear on the descriptive procedure.

My advice is, start with a clear idea of what you are looking for, hence the next maxim:

(5) LET THE THEORY DO THE WORK

You may have to start with an extremely complex theory, which interrelates a whole set of hypothetical underlying processes, and which has no empirical justification. That is all to the good! One has to procede *counterinductively:* if we only employed well-confirmed hypotheses science would never progress. In any case, as Feyerabend has pointed out (1975:31), "There is not a single interesting theory that agrees with all the known facts in its domain." The purpose of this maxim is to save observational labour by investing in mental labour. As a spin-off, the maxim will increase the fund of ideas investigators can draw on.

I will end with a paradox, if not a contradiction. Although maxim (5) says that one should approach the phenomena with clear ideas, the phenomena themselves should help provide those ideas. I was convinced by watching Anita Pomerantz, an ethnomethodologist, analyse a videotape of the Nixon-Frost interviews that one should start by approaching a conversation without preconceptions. Hence I was led to the counter-maxim.

(6) LET THE PHENOMENA GUIDE THE THEORY

The paradox reflects, simply, the dilemma the student of conversations inevitably finds himself in, providing he is doing his work responsibly.

NOTES

* This paper has emerged from discussions which took place during the conference *Epistemology for Practising Social Scientists* at SUNY Buffalo in March of this year. It turned out that I found myself having to defend the way in which I studied conversations rather than the substantive results and theories that I had advanced in my formal presentation. One of the organisers, Prof. Madeleine Mathiot, therefore suggested that I submit my defence in writing for further scrutiny, which I now do. I should add that what I have to say here is not identical with what I said in Buffalo. Many of the propositions which there I held tacitly, I here try to make explicit, and some of the positions advanced in the heat of debate have been repented at leisure.

I would like to thank my co-conferees for much stimulating discussion, especially Ken Abrams, Paul Garvin, Albert Scheflen, Ray McDermott, Jim Schenkein, and Madeleine Mathiot (who also forced me to write a methodological paper). My appreciation to Prof. Paul Feyerabend and the late Prof. Imre Lakatos for a series of seminars on the philosophy of science at University College London in 1970; attending it was one of my most enjoyable intellectual experiences. My thanks also to my former student, Geoffrey Beattie, who provided ammunition I hope I haven't squandered.

[1] An attempt to bring some of these disciplinary approaches together in the service of understanding language production can be found in Butterworth (1979).

[2] If the 'continuous' hypothesis were true, it would be expected that pauses in homologous locations in each trial (e.g., before a given low probability word) would become briefer from trial to trial. In fact, it turns out that such pauses simply disappear (Thomas 1975) suggesting that the low probability words, say, become part of the already-organised component.

[3] In quantum theory, probabilities are part of the theoretical apparatus, and the 'Copenhagen School', at least, is perfectly happy to maintain that the universe is basically indeterministic. Sampson (1975) has cited quantum theory as a precedent for the Variable Rule Descriptions in the Labovian theory. Gazdar

(1977), however, points out that this would entail importing the Uncertainty Principle into linguistics, in which case "it might no longer be meaningful to say that an unobserved sleepwalking Cockney had dropped (or had not dropped) an /h/". Nagel (1961) notes that many distinguished quantum theorists – Planck, de Broglie, etc. – are not happy with indeterministic accounts. As Einstein pungently put it: "God does not play dice." And von Neumann has a controversial proof, that it is not possible simply to add a 'hidden variable' to quantum theory to make it deterministic. The problem is that, however, much one may dislike indeterministic theories, quantum theory is the best we have got, and we cannot just ignore it. Maybe the same is true of Labovian theory.

[4] By 'Interestingly competing theories' I mean theories which differ in more than just the values assigned to variables. That is, they cannot be converted into each other by simple mathematical transformation.

[5] It is arguable that Chomsky's greatest, though most controversial, contribution to linguistics is his introduction of rigorous exclusion principles. For example, those which remove 'semantic' and 'performance' factors from the description of a language.

[6] Harold Macmillan says that his advice to novice public speakers is to make sure the gesture does not follow the word (Interview with Robert Mackenzie, BBC TV), which suggests that in rehearsed speech there may be counter-examples to the Gesture First Law. However, Macmillan is notorious for gesturing *after* the word, and he may be generalising from a single, unrepresentative example: himself.

[7] It is important to distinguish between turn-changing in an agreed and orderly way from interruption. Usually, alas, authors do not make this distinction, and define a 'turn' as the point where the second speaker takes the floor. For a systematic and enlightening use of the distinction see Beattie (1977).

[8] Beattie (1977) found that "Significantly more immediate speaker switches were found when the utterance terminated with no gaze (at the listener) than when the utterance terminated with gaze". These, and other data he presents, casts yet more doubt on the signalling hypothesis.

[9] Argyle says 'found' in 1975, when in fact his colleagues Cook and Lalljee (1972) could obtain no data to support this claim. I therefore intend to treat ths statement as a prediction.

REFERENCES

Abercrombie, D.
1968 "Silent Stress", Department of Phonetics and Linguistics, Edinburgh University, *Work in Progress*, No. 2, 1-10.

Argyle, M.
1972 "Nonverbal communication in human social interaction", in R. Hinde, ed., *Nonverbal Communication*. Cambridge: Cambridge University Press.
1975 *Bodily Communication*. London: Methuen.

Beattie, G.
1977 "Floor apportionment and gaze in conversational dyads", *British Journal of Social and Clinical Psychology*, to appear.

1978 "Sequential temporal patterns of speech and gaze in dialogue", *Semiotica*, to appear.

Broadbent, D.
1958 *Perception and Communication*. London: Pergamon Press.

Butterworth, B.
1975 "Hesitation and semantic planning in speech", *Journal of Psycholinguistic Research* 4, 75-87.
1976 "Semantic planning, lexical choice and syntactic organisation in spontaneous speech", Internal Report, Psychological Laboratory, Cambridge University.

Butterworth, B., ed.
1979 *Language Production*. London: Academic Press (to appear).

Butterworth, B., and Beattie G.
1977 "Gesture and silence as indicators of planning in speech", in *The Stirling Psychology of Language Conference*, R. Campbell and P. T. Smith, eds. New York: Plenum (in press).

Butterworth, B., Hine, R. and Brady, K.
1977 "Speech and interaction in sound-only communication channels", *Semiotica* 20/1-2, 81-99.

Chomsky, N.
1965 *Aspects of the Theory of Syntax*. Cambridge, Mass.: M.I.T. Press.

Cook, M., and Lalljee, M. G.
1972 "Verbal substitutes for visual signals in interaction", *Semiotica* 6, 212-21.

Duncan, S.
1972 "Some signals and rules for taking speaking turns in conversations", *Journal of Personality and Social Psychology* 23, 283-92.
1973 "Toward a grammar for dyadic conversation", *Semiotica* 9, 29-47.

Feigl, H.
1970 "The orthodox view of theories", in *Analyses of Theories and Methods of Physics and Psychology*, . . . Radner and . . . Winokur, eds. Minneapolis.

Feyerabend, P.
1970 "Consolations for the specialist", in Lakatos and Musgrave 1970: 197-230.
1975 *Against Method*. London: New Left Books.

Gazdar, G.
1977 "Mr Sampson and Mrs Higgins", in *York Papers in Linguistics*, 175-79.

Goffman, E.
1959 *The Presentation of Self in Everyday Life*. Harmondsworth: Penguin (1975).

Goldman-Eisler, F.
1958 "Speech production and the predictability of words in context", *Quarterly Journal of Experimental Psychology* 10, 96-106.
1961 "Hesitation and information in speech", in *Information Theory*, ed. by C. Cherry. London: Butterworth.
1968 *Psycholinguistics*. London: Academic Press.
1972 "Pauses, clauses, sentences", *Language and Speech* 15, 103-13.

Hick, W. E.
1952 "On the rate of gain of information", *Quarterly Journal of Experimental Psychology* 4, 11-26.

Jackson, J. Hughlings
1878 "On affections of speech from disease of the brain", reprinted in *Selected Writings of Hughlings Jackson* II. New York: Basic Books.
Kendon, A.
1967 "Some functions of gaze-direction in social interaction", *Acta Psychologica* 26, 22-63.
Kuhn, T.
1962 *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.
Labov, W.
1966 *The Social Stratification of English in New York City.* Washington, D.C.: Center for Applied Linguistics.
Lakatos, I.
1970 "Falsification and the methodology of scientific research programmes", in Lakatos and Musgrave 1970.
Lakatos, I., and A. Musgrave
1970 *Criticism and the Growth of Knowledge.* Cambridge: Cambridge University Press.
McLeod, P.
1977 "A dual task response modality effect: support for multiprocessor models of attention", *Quarterly Journal of Experimental Psychology* 29, 000-000.
Mill, J. S.
1843 *A System of Logic.*
Nagel, E.
1961 *The Structure of Science.* London: Routledge and Kegan Paul.
Pittinger, R. E., C. F. Hockett, and J. J. Danehy
1960 *The First Five Minutes: A Sample of Microscopic Interview Analysis.* Ithaca, N.Y.: Paul Martineau.
Popper, K.
1959 *The Logic of Scientific Discovery*. London: Hutchinson.
Sampson, G.
1975 *The Form of Langauge.* London: Weidenfeld and Nicholson.
Scheflen, A.
1964 "The significance of posture in communication systems", *Psychiatry* 27, 316-31.
Thomas, C.
1975 Unpublished MS. University of Cambridge.
Trudgill, P.
1974 *The Social Differentiation of English.* Cambridge: Cambridge University Press.
Zeman, E. C.
1970 "Catastrophic Theory", *Scientific American* 234/4, 65-83.

Brian Lewis Butterworth (b. 1944) is University demonstrator in Psychology at the University of Cambridge, England. His major research interest is concerned with cognitive processes as reflected in speech production. Among his major publications are: "Hesitation and Semantic Planning in Speech" (1975); "Short-term Memory Impairment and Spontaneous Speech" (with T. Shallice, to appear); and "Speech and Interaction in Sound-Only Communication Channels" (with R. Hine and K. Brady, 1977).