PROBLEMS IN THE PHILOSOPHY OF LINGUISTICS

by

PAUL MARTIN MELLEMA

A.B., Calvin College

(1968)

SUBMITTED IN

PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF

TECHNOLOGY

June, 1973

Department of Philosophy, May 16, 1973

Certified by

Thesis Supervisor

Accepted by

Chairman, Departmental
Committee on Graduate Students

ABSTRACT

Problems in the Philosophy of Linguistics
Paul Martin Mellema
Submitted to the Department of Philosophy
on May 16, 1973 in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The three chapters of this thesis deal with three distinct problems in the philosophy of linguistics. Chapter 1 examines the theory of case grammar, Charles Fillmore's proposed revision of the standard theory of transformational grammar. Grounds are given for rejecting Fillmore's semantic, syntactic, and methodological arguments in support of case grammar, and some semantic and syntactic arguments against case grammar are advanced.

Chapter 2 concerns Chomsky's claim that children have innate knowledge of the principles of universal grammar, and make use of that knowledge in learning languages. The author rejects an argument to the effect that innate knowledge would require, but not have, justification, as well as an argument to the effect that certain of our purported innate beliefs lack truth-value, and fail to qualify as innate knowledge for that reason. Certain difficulties are seen, however, in the implication that we innately believe the principles of universal grammar. The Appendix questions the explanatory value of the tacit-knowledge theory.

Chapter 3 defends rationalist claims about the psychology of language acquisition. (Chapter 2, by way of contrast, is about rationalist claims in epistemology.) It has been charged that rationalism in learning theory generates a vicious infinite regress; that rationalist theories of learning are empirically indistinguishable from empiricist learning theories; and that a new, enriched theory of inductive inference invalidates the rationalist critique of empiricist learning theory. All three of these charges are rejected in Chapter 3.

Thesis Supervisor: Sylvain Bromberger
Title: Professor of Philosophy

Dedication

To God, with gratitude for its completion

Table of Contents

Chapter 1

A BRIEF AGAINST CASE GRAMMAR

In this paper I wish to review some of the arguments which Charles Fillmore has put forward in support of the theory of case grammar. I hope that I have not neglected any argument which case grammarians regard as weighty and significant, and that none of the arguments I discuss will strike them as insignificant or readily dispensable. Fillmore's writings are typically not polemical; they tend simply to explain how this or that linguistic phenomenon may be described within the framework of case grammar, without examining in detail the question of whether (or how economically) the standard theory of transformational grammar could describe the same phenomenon. Fillmore leaves his readers to draw their own conclusions about what constitutes evidence for case grammar and against the standard theory, and about the relative importance of the various arguments for case grammar. If my judgment in these matters has been faulty, I rely on the advocates of case grammar to set the record straight.

Fillmore has argued that "subject/object" grammars suffer from certain semantic shortcomings. I shall argue, in section I, that Fillmore has ignored certain crucial aspects of Katzian semantic theory, and that when these aspects of semantic theory are taken into account, the alleged deficiencies disappear. I shall also point out two sorts of defects in the semantic theory associated with case grammar. Section II deals with various syntactic arguments intended to show that the theory of case grammar is preferable to the standard theory. Finally, in section III, I shall discuss a sort of methodological argument for case grammar, to the effect that case grammars (unlike those provided under the standard theory) properly distinguish facts about semantic structure from facts about surface syntax.

I. CASES IN SEMANTICS

It is a well-known feature of the standard theory that it uses tree graphs to present two kinds of information about deep structures. In addition to providing information on the grammatical categories of the various constituents in a sentence, a tree diagram indicates what grammatical relations hold between constituents. For example, besides

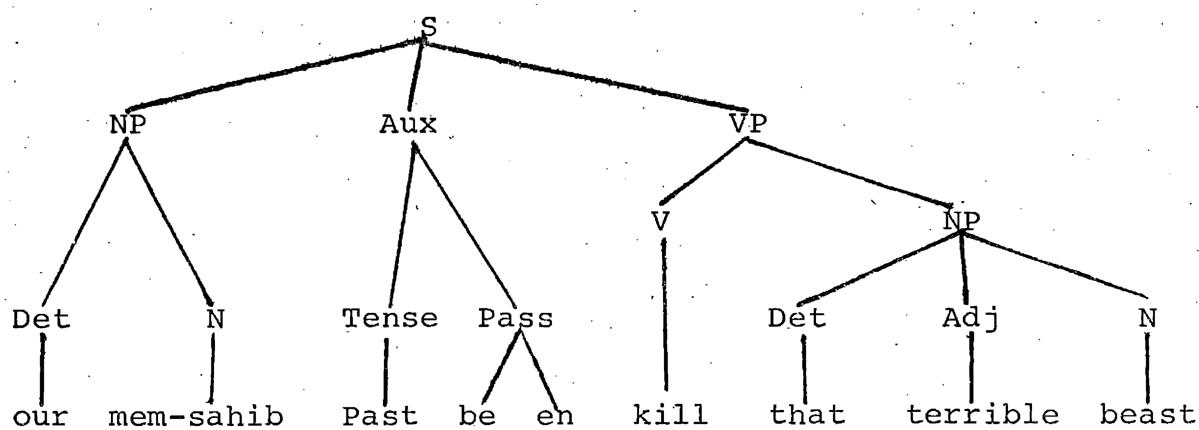

Figure 1

telling us that the phrase *our mem-sahib* is a noun phrase, Figure 1 tells us that this phrase bears the subject relation to the VP in the sentence there represented. The phrase *that terrible beast* can similarly be recognized as the direct object of the sentence whose deep structure is represented by this tree diagram.

This kind of relational information is derivable from tree graphs by virtue of certain definitions laid down by Chomsky

(1965, 69-74). One of these definitions states that the subject of a sentence is that noun phrase which is dominated by the NP node which lies directly under the node S; another definition tells us that the NP node immediately dominated by the node VP is the direct object NP node. Thus if a sentence is to have a deep structure subject, in Chomsky's sense, there must be a node NP which is immediately dominated by S; similarly, if the deep structure tree of a sentence does not contain a node NP which is directly dominated by VP, then that sentence does not have a direct object in Chomsky's sense.

The principal function of grammatical relations in deep structure, under the standard theory, is to direct the process of semantic interpretation. Without information

(1) That terrible beast was killed by our mem-sahib.

concerning grammatical relations in deep structure, we would be unable to determine whether it was the lady or the tiger that perished in the encounter described by sentence (1). For purposes of semantic interpretation, the fact that the phrase <u>that terrible beast</u> is the topic, or surface structure subject, of sentence (1), is of little consequence.

(2) *Him was killed by she.

(3) He was killed by her.

Information about grammatical relations in surface structure is important in assigning the proper case features or morphemes to constituents in surface structure, as sentences (2) and (3) illustrate, but the standard theory claims semantic relevance primarily for grammatical relations at the level of deep structure.

Now Fillmore has argued that even at the level of deep structure, the grammatical relation terms _subject_ and _direct object_, as defined by Chomsky, are inadequate to the needs of a semantic theory for English. First, Fillmore (1966, 363) observes that the deep subject of one sentence may

(4) The door will open.[1]

(5) The janitor will open the door.

have the same semantic role as the deep object of another sentence. In sentences (4) and (5), for example, "there is a semantically relevant relation between _the door_ and _open_ that is the same in the two sentences, in spite of the fact that _the door_ is the subject of the so-called intransitive verb and the object of the so-called transitive verb". To anticipate, the semantic role of _the door_ in both these sentences is what Fillmore terms the Objective case. Evidently the distinction between subject and object in deep structure does not always reflect a semantic difference. Otherwise put, if we retain this distinction in deep structure, then the semantic component will have to neutralize this distinction when it interprets pairs of sentences like (4) and (5).

On the other hand, Fillmore argues that both the

(6) a. The boy slapped the girl.

b. The boy fell down.

c. The boy received a blow.

d. The boy has a toothache.

e. The boy has blue eyes.

f. The boy [= his appearance] shocked me.[2]

(7) a. I smashed the pumpkin.

b. I grew the pumpkin.

c. I like the pumpkin.

d. I imagined the pumpkin.

e. I made the pumpkin into a mask.

f. I made a mask out of the pumpkin.

subject relation and the direct object relation may correspond to any of a number of distinct semantic relations, depending on the verb and on other constituents in the sentence. Commenting on sentences (6) a-f, Fillmore (1970, 252-253) writes: "... the semantic role of deep structure subjects appears not to be univocal ... The involvement of the entity named by the subject NP ... appears to be quite different in each case ... there appears to be no common notional property of 'subjectness' which semantic descriptions of these sentences can exploit." Similarly, sentences (7) a-f are intended to show that "the direct object relation is not semantically univocal" (Fillmore 1970, 255).

To anticipate again, the subject NP _the boy_ of (6) is semantically an Agent in (6) a, an Object in (6) b, a Dative or Experiencer in (6) d, and an Instrument in (6) f. The direct object of (7) a-e, _the pumpkin_, is Objective in (7) a, but Factitive or Goal in (7) b. Each of these two grammatical relations, then, hides a multiplicity of distinct semantic roles. To say of the phrase _the boy_ only that it bears the subject relation to each of the sentences (6) a-f is to obscure the diversity which in fact characterizes the relations between this phrase and its containing sentences on the semantic level.

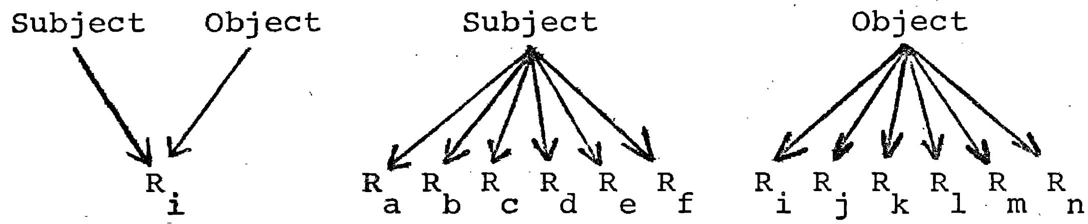

Figure 2

Figure 2 provides a schematic summary of Fillmore's argument thus far. He has presented evidence that the semantic component, in mapping grammatical relations into semantic ones, must on some occasions map the grammatical subject relation and the grammatical object relation into exactly the same semantic role, say R_i. Moreover, it appears that the subject relation may be mapped into any of a half-dozen or so semantic roles, and the same applies to the object relation.

Having shown that the mapping from grammatical relations to semantic relations is neither one-to-one nor functional, Fillmore proposes to change the form of the base rules so as to mark the semantic roles of constituents directly in deep structure. Deep structures will no longer contain subjects or objects in Chomsky's sense, so no formal mechanism will be needed to map these grammatical relations into the semantic roles, during the process of semantic interpretation. Specifically, Fillmore (1968, 24)

(8) a. S $\rightarrow$ NP + Aux + VP

b. VP $\rightarrow$ V (NP)

(9) a. S $\rightarrow$ Modality + Proposition

b. Proposition $\rightarrow$ V + C_1 + ... + C_n

proposes to replace the usual phrase structure rules (8) a-b with the rule (9) a and the rule schema (9) b. Clearly, no sentence generated by Fillmore's rules will contain a subject in deep structure, in Chomsky's sense, since by rule (9) a the node S will never immediately dominate an NP node. Similarly, Fillmore's base will not generate any deep structure objects, since Chomsky defined the object relation in terms of the node VP, and this node will not appear at all in any of Fillmore's deep structures.

The rule schema (9) b calls for a little explanation. Each of the C_i's in this rule schema stands for one or another case symbol, representing one of the semantic roles. The schema tells us that in a deep (case) structure, the node Prop immediately dominates the node V, followed by a string of one or more case symbols. No case symbol may occur more than once directly under a given Prop node.[3]

The list of cases now available for use in connection with rule schema (9) b is as follows: Agent, Experiencer, Instrument, Object, Source, Goal, Location, Time, and Path (Fillmore 1971, 42, 50-51). The order of cases on this list is significant; "the left-to-right order of the cases in the deep structure representation of individual sentences" must reflect the order imposed by this list, since in particular sentences it is "the leftmost noun phrase in the list" which becomes the surface subject of the sentence, via "the subject selection process" (Fillmore 1971, 37).

I shall have more to say about this "subject choice heirarchy" in section II of this paper.

Fillmore's proposed changes in the structure of the base, and in particular the introduction of semantic case nodes in deep structure, seem to be offered as a remedy for a defect which Fillmore sees in the treatment of semantics under the standard theory. Any semantic theory must, according to Fillmore, provide complete information about the semantic roles played by NP's and embedded S's, for any sentence in the language being described. The grammatical relations defined in deep structure, under the standard theory, do not provide us with information on semantic roles. Therefore, Fillmore seems to imply, we need to alter the structure of the base so that deep structures do provide information about semantic roles.[4]

This argument assumes, quite without justification, that if the standard theory cannot provide semantic role information simply on the basis of grammatical relations in deep structure, then it cannot provide such information at all. I wish to argue that the standard theory can provide complete role information at the level of semantic representation.[5] Lexical readings for verbs and prepositions, as well as grammatical relations, must be taken into account in determining the semantic roles of constituents.

To argue this thesis in detail, it will be necessary to examine Fillmore's definitions for the various cases. First, if I may proceed in violation of the subject selection heirarchy, consider the Experiencer case. This case is associated with verbs describing "a genuine psychological event or mental state" (Fillmore 1971, 42). Formerly, Experiencers were considered to be in the Dative case, "the case of the animate being affected by the state or action identified by the verb" (Fillmore 1968, 24), but Fillmore (1971, 42) "no longer confuse [s] selection restrictions to animates with true case-like notions".

To identify Experiencers in semantic representations of sentences, we must be able to recognize mental state and psychological event verbs. Now as it happens, Katz has proposed a semantic marker for verbs that seems admirably suited to this purpose. In a discussion of the verb <u>chase</u>, Katz (1967, 169) observes that in addition to being characterized as an activity verb, this verb must have in its reading the semantic marker (Physical), to distinguish chasing from <u>mental</u> activities, such as thinking and remembering. Thus it seems plausible to suppose that verbs which Fillmore would describe as taking an Experiencer, will contain the semantic marker (Mental) in a Katzian lexicon.

There remains the problem of deciding which argument of such a verb is the Experiencer, giving only a semantic

(10) I saw the rock

(11) *The rock saw me.

representation of the kind envisioned by the standard theory. The verb <u>see</u> takes an Experiencer, but it also takes an Instrument phrase referring to the stimulus, as in sentence (10). We must find some way to determine which NP designates the Experiencer, and it seems to me that Fillmore's definition of the old Dative case affords a suggestion worth pursuing. From the semantic anomaly of sentence (11), it is evident that the verb *see* requires that one of its arguments be semantically animate.[6] This fact will be reflected in the lexical reading for *see*, in

(12) *see*, (Event) (Mental) (Visual) (Perception of $\underset{\langle \text{Physical object} \rangle}{\overset{[NP,VP,S]}{X}}$ by $\underset{\langle \text{Animate} \rangle}{\overset{[NP,S]}{X}}$)

that the variable whose value is the semantic reading of the subject NP will have the selection restriction *⟨Animate⟩* underneath it, as in (12). Violation of this selection restriction results in the semantic anomaly of (11), where the semantic reading of the subject NP *the rock* does not include the semantic marker (Animate).

It is likely that every psychological verb will have in its lexical reading at least one argument that is marked as obligatorily animate. If, as I suspect, such verbs have exactly one argument so marked, then that argument will usually be the Experiencer in Fillmore's sense. Under circumstances shortly to be described, such arguments will be Agents. The verb *add* is a psychological verb, for example, but its obligatorily animate argument is an Agent rather than an Experiencer.

Next, let us consider how we might be able to recognize Location and Time phrases, without benefit of cases in deep

(13) The olive hit the windshield

(14) The operation lasted two hours

(15) *The olive hit two hours.

(16) *The operation lasted the windshield.

structure. The Location case is examplified by the windshield in (13), and the Time case by two hours in (14). The semantic anomaly of (15) and (16) calls attention to the fact that not all nouns are capable of appearing as heads of Location or Time phrases. Within the standard theory, the natural way to explain this fact is to posit (Time) and (Location) as semantic markers, and to indicate that the verb hit requires an object that includes the marker (Location) in its reading, on pain of semantic anomaly. A similar analysis applies to the verb last, and to prepositions like during, before, after, behind, inside, under, etc. Location phrases, then, will be NP's that are required to be semantically locative by the lexical reading of the verb or preposition involved, and similarly for Time phrases.

Agents and Instruments are both associated with causative verbs. The Instrumental case is "the case of the immediate cause of an event, or, in the case of a psychological predicator, the 'stimulus', the thing reacted to" (Fillmore 1971, 42). In former days, an Agent was defined as an animate cause, and an Instrument as an inanimate cause, but as I indicated above, animacy is no longer allowed to play a role in defining the cases. Fillmore (1971, 43-44) suggests that we might define the Agentive as the case of the principal cause of an event, but then he gives several reasons for rejecting this proposal, and lets the question drop, without coming to any conclusions about the proper definition of this case.

The amorphous character of Fillmore's present position in regard to the Agentive case makes it difficult to tell what would count as a satisfactory definition of this case within the vocabulary of the standard theory. Nonetheless, I believe that it is possible to make a beginning on the task, and the related job of devising a way to recognize Instruments. The semantic marker (X causes ...) has established itself in the vocabulary of Katzian semantics, and whenever this marker appears in the semantic reading for a verb, it is safe to conclude that the value of the variable X will be the reading of an NP or S which is in either the Instrumental or the Agentive case. Moreover, whenever this variable is marked to *require* a value which includes the semantic marker (Animate), we may feel quite confident in concluding that we are dealing with the Agent case, and not the Instrumental. I would further conjecture that whenever the value of the variable includes (Animate) *as a matter of fact*, then the phrase in question has an Agentive reading (perhaps in addition to an Instrumental reading).

There is, of course, another approach which the standard theory could take to the problem of distinguishing Agents from Instruments. The approach I have in mind is highly *ad hoc*, and represents no insight into the nature of agency beyond those provided by Fillmore's theory. Nonetheless, this "brute force" approach serves to illustrate a general formal point which has some significance, I think. To identify Agents, we might first introduce a new semantic marker, (X do), into the standard semantic theory. This new marker is to be slightly stronger than the marker (X cause), in some unspecified way. We then write the lexicon in such a way that the arguments of (do) are always Agents in Fillmore's grammar, and conversely. The formal point to be made is that any case notion can be defined over Katzian semantic representations, simply by introducing a suitably constrained new semantic marker. Whatever the insights of case grammar may be, these insights can be captured with no change in the structure of the base, and in particular, with no case symbols in deep structure.

The Source, Goal, and Path cases seem to me to be adjuncts of process verbs of motion. For this

$$\text{(17)}\ \underline{\text{open}},\ -\ [\_\_\text{NP}]\ ;\ \text{(Condition)}\left(\begin{matrix}[\text{NP},\text{S}]\\ \text{X}\\ \langle\text{Barrier}\rangle\end{matrix}\ \ \begin{matrix}\text{closed at}\\ \text{time } \underline{t}\end{matrix}\right),$$

$$\ldots,\ \text{(Condition)}\left(\begin{matrix}[\text{NP},\text{S}]\\ \text{X}\\ \langle\text{Barrier}\rangle\end{matrix}\ \ \text{open at } \underline{t} + \underline{n}\right)$$

reason, it may be worth our while to look briefly at the kind of lexical entry used for process verbs, under the standard theory. The intransitive verb open has a lexical entry roughly like (17);[7] the process of opening is represented as a sequence of conditions or states, beginning with a state in which some barrier is closed, and ending with a state in which that barrier is open. Each stage

skip next page!

of the process is ordered in time with respect to the other stages, by means of a time variable.

Fillmore (1971, 41) states that Source and Goal may be interpreted as "earlier and later locations, earlier and later states, or earlier and later time points", depending on the verb in the sentence. An embedded sentence in the Goal case may identify "the resulting state or event in a causative construction" (Fillmore 1971, 42). Goal is also "the end-result role of a thing which comes into existence as a result of the action identified by the predicator, as in _I wrote a poem_" (Fillmore 1971, 42).

(18) He walked from the dormitory to the chapel.

Sentence (18) provides an example of a locational Source phrase (_from the dormitory_) and a locational Goal phrase (_to the chapel_). Under the standard theory, the semantic readings for these two phrases would contain time variables. For example, the reading for _from the dormitory_ might contain simply the variable _t_, whereas the reading for the phrase _to the chapel_ might contain the time variable _t+n_. Since $t < t+n$, the time variables in the readings for these phrases makes it clear that the walker referred to was in the dormitory before he was in the chapel.

The device of time variables does not suffice, however, to identify _from the dormitory_ as a Source phrase. In a sentence with the verb _walk_, a Source phrase identifies the location (or time) at which the walking begins, and for all the time variables have told us thus far, the walker of sentence (18) was already in transit when he passed through the dormitory. To identify the dormitory as the initial point of his walk, we must look at the semantic representation of the sentence as a whole. Since _walk_ is a process verb, the semantic representation for the sentence will be a sequence of state descriptions. And just as (17) shows an initial state in which some barrier is closed, so the semantic interpretation for (18) will show an initial state in which the walker's location is the dormitory. What brands _from the dormitory_ as a Source phrase, then, is the fact that its semantic reading appears in the first of the state descriptions which make up the semantic representation for the whole of sentence (18).

(19) I persuaded him to fill my tank.

Another kind of Goal is exemplified in sentence (19), where the clause underlying _him to fill my tank_ is the complement of the causative verb _persuade_. Now the semantic reading for _persuade_, under the standard theory, will surely contain the semantic marker (X causes Y), and the variable Y in this reading will be replaced by the semantic reading for the underlying clause _he fill my tank_. But once the reading for this clause has taken its place as an argument of the semantic marker (X causes Y), we know that this clause is a Goal, by Fillmore's definition.

$$\text{(20) write, } + \left[\_\_\,\mathrm{NP}\right] ;\ \Big(\underset{\langle \mathrm{Inscription} \rangle}{\overset{[\mathrm{NP,VP,S}]}{X}} \text{ did not exist before } \underline{t}\Big)$$

$$\&\ \Big(\Big(\underset{\langle \mathrm{Human} \rangle}{\overset{[\mathrm{NP,S}]}{X}} \text{ produces symbols on a surface at } \underline{t}\Big)$$

$$\text{causes } \Big(\underset{\langle \mathrm{Inscription} \rangle}{\overset{[\mathrm{NP,VP,S}]}{X}} \text{ exists after } \underline{t}\Big)\Big)$$

Finally, consider Fillmore's last kind of Goal, the kind exemplified by *a poem* in *I wrote a poem*. The semantic interpretation of this sentence will be guided by the lexical reading for the verb *write*, given in crude form above as (20). An inspection of (20) shows that the semantic reading for the direct object of *write* (in this case, *a poem*) appears twice in the semantic representation of the sentence. Once, the semantic reading for *a poem* appears in the context ... *did not exist before t*; later, the reading for *a poem* occurs in the context ... *exists after t*. Thus the semantic representation for the sentence *I wrote a poem* tells us that the poem is "a thing which comes into existence as a result of the action identified by the predicator" (Fillmore 1971, 42). Every kind of Source and Goal recognized by Fillmore can thus be identified on the basis of information provided by semantic representations under the standard theory.

As one might suspect from its name, the Path case is

(21) He walked down the hill across the bridge through the pasture to the chapel.

used with verbs of motion; Fillmore (1971, 51) gives as an example the phrase down the hill across the bridge through the pasture in sentence (21). This phrase gives us a temporal sequence of locations within which the activity of walking took place, and this sequence can easily be represented through the use of the format for representing the sequence of stages in process verbs. An NP is in the Path case, then, just in case it gives the location within which motion takes place, without implying that the motion began or terminated there. (Initial and final locations are Sources and Goals, respectively.)

This leaves only the Objective case, and by Fillmore's (1971, 42) own admission, this is a wastebasket case. This means that if an embedded S or NP is not in any of the other cases already defined, then it is an Object. But since we already know how to recognize all the other cases, we can easily identify Objects as well. If an NP or S does not belong to any of the other cases, we know that it belongs in the "wastebasket" Objective case.

Fillmore argued successfully that the grammatical relations defined by standard deep structure representations do not suffice to provide complete information about the semantic roles of constituents. We have just seen, however, that standard semantic representations provide just as much information about semantic roles as is provided by Fillmore's deep structure graphs with case nodes. Thus far we have found no reason to suppose that the use of cases in deep structure will remedy any deficiencies in the semantic theory associated with the standard theory of transformational grammar. I wish to now argue that the theory of case grammar meets with certain semantic difficulties that do not attend the standard theory.

(22) a. John smeared paint on the wall.

b. John smeared the wall with paint.

(23) a. Bees are swarming in the garden.

b. The garden is swarming with bees.

Sentences (22) a and (22) b receive the same case structure, according to Fillmore (1968, 48). There is, however, a semantic difference between these two sentences.

(24) Most of the wall didn't get any paint on it.

Sentence (22) a is logically consistent with (24), but (22) b is not consistent with (24). Similarly, sentences

(25) Most of the garden has no bees in it.

(23) a-b are derived from identical case structures, but (23) a is consistent with (25), while (23) b is not. Steve Anderson (1971, 389) describes this semantic difference by saying that the Locative phrases in (22) b and (23) b receive a holistic interpretation; in (22) a and (23) a, the interpretation of the Locative phrases is partitive.

On the basis of (22) and (23), one is tempted to think that a Locative phrase receives a holistic interpretation if and only if that phrase occupies either the subject position or the direct object position.[8] Anderson (1971,

(26) a. The press secretary read his prepared speech.
b. The press secretary read from his prepared speech.

(27) a. John painted Bill's portrait this morning.
b. John painted on Bill's portrait this morning.

391) provides some examples that show that the relationship between holistic interpretation and grammatical relations is not limited to NP's in the Location case. Presumably _his prepared speech_ is Objective in both (26) a-b, but this phrase receives a holistic interpretation only in (26) a, where it functions as the direct object. The Goal phrase _Bill's portrait_ in (27) a-b receives a holistic interpretation only in (27) a, where it occupies the object position.

Now case grammars treat grammatical relations as surface phenomena, so it is natural for Fillmore (1968, 48-49n) to suggest that holistic interpretations be assigned by some process of surface structure interpretation.

(28) The garden is where bees are swarming.

(29) It is bees that the garden is swarming with.[9]

However, the generalization formulated one paragraph back is false in both directions, at the level of surface structure. In sentence (28), the garden does not receive a holistic interpretation, though it is the surface subject of the sentence. In (29), the garden is in neither subject nor object position in the sentence, but still it requires a holistic interpretation. Other sentences cited by Anderson (1971, 388, 390) suggest that no generalization about holistic interpretations can be stated in terms of grammatical relations in surface structure.[10]

We have seen that the semantic difference between (22) a and (22) b cannot be systematically accounted for in deep structure, nor in surface structure, under the theory of case grammar. Anderson (1971, 395) argues that there is only one level in the derivation of sentences at which Fillmore can state the appropriate generalization. This level can be defined as the output of the transformation which freely selects either paint or the wall as direct object, in the sentences of (22).[11] Let us call this transformation the object selection rule, and let us ask whether the grammar of English ought to contain this rule. If there are good reasons to reject this rule, then case grammar will be deprived of the only level at which it is possible to assign holistic readings correctly.

Steve Anderson (1971, 393-395) notes that while smear undergoes the object selection rule freely, there are other

(30) a. *John covered jelly on the slice of bread.

b. John covered the slice of bread with jelly.

(31) a. John threw paint on the canvas.

b. *John threw the canvas with paint.

verbs which do not. Like smear, both cover and throw take Instrumental and Location adjuncts. The Instrumental argument of cover cannot be chosen as object, however, and the Locative argument of throw is similarly barred from objecthood; witness sentences (30) a and (31) b. As long as a grammar contains Fillmore's object selection rule, cover and throw will have to be marked in the lexicon as exceptions to that rule. By including this rule in the grammar, Fillmore complicates the lexical entries for certain verbs. This complication can be eliminated by handling object selection in the base.

We have just seen that there are syntactic grounds for having direct objects in deep structure. There is also a semantic motivation for this, of course; we saw earlier that holistic and partitive interpretations can be predicted by a very simple rule, if that rule can refer to grammatical relations as defined in deep structure. Anderson's paper shows that holistic and partitive interpretations cannot be assigned on the basis of any well-motivated level of syntactic representation in a case grammar.

Before going on, I should like to discuss a certain objection which might be made to Anderson's argument. If we drop the rule of object selection, and assume that the object of a sentence will always be indicated in deep structure, this means that the lexicon will have to contain two lexical entries for the verb smear. One of these "two" verbs will take a locative object, and the other will not. Presumably, there are many verbs like smear. Consequently, the lexicon will (under Anderson's analysis) contain many "elaborate and unexplained examples of homonymy".[12] Otherwise put, this argument charges Anderson with treating every verb as irregular in respect to the object selection process. Fillmore's analysis at least has the virtue of treating some verbs as regular, even if cover and throw turn out to be irregular.

This objection, like Anderson's argument, appeals to our concern for simplicity in linguistic descriptions. According to the objection, a case grammar of English is better because it is shorter (has less lexical entries) than a non-case grammar. But according to Chomsky (1965, 42-44), the complexity (length) of a rule ought to reflect its naturalness from a psychological point of view. Long rules ought to be more difficult to learn, and to use in speech production and perception, than short rules. If Fillmore means for case grammars to have psychological reality, then it may be appropriate to choose between Fillmore's analysis and Anderson's by means of psycholinguistic experimentation, rather than by appeals to simplicity.

Unfortunately, it is difficult to say just what sort of experimentation would be relevant to a choice between these two analyses, because Fillmore has so far as I know never made it clear just what sort of psychological reality he claims for case grammars, and for his version of universal grammar. I wish to discuss two possible positions on this issue, which Jerry Fodor (1971) terms the strong psychological reality position and the weak psychological reality position. I shall argue that if Fillmore holds either of these two positions, then psycholinguistic considerations can in principle be brought to bear on the choice between Anderson's analysis and Fillmore's. Of course, if Fillmore does not claim psychological relevance for case grammars, or only some very weak sort of psychological relevance, then the psychological considerations I point out will be altogether beside the point.

According to adherents of the "weak psychological reality" position, a hearer understands a sentence by computing the structural descriptions of that sentence, but the speaker does not necessarily use the rules of the grammar in computing this structural description. On this view, a speaker must determine the case structure (deep structure) of a sentence in order to understand the sentence, but perhaps he does this by means of a set of heuristics that operate on surface structures. On the other hand, those who hold a "strong psychological reality" position believe that speakers understand sentences by computing their structural descriptions, and that this computational process is carried out by applying the rules of the grammar. Analysis-by-synthesis models of speech recognition thus embody the "strong psychological reality" position.

Suppose, now, that Fillmore accepts a strong view of the psychological reality of grammars. Anderson's argument showed that cover and throw have more complex lexical entries than does smear, since these verbs (unlike smear) impose idiosyncratic constraints on the application of a certain transformation rule (viz., object selection). Since cover and throw exhibit these transformational irregularities, sentences containing these verbs ought to be harder to understand than sentences containing regular verbs like smear, all other things being equal.

Anderson's analysis, coupled with a strong view of psychological reality, yields an opposite prediction. For smear appears twice in the lexicon, whereas cover and throw each appear only once. To determine the meaning of a sentence containing smear, a hearer must decide which of the two lexical entries for this verb is appropriate to the interpretation of the sentence in question. On some occasions, he might happen to guess right on his first attempt. Such good luck is not to be expected all the time, however. Indeed, it might be possible to induce wrong first guesses, by exposing subjects to sentences with verbs like cover, then to a sentence in which smear has its throw-like meaning.

If we adopt the "strong psychological reality" view, then Fillmore's grammar predicts that _cover_ and _throw_ will pose greater difficulties to the hearer than _smear_; Anderson's proposal, on the contrary, predicts that sentences with _smear_ will on the whole be more difficult to process than sentences with either of the "univocal" verbs, _cover_ or _throw_.

Now let us suppose that Fillmore holds the weak view discussed by Fodor. This means that only structural descriptions, and not grammatical rules, have psychological reality, and the difficulty in understanding a sentence depends on the degree of divergence between surface structure and deep (case) structure. Now according to Fillmore, Goal phrases associated with _smear_, _cover_, and _throw_, do not appear in postverb position in deep structure; rather, it is the Instrument phrase which occurs immediately to the right of the verb in deep structure. Thus sentences containing _throw_ have the same order of constituents in surface structure as in deep structure. The same is true of sentences with _smear_ to which the object selection rule has not applied. Surface structure and deep structure show different constituent order, however, in sentences containing _cover_, and in sentences with _smear_ to which the object selection rule has (optionally) applied. On a "weak psychological reality" view, then, Fillmore's analysis implies that sentences with _throw_ should be easier to understand than sentences with *cover*, and that untransformed *smear* sentences should be easier to understand than sentences with *smear* to which the object selection rule has applied.

On the same view of psychological reality, Anderson's analysis implies that there will be no difference in perceptual complexity between the two uses of *smear*, and no difference between sentences with *cover* and sentences with *throw*. This is because Anderson does not posit any single order of constituents in deep structure. Rather, surface constituent order conforms to deep constituent order for *cover* as well as for *throw*, and for both uses of *smear*. For Anderson's analysis, the cleavage is still between *smear* on the one hand, and *cover* and *throw* on the other. This is because *cover* and *throw* both occur in only one kind of deep structure, thereas *smear* may appear in either of two kinds of deep structure context.

Whether Fillmore adopts a strong or a weak view on the psychological relevance of grammars, then, his analysis yields psychological predictions which are incompatible with the psychological claims which are implicit in Anderson's analysis. So far as I know, there is no evidence available to decide the issue, but since such evidence is in principle available, an appeal to simplicity seems somewhat slothful, if not downright irrelevant.[13]

page 32 precedes page 19

OUT OF ORDER

We have seen that the semantics of the standard theory is able to determine the semantic roles of constituents, in Fillmore's sense, and that the theory of case grammar is not able to provide an adequate treatment of the distinction between holistic and partitive Locatives. I wish now to argue that certain entailment relations can be handled in case grammar only at the cost of added complexity.

The entailment relations I have in mind are of two kinds. Some of them involve Agentive and Instrumental phrases; the others arise out of Goal phrases of the kind which Fillmore used to call Factitives. Causality is part

(32) a. Helen broke the window.

b. Helen caused the window to break.

(33) a. The ball broke the window.

b. The ball caused the window to break.

of the definition of both the Agentive and the Instrumental case. Since <u>Helen</u> is Agentive in sentence (32) a, this sentence entails (32) b; and since <u>the ball</u> is Instrumental in (33)a, (33) a entails (33) b. The Goal case is defined

(34) a. I wrote this poem today.

b. This poem did not exist before today.

to include NP's designating objects that are said to come into existence as a result of some process. In sentence (34) a, <u>this poem</u> is a Goal of this kind, and of course (34) a entails (34) b.

Now under the standard theory, the semantic reading for transitive <u>break</u> will play an important part in predicting the entailments from (32) a to (32) b, and from (33) a

$$\text{(35)}\quad \underline{\text{break}},\ +\ [\_\_\text{NP}]\ ;\ \Big(\Big(\ \underset{\left\langle \substack{\text{Physical}\\ \text{object}}\ \vee\ \substack{\text{Physical}\\ \text{event}} \right\rangle}{\overset{[\text{NP,S}]}{X}}\ \ \text{causes}$$

$$\Big((\text{Condition})\Big(\ \underset{\left\langle \substack{\text{Physical}\\ \text{object}} \right\rangle}{\overset{[\text{NP,VP,S}]}{X}}\ \ \text{whole at time } \underline{t}\Big),\ \ldots,$$

$$(\text{Condition})\Big(\ \underset{\left\langle \substack{\text{Physical}\\ \text{Object}} \right\rangle}{\overset{[\text{NP,VP,S}]}{X}}\ \ \text{broken at } \underline{t} + \underline{n}\Big)\Big)\ \Big)$$

to (33) b. This lexical reading will indicate that in sentences where <u>break</u> is used transitively, the subject of the sentence is said to cause a certain process to take place. By inspecting the sequence of conditions or stages spelled out in (35), we see that the process is one in which a physical object starts out whole, and ends up broken.

$$\text{(36)}\quad \underline{\text{break}},\ -\ [\_\_\text{NP}]\ ;\ (\text{Condition})\Big(\ \underset{\left\langle \substack{\text{Physical}\\ \text{object}} \right\rangle}{\overset{[\text{NP,S}]}{X}}\ \ \text{whole at } \underline{t}\Big),$$

$$\ldots,\ (\text{Condition})\Big(\ \underset{\left\langle \substack{\text{Physical}\\ \text{object}} \right\rangle}{\overset{[\text{NP,S}]}{X}}\ \ \text{broken at } \underline{t+n}\Big)$$

Consider next the lexical entry for intransitive <u>break</u>, which is given above as (36). The semantic marker (X causes Y), which appeared in the lexical entry for transitive <u>break</u>, is conspicuously absent from (36). However, (36) is virtually a carbon copy of the process portion of (35); both verbs describe a sequence of conditions, the first of which involves a whole object, and the last of which involves that same object in a broken state.

Looking back now to (32) and (33), we notice that the surface verb <u>cause</u> occurs in sentences (32) b and (33) b, along with the intransitive verb <u>break</u>. Consequently, the

(37) Helen causes ((Condition) (The window whole at <u>t</u>),

..., (Condition) (The window broken at <u>t+n</u>))

the semantic representation of (32) b will be rather like (37). But now consider (32) a. This sentence contains <u>break</u> in its transitive sense, and according to (35), the semantic marker (X causes Y) is part of the meaning of transitive <u>break</u>. Consequently, even though (32) a does not contain an occurrence of the surface verb <u>cause</u>, the semantic representation of (32) a will be very much like (37).[14] The standard theory thus predicts the entailment from (32) a to (32) b in terms of the close similarity between the semantic representations of these two sentences.

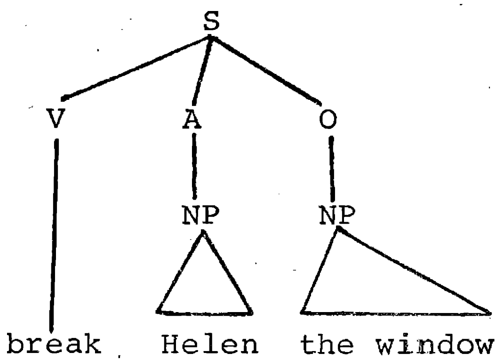

Figure 3

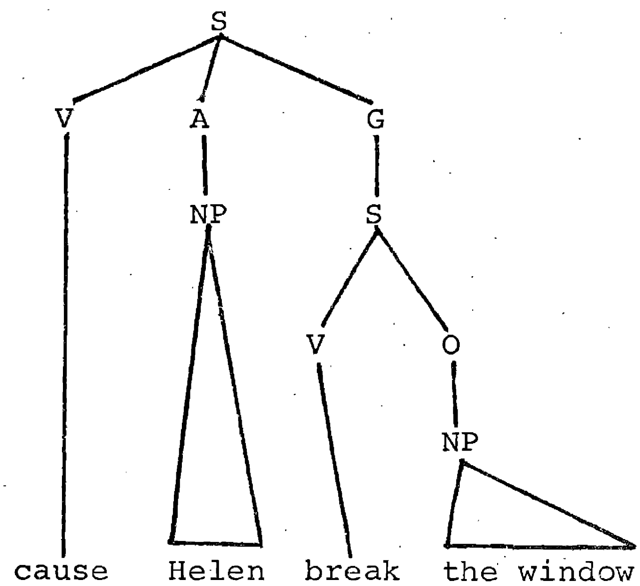

Figure 4

How might a case grammar account for the fact that (32) a entails (32) b? (The entailment from (33) a to (33) b can presumably be handled, _mutatis mutandis_, in the same way.) I shall assume that the case structures of (32) a-b are roughly those given in Figures 3 and 4, respectively.[15] Now a natural way to proceed would be to devise some rule for converting Figure 3 into a structure resembling Figure 4. Specifically, this new structure should contain an explicit occurrence of _cause_ (or the associated semantic marker), just as Figure 4 does. Since Fillmore is inclined to define the Agentive case partly in terms of causality, it is also natural to suggest that this rule of semantic interpretation be sensitive to the presence of Agent nodes.

The details of this rule need not concern us. The point is simply that some such rule is needed. Fillmore has given us some informal remarks about Agents, Instruments, and causality, but in order to predict the entailments in question here, some additional formal apparatus is needed. This formal apparatus must be reckoned in the cost of adopting the theory of case grammar. In the standard theory, these entailments are predicted without the need for any special interpretation rules, just as subjects and objects are selected without the need for subject and object selection transformations. The use of cases in deep structure thus appears to require the addition of extra semantic, as well as syntactic, paraphenalia. This suggests that the cases obscure, rather than clarify, linguistic structure.

?

A case grammar also needs an extra rule of semantic interpretation in order to capture entailments of certain sentences containing Goal phrases. For example, sentence (34) a above contains a Goal phrase (<u>this poem</u>) of the kind Fillmore used to call Factitives, and for this reason, (34) a entails (34) b. Presumably the case structure for

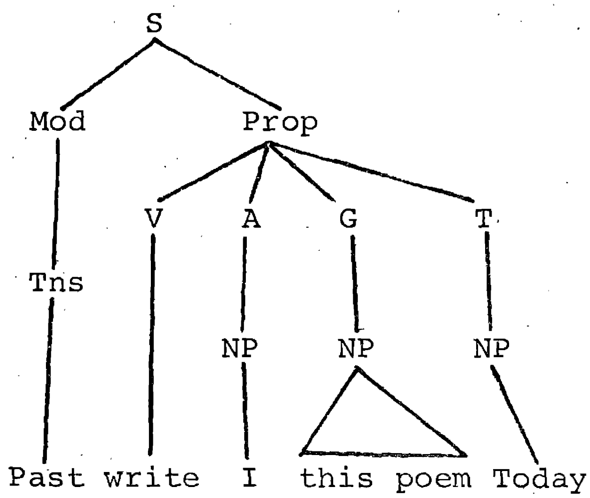

Figure 5

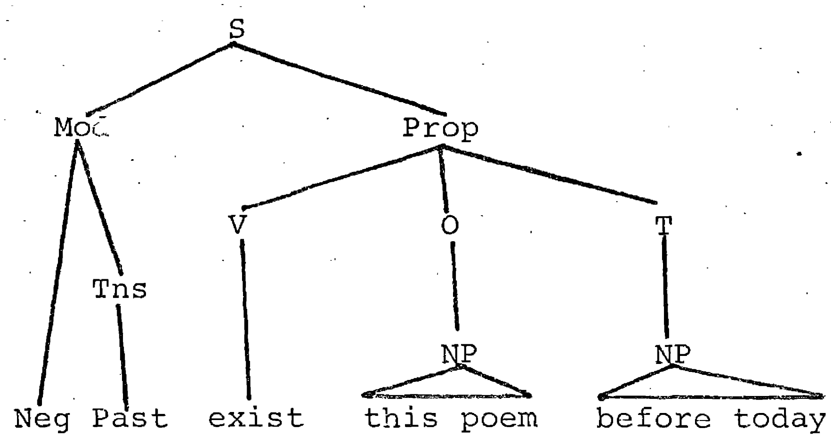

Figure 6

sentence (34)a can be represented by the tree graph of Figure 5, and the case structure of (34) b by the tree in Figure 6. Now the entailment relation between (34) a and (34) b must be predicted in some way, but as they stand, Figures 5 and 6 surely do not make it obvious that such an entailment relation holds. To exhibit the entailment relation, we must presumably show that the case structure of (34) b is derivable from the case structure of (34) a in some way.[16] But the rule for carrying out such a derivation must be quite complex, for note how different Figure 6 is from Figure 5: The verb of Figure 5 has been replaced in Figure 6 by another verb altogether; the Modal constituent of Figure 6 contains a Negation element not to be found in Figure 5; the Time phrase of Figure 6 contains the preposition <u>before</u>, which was absent from the Time phrase in Figure 5; and the Agent phrase of Figure 5 disappears from Figure 6.

Figures 5 and 6 differ more radically than do Figures 3 and 4, which we examined a moment ago. It would appear, then, that the semantic rule we need for predicting the contraexistential entailments of sentences with factitive Goal phrases must be more complex than the rule needed to predict the entailments of sentences containing Agents and Instruments. Fillmore does not seem to have appreciated the need for either of these semantic rules in a case grammar. Within the standard theory, however, no extra machinery is needed to show that sentence (34) a entails sentence (34) b.

$$\text{(38)}\quad \underline{\text{write}},\ +\left[\_\_\text{NP}\right]\ ;\ \Big(\underset{\langle\text{Inscription}\rangle}{\overset{[\text{NP},\text{VP},\text{S}]}{\text{X}}}\ \text{did not exist before } \underline{t}\Big)$$

$$\&\ \Big(\Big(\underset{\langle\text{Human}\rangle}{\overset{[\text{NP},\text{S}]}{\text{X}}}\ \text{produces symbols on a surface at } \underline{t}\Big)$$

$$\text{causes}\ \Big(\underset{\langle\text{Inscription}\rangle}{\overset{[\text{NP},\text{VP},\text{S}]}{\text{X}}}\ \text{exists after } \underline{t}\Big)\Big)$$

For recall the lexical entry for the verb <u>write</u>, given above as (20) and repeated here for convenience as (38). According to (38), sentences having <u>write</u> as their main verb receive semantic representations which have the form of a conjunction of two semantic structures. The first conjunct of their semantic representations is always a reading for a contraexistential statement, such as (34) b. In explaining the entailment from (34) a to (34) b, then, the standard theory appeals to the semantic representations of these sentences, and to the rule of simplification: from a statement of the form <u>p&q</u>, to infer a statement of the form <u>p</u>. This rule of inference is of course familiar from the sentential calculus.

In summary, the semantic case for case does not succeed in impeaching the standard theory, since all the semantic information provided by case labels is also available in the semantic representations generated under the standard theory. On the other hand, the standard theory is conspicuously better than case grammar at predicting whether a given Locative (or Objective, or Goal) phrase is to receive a holistic interpretation. Case grammars are also at a disadvantage in accounting for certain entailments of sentences containing Agents, Instruments, and Goals. To exhibit these entailments, a case grammar must be supplemented by (presumably universal) rules of semantic interpretation that are not needed in the standard theory. On balance, the semantic evidence seems to weigh against case grammar, rather than in its favor.

II. CASES IN SYNTAX

In this section, we shall consider four syntactic arguments for case grammar. First, Fillmore (1968, 31) has argued that in a case grammar, it is possible to eliminate Lakoff's (1966) feature [±stative], without losing the insights statable in terms of that feature. Second, Fillmore (1971, 37) maintains that there is a universal relationship between the cases appearing in a sentence, and the normal choice of surface subject in that sentence. Third, Fillmore (1968, 15) appears to regard the assignment of prepositions and other surface case markers as a task for which case grammars are particularly well suited. Finally, Fillmore (1968, 29-30) argues that case grammar permits certain economies in the lexicon.

With regard to stativity, Fillmore (1968, 31) inquires "whether Lakoff's features are primitives in the lexical entries for verbs, or whether they permit reduction to case concepts". He suggests that rules sensitive to non-stativity may just as well be stated in terms of the Agentive case. "The transformation which accounts for the 'true imperatives' can apply only to sentences containing A's", and the selection of a Benefactive phrase, or of the Progressive aspect, is permissable only in sentences containing Agents, Fillmore suggests.

I wish to present three arguments relating to this suggestion. First, I shall argue that if non-stativity can be defined away in terms of the Agentive case, then the standard theory can also do without the feature [$\pm$stative] in its lexical entries for verbs. Second, I shall present several counterexamples to the claim that non-stative verbs are able to take Agentive arguments. Finally, I shall cite some evidence which suggests that the feature [$\pm$stative] does not play a role in the grammar of English; if this is correct, then it is no virtue for a theory to be able to define this feature away.

Let us assume for the sake of argument that there are important generalizations to be stated in terms of the feature [$\pm$stative], and that these generalizations can also be stated in terms of the presence or absence of an Agent phrase in a sentence. This will count in favor of the theory of case grammar, and against the standard theory, only if the standard theory is compelled to incorporate [$\pm$stative] as a primitive term. But in section I, we noted that the Agentive case was in part definable in terms of the semantic marker (Cause), and I conjectured that if a verb contains this marker in its lexical reading, then that verb can have an Agentive argument. Let us assume that this conjecture is correct.[17] What follows, then, is that a verb is [-stative] just in case its semantic reading includes the semantic marker (Cause). The standard theory can use this semantic marker to dispose of Lakoff's feature, just as case grammar uses the Agentive case to define the feature away.

It may be objected that transformations, such as imperative formation, cannot be stated in terms of a semantic marker such as (Cause). Semantic markers are not, after all, part of deep structure, and transformations are generally thought to apply to uninterpreted deep structures.

Katz (1970, 231), in discussing the controversy between generative and interpretative semantics, has argued that the standard theory is "independent of any claims about the type of information transformations utilize". He points out that under the standard theory, "transformations can be written that apply either to underlying phrase markers or semantically interpreted underlying phrase markers, and the process of transformational development can be made logically posterior to the process of interpreting underlying phrase markers". If need be, the standard theory can define away the feature [±stative] . All that is needed is a definition of the Agent case, in terms of semantic markers. Such a definition was sketched in section I.

But as it turns out, Fillmore was mistaken in supposing that non-stativity could be defined as presence of an Agent.[18] Several of Fillmore's favorite verbs are counterexamples to this claim. One of the principal marks of a non-stative verb is its ability to occur in true imperative sentences.

(39) Die gloriously for the fatherland.

(40) Fall on your knees when you approach the emperor.

(41) Rotate quickly on the ball of your left foot.

(42) Don't break when they question you.

(43) Imagine the Jolly Green Giant dressed as Santa.

(44) Don't believe a word the recruiter told you.

Sentences (39)-(44) appear to be true imperatives, yet the verbs involved do not have Agentive arguments, even in deep structure. Die differs from kill, according to Fillmore (1968,30), precisely in that die cannot take an Agent, whereas kill can. However, die is [-stative] by two tests. Sentence (39) is an imperative, and it also contains a Benefactive phrase. Fall, in (40), is a verb of motion, and since "the Object case is that of the entity which moves" (Fillmore 1971, 42), the deleted subject of (40) is Objective rather than Agentive. Similarly, the intransitive forms of rotate and break, which occur in (41) and (42), do not take Agents, but Objectives. Finally, imagine and believe, which occur in (43) and (44), are psychological-state verbs, and as such they take Experiencer subjects, not Agents.

We have just seen that verbs may be non-stative, without taking an Agentive argument, and presumably without having the semantic marker (Cause) in their readings. This shows that case grammar is not able to define away Lakoff's feature in the way Fillmore suggested, and that the standard theory cannot use (Cause) to do the work of this feature. I wish now to argue that the feature [+stative] cannot be used to do all the syntactic work Lakoff had in mind for it. Lakoff posited this feature in order to account for the occurrence of nine or so different phenomena, which he believed to be related. He cited evidence to show that if a verb could appear in true imperative constructions, then it could also appear in the Progressive aspect, with Benefactive phrases, with the do-so construction, etc. He further hypothesized that if a particular verb rejected one of these constructions, it rejected them all.

Closer examination reveals, however, that verbs do not always accept or reject these constructions as a block.

(45) *Ben is believing what the recruiter told him.

Sentence (43) shows that believe can occur in true imperatives, but believe rejects the Progressive aspect, as illustrated in (45). Another verb that seems to share these properties

(46) a. Don't despise the gypsy's advice.

b. *I am despising her suggestion.

is despise (cf. (46) a-b), and there may be others.

Mike Harnish (personal communication) has pointed out a class of verbs that violates Lakoff's hypothesis in a different way. Certain success verbs behave like non-statives

(47) Judy found a gem, and Horace did so too.

(48) Carl discovered the proof, and Susan did so too.

(49) *Judy is finding a gem.

(50) *Carl is discovering the proof.

with respect to the do-so construction, but like statives, they reject the Progressive aspect. As we can see from (47)-(50), find and discover are two such verbs.

The existence of verbs like believe and find shows that a single feature is not enough to account for the behavior of verbs with respect to the imperative, Benefactive, Progressive, do-so, and other constructions. These constructions are apparently not as closely related as Lakoff at first thought. But the utility of [±stative] depends heavily on the supposition that all these constructions are closely related. Even if Fillmore's case-theoretic definition of non-stativity had been valid, it would not have constituted a particularly impressive vindication of case grammar.

We turn next to Fillmore's claim that the process of subject selection is predictable in terms of a universal case heirarchy. If the standard theory is unable to capture the generalization underlying this claim, and if the claim is true, then case grammar would enjoy an advantage over the standard theory. I shall argue that it is possible to state Fillmore's hypothesis about the case heirarchy, using terms available in the standard theory. I shall also point to some rather dubious implications of the heirarchy hypothesis.

The case heirarchy is a ranking of the cases, in which the Agentive is the highest-ranked case. The Experiencer, Instrumental, Objective, Source, Goal, Location, Time, and Path cases follow, in that order. Fillmore's heirarchy hypothesis states that ordinarily, the constituent with the highest-ranked case becomes the subject in surface structure. Thus if a sentence contains an Agentive phrase, that phrase is normally chosen as the subject, whatever other cases may be present in the sentence. This is because an Agent outranks every other case. If a sentence has an Experiencer, but no Agent, then the Experiencer is the most likely choice for subject, and so on. If we do not choose the highest-ranked NP as subject, then we must indicate this deviation by using a passive verb form, in English and many other languages.

At first glance, this hypothesis does not appear to be statable within the standard theory. The standard theory recognizes neither cases nor the subject selection rule, but this hypothesis concerns the relationship between cases and the subject selection rule. But semantic interpretation, in the standard theory, includes a process which is roughly the inverse of the subject selection process. Fillmore's subject selection rule takes an NP with a particular semantic role, and moves that NP into subject position. Semantic interpretation, on the other hand, operates on the semantic reading of the subject NP in a sentence. The subject NP reading is moved into a particular spot in the lexical reading of the verb, and as we saw in section I, the semantic role of the subject is determined by the position it comes to occupy in the lexical reading of the verb. Arguments of the semantic marker (Cause) are either Agentive or Instrumental; obligatorily animate arguments of (Mental) verbs are Experiencers; and so on.

The place in a verb reading where the subject NP reading belongs is marked by means of a categorized variable. In (35), for example, we saw that the lexical entry for transitive *break* starts out: "(($\overset{[NP,S]}{\underset{\left\langle \substack{\text{Physical}\\ \text{object}} \vee \substack{\text{Physical}\\ \text{event}} \right\rangle}{X}}$ causes..."

The subject-relation sign, [NP,S] , above the variable X indicates that the reading of the subject NP must be inserted as the value of the variable X. Now if Fillmore is right in claiming that there is a systematic relationship between subject position and semantic role, it should be possible to eliminate the use of the subject-relation sign from over many categorized variables. In the lexical entry for transitive *break*, for instance, the semantic marker (Cause) tells us that we are dealing with a sentence in which an Agent or Instrument is present. The second-ranked case, Experiencer, is not present, since the lexical entry does not include the semantic marker (Mental). Evidently, then, the subject NP must be either Agentive or Instrumental. In either case, its semantic reading goes to the left of the semantic marker (Cause).

Similarly, suppose that we have a verb whose semantic reading includes (Mental). This semantic marker alerts us to the presence of an Experiencer, and if (Cause) is not part of the meaning of the verb, we know that no Agent is present. Consequently, the highest-ranking case is Experiencer, and the reading of the subject NP must be put in place of the variable that is marked as obligatorily animate. The reader will recall that obligatory animacy is a mark of the Experiencer, where psychological verbs are concerned.

The general strategy is to examine the semantic reading for the verb of a sentence, in order to determine which cases are present in the sentence.[19] Having determined this, we pick out the highest-ranked case, in terms of Fillmore's heirarchy. This case will be associated with a specific variable in the reading for the verb, and we can proceed to replace that variable with the semantic reading for the subject NP. There is no need for explicit marking of the relationship between this variable and the subject NP, if Fillmore's hypothesis is correct.

Some verbs, however, are exceptions to the heirarchy hypothesis. These verbs normally (that is, in the active voice) take subjects which are not in the highest-ranked case present in the sentence. Frighten, for example, takes

(51) Honesty frightens a con man.

(52) A con man is frightened by honesty.

an Experiencer and an Instrument. The heirarchy leads us to expect that the normal subject of frighten will be the Experiencer, but (51) and (52) indicate that the Instrument is the normal subject of frighten. Whenever the higher-ranked Experiencer is chosen as subject, the verb must be passive. Nor is frighten an isolated example. Fillmore (1971, 42) maintains that the whole class of "so-called Psych-Movement verbs ... require a transformation which moves the highest ranked non-Experiencer noun-phrase into the first position". Presumably these verbs must be marked in the lexicon with a special subject selection feature which triggers the re-ranking transformation.

Converse pairs of relation terms provide a more random collection of exceptions to the heirarchy hypothesis. Like

(53) A likes B.

(54) B pleases A.

and please are converses of each other. This means that any sentence of the form (53) is logically equivalent to the corresponding sentence of form (54). For example, sentences

(55) Linus likes Beethoven.

(56) Beethoven pleases Linus.

(55) and (56) are logically equivalent, and indeed synonymous, because they are formally related in the prescribed way. Fillmore (1968,30) describes like and please as "synonymous; ... they differ only in their subject selection features".

Now presumably, like need not have any special subject selection feature in its lexical entry, since its subject is normally the Experiencer phrase, its object the Instrumental phrase in the sentence. Please, however, is an exception, and does require a subject selection feature, since it takes the Instrumental phrase as its subject. This pattern will be repeated for every pair of converses: buy and sell, give and receive, teach and learn, etc. At most one member of each such pair will be regular.[20] The other will have to be marked in the lexicon as an exceptional verb.

It is significant that a large number of verbs should violate the heirarchy hypothesis, even if Fillmore does not mean to claim psychological relevance for this hypothesis. But if the subject selection heirarchy is intended to reflect some kind of psychological reality, then it may actually be possible to refute the heirarchy hypothesis. Whether we take a strong or a weak position on the psychological reality of grammars (see p. 28 above), the heirarchy predicts that sentences with frighten will be more difficult to process than sentences with fear. Suppose, first, that we take a "strong psychological reality" position, according to which the hearer computes the meaning of a sentence by some process which uses the rules of the grammar. Sentences containing _frighten_ must undergo an extra transformation, a minor rule which selects the Instrument rather than the Experiencer as subject; also, the lexical entry for _frighten_ must contain a subject selection feature to trigger this special transformation. _Fear_, on the other hand, is regular with respect to the subject selection process; the heirarchy hypothesis correctly predicts which argument of _fear_ will become the subject, without the need for extra machinery in the grammar. If Fillmore takes a strong view of the psychological implications of grammar, his heirarchy hypothesis predicts that sentences with _fear_ will be easier to understand than sentences with _frighten_.

Next, suppose Fillmore takes a "weak psychological reality" position; that is, suppose he believes that hearers must recover the case structures of sentences in order to understand them, but that this recovery process does not require application of the rules of the grammar. Even on this relatively weak assumption, _fear_ ought to be easier for hearers to handle than _frighten_. For is the heirarchy hypothesis is correct, then hearers normally use a heuristic which tells them: take the highest-ranked case mentioned in the case-frame feature of the verb, and assign that case to the surface subject of the sentence. This heuristic yields correct results for _fear_, but not for _frighten_. Consequently, the hearer would have to use some other, less familiar routine to determine the case of the surface subject of frighten, perhaps after applying the normal heuristic, and finding it to fail.

Under the standard theory, on the other hand, there is no significant difference between fear and frighten. Both verbs impose certain selectional constraints on their subjects and objects, but these constraints are not claimed to be more regular or natural for one verb than for the other. Hence the standard theory predicts that sentences with frighten will be no more difficult to understand than sentences with fear, all other things being equal. This prediction follows whether we take a strong or a weak position on psychological reality. I am not aware of any psycholinguistic evidence that supports either this view or the contrary claims of case grammar; but if Fillmore means to claim psychological significance for his theory, then clearly we ought to seek psycholinguistic evidence bearing on the heirarchy hypothesis. We may expect to find such evidence by examining the relative perceptual complexity of converse pairs of verbs, such as frighten and fear, please and like, etc.

We turn next to Fillmore's treatment of surface case systems. In "The Case for Case" (Fillmore 1968, 15-16), Fillmore contends that the rules for assigning prepositions in English, postpositions in Japanese, and case suffixes in Latin, all depend on case information in deep structure.

The semantic roles, _Agentive_, _Experiencer_, and so on, are supposed to be relevant to the assignment of case features or morphemes in surface structure. This suggests that Fillmore regards the process of preposition insertion as a phenomenon particularly amenable to case-grammatical description.

In an earlier paper, Fillmore (1966, 367-368) divides English prepositions into three classes, according to the kind of information required by the rules that insert them. "Some prepositions may be filled in by optional choices from the lexicon. In Locative phrases ... generally the choice is optional: _over_, _under_, _in_, _on_, _besides_, etc. These are the prepositions that bring with them semantic information." Even here, though, case seems to play a role in limiting (if not in uniquely determining) choice of preposition. The Location case permits insertion only of those prepositions with the feature [+locative] , and excludes such prepositions as _during_.

Prepositions of the second class depend more directly on the case of the NP to which they are prefixed, but these prepositions are governed by the occurrence of specific lexical items, as well as by case information. "Thus _blame_ requires the Objective preposition to be _for_, the Dative preposition to be _on_" (Fillmore 1966, 368). This accounts

(57) Gilligan blamed the tidal wave on the tycoon.

(58) Gilligan blamed the tycoon for the tidal wave.

for the fact that on occurs in (57), while for occurs in (58). The case grammar of English contains a general rule which deletes the preposition of an NP in post-verbal position, so that no sentence containing blame is likely to contain both on and for.

Finally, there are prepositions that can be inserted solely on the basis of information about the cases found in a sentence. "Thus, the Objective preposition is of if it is the only actant in the preposition or if the preposition contains an Instrumental or Agentive; otherwise it is with. The Instrumental preposition is with just in case the Agentive co-occurs; otherwise it is by. The Agentive preposition is by" (Fillmore 1966, 368).

In order to assess the claim that case grammar handles preposition insertion more economically than the standard theory, let us consider the means at our disposal for handling these three classes of prepositions without benefit of cases. Fillmore's first class of prepositions "bring with them semantic information"; this class includes "over, under, in, on, beside, etc.", prepositions which according to Fillmore have the feature [+locative] . Now in section I, we saw that the standard theory needs a semantic marker (Location), and surely this semantic marker will appear in the lexical readings for prepositions like over and under.

Consider now the verb *sit*. Fillmore's lexicon presumably gives +[__Loc] as a frame feature of this verb, and in this way he assures that his grammar will generate sentences like

(59) Jack sat under the awning.

(60) *Jack sat of the awning.

(59) but not sentences like (60): the preposition *of* does not have the feature [+locative] , and hence this preposition cannot be part of a constituent in the Locative case. But the standard theory can also prevent *of* from being inserted in this context, without referring to the Locative case. The trick is to use the semantic marker (Location) to state a restriction on the meaning of the PrepP involved,

(61) *sit*, $\overset{[\mathrm{NP,S}]}{\underset{\langle\cdot\cdot\rangle}{\mathrm{X}}}$ rests on buttocks or haunches at $\overset{[\mathrm{PreP,VP,S}]}{\underset{\langle\mathrm{Location}\rangle}{\mathrm{X}}}$).

as in the lexical entry (61). Since the semantic reading of the phrase *of the awning* does not contain the semantic marker (Location), a grammar containing the lexical entry (61) would label sentence (60) as semantically anomalous. Perhaps Fillmore prefers to describe (60) as syntactically rather than semantically ill-formed, but surely this preference is a difficult one to argue for. With respect to the first class of prepositions, then, it appears that the standard theory and the theory of case grammar do not differ significantly in complexity.

(62) $\Delta \rightarrow \underline{on}/\underline{blame} + X + {}_{D}\Big[\ {}_{NP}\big[\ {}_{P}[\ \text{—}\ ]\big]\Big] + Y$

(63) $\Delta \rightarrow \underline{for}/\underline{blame} + X + {}_{O}\Big[\ {}_{NP}\big[\ {}_{P}[\ \text{—}\ ]\big]\Big] + Y$

Consider next rules (62)-(63), which would insert <u>on</u> and <u>for</u> in the appropriate complements of <u>blame</u>, according to Fillmore's analysis. Rule (62) states that a dummy symbol is to be replaced by <u>on</u> in any <u>D</u> (ative) complement of <u>blame</u>. According to (63), <u>for</u> replaces the dummy symbol in <u>O</u>(bjective) complements of <u>blame</u>. Taken together, these

(64) $\underline{\text{Blame}}\ \#\underline{\text{for}},\quad + \Big[\ \_\_\ \begin{matrix}\text{NP}\\ [+\text{animate}]\end{matrix} + \text{PrepP}\Big]$

(65) $\underline{\text{blame}}\ \#\underline{\text{on}},\quad + \Big[\ \_\_\ \text{NP} + \text{P} + \begin{matrix}\text{NP}\\ [+\text{animate}]\end{matrix}\ \Big]$

two rules use one less symbol than the partial lexical entries (64)-(65) which the standard theory might use to achieve the same purpose.[21]

To be sure, the standard theory needs an additional rule in order to permute <u>for</u> and <u>on</u> with the direct object, but the case grammar also needs an extra rule to get rid of the preposition on the direct object NP. Thus a count of the symbols used under the two theories suggests that the case-grammatical solution is somewhat more economical than the analysis provided under the standard theory.

But a symbol count can be used to compare two analyses only if these analyses are couched in the same vocabulary.

The point is often made in phonology that symbol counting leads to absurd consequences if, for example, one analysis uses only features while the other uses phoneme symbols as well as features. In our present circumstances, we are attempting to compare a solution using only grammatical categories and features, with a solution in which case symbols are also allowed. And since the difference in the number of symbols required was so small, it seems safe to conclude that Fillmore's treatment of this second class of prepositions is not significantly more economical than that suggested by Chomsky (1965,191).

There are certain curious features about the class of case-governed prepositions, Fillmore's third class. First, in stating the distribution of <u>of</u>, Fillmore finds it necessary to refer to propositions which contain either an Instrumental or an Agentive phrase. From section I, we know that both these cases are characterized partly in terms of the semantic marker (Cause). If Fillmore has a correct generalization here, the generalization can be stated in terms of a single semantic marker, though Fillmore has to refer to two cases. In this instance, the case terminology seems more a hindrance than a help.

I have nothing of interest to say about Fillmore's treatment of <u>with</u>. His remarks about the distribution of <u>by</u>, however, call for some comment. Fillmore regards <u>by</u> as a case-governed preposition, whose distribution can be stated solely in terms of case environments. Under the standard theory, by is introduced by the passive rule, in a formulation which naturally does not use case terminology. According to Fillmore, there are two positions in which by may occur. This preposition may mark either the Agentive or the Instrumental case, he says. But evidently this is

(66) The UFO was seen by the ground controller.

not the whole story, for sentence (66) shows by associated with an Experiencer, though the sentence contains an Instrument.[22]

In the light of sentences like (66), we can amend Fillmore's account of by, as follows. If a sentence contains an Agentive phrase, then by marks the Agentive. If a sentence does not contain an Agent, then by marks the Experiencer; and if no Experiencer is present, then by marks the Instrument. This ranking of cases - Agentive, Experiencer, Instrumental - is exactly the heirarchy which is supposed to govern the normal choice of subjects. It would be odd if the rules of English had to duplicate the statement of this heirarchy in universal grammar, in order to insert the preposition by where it belongs.[23]

Evidently, by is not specifically associated with any one case. Rather, it marks whatever NP would normally be chosen as the subject of a given sentence. In the standard theory, the NP which Fillmore would call the normal subject choice is invariably the subject in deep structure, and the passive rule attaches by to the NP which occupies the subject position in deep structure. The occurrence of by is not predictable directly in terms of case, but rather in terms of syntactic position. Fillmore (1971, 42-43) has come to recognize this, and he now has a passive transformation which handles the insertion of by. His analysis has thus come to look very much like that provided by the standard theory, and he is hardly in a position to argue that his treatment of by is more elegant than the usual one.

So far, we have been considering only one kind of surface case system, the prepositions of English. We have not thus far seen any reason to believe that case grammars provide a more efficient apparatus for dealing with surface cases than does the standard theory. Before we take up Fillmore's arguments about lexicon simplicity, let us cast a glance at one other kind of surface case system.

Fillmore (1968, 17) observes that case affixation often depends on the status of an NP as subject or direct object in surface structure. The case of pronouns in English is determined in this way, for example. Now surface subjects and objects are selected partly on the basis of case information in deep structure, according to Fillmore. Thus semantic roles are indirectly involved in determining the cases of English pronouns. But in order to link semantic roles with surface cases of this kind, Fillmore needs subject and object selection rules, and the standard theory does not need such rules at all. Quite apart from this, however, subject selection rules may be objectionable on formal grounds. Emonds (1970) has argued that most generally accepted transformation rules appear to preserve structure, and that perhaps all transformations must preserve structure, under a universal constraint. Chomsky has pointed out to me in conversation, however, that Fillmore's subject selection process is not a structure-preserving transformation. If Emond's constraint is correct, then subject selection rules violate a principle of universal grammar.

We may conclude, I think, that the study of surface case systems does not support the theory of case grammar. We turn now to Fillmore's claim that case grammars have simpler lexicons than do grammars provided by the standard theory. Fillmore (1968, 29-31) defends this claim rather clearly in "The Case for Case". There he points out that in a case grammar, <u>like</u> and <u>please</u> can have identical lexical readings. Aside from their phonological differences, these verbs differ only syntactically, in their subject selection properties. There is no need for any semantic distinction between these two verbs. Consequently, "it is possible to reduce the number of semantic entry <u>types</u>" (in contrast with semantic entry tokens) in the lexicon of English (Fillmore 1968, 30).

According to this argument, _like_ and _please_ are synonyms, and it is implied that this has some bearing on the complexity of the lexicon. Fillmore seems to be invoking some general principle, to the effect that given two lexicons with the same number of entries, the one with the greater number of synonyms is the simpler of the two. I see no initial plausibility in this principle, and Fillmore provides no argument to support it. It is not possible to have just one lexical entry for _like_ and _please_, simply because these words are phonologically distinct. Whether or not cases are incorporated into deep structures, then, the lexicon will have to include two lexical entries, and Fillmore has given us no reason to believe that these entries will be any simpler under the theory of case grammar than under the standard theory.

Similar considerations can be brought against Fillmore's arguments concerning _die_ and _kill_, _see_ and _show_, etc. He argues that a case grammar can treat these pairs of verbs as synonyms, differing only in their case structure. But since these pairs of verbs are phonologically distinct, each member of a pair must have its own lexical entry, in a case grammar as in the standard theory.

Fillmore has other proposals, however, which are not vulnerable to this objection. He suggests, in particular, that a case grammar needs only one lexical entry for the transitive and intransitive senses of _open_. The standard theory appears to need two entries; one must contain the semantic marker (Cause), while the other cannot. Besides, these two verbs have different contextual features. The causative open takes an object, but the other does not. Fillmore seems to have brought to light a genuine difference in complexity between his lexicon and its counterpart in the standard theory.

Katz (1972, 357-361) has pointed out, however, that it is possible to devise a notation for collapsing the two lexical entries for open. This notational innovation uses only information already available in a standard grammar. In particular, there is no need to use case terminology. An

$$
\begin{aligned}
&(67)\quad \underline{\text{open}},\ +\left[\underline{\quad}(\text{NP})\right],\quad \Big(\Big(\underset{\langle\cdots\rangle}{\overset{[\text{NP},\overline{\text{S}}\ \underline{\text{NP}},\text{VP},\text{S}]}{X}}\quad \text{causes} \\
&\qquad\qquad \Big(\Big(\text{Condition}\Big)\Big(\underset{\langle\cdots\rangle}{\overset{[\text{NP},\text{VP},\overline{\text{S}}\ \underline{\text{NP}},\text{S}]}{Y}}\quad \text{closed at } \underline{t}\Big)\ \ldots, \\
&\qquad\qquad \Big(\text{Condition}\Big)\ \underset{\langle\cdots\rangle}{\overset{[\text{NP},\text{VP},\overline{\text{S}}\,\underline{\text{NP}},\text{S}]}{Y}}\quad \text{open at } \underline{t} + \underline{n}\Big)\Big)\Big)
\end{aligned}
$$

abbreviated form of this collapsed lexical entry is given here as (67).

This notation is to be understood in the following way. The variable X has the reading of the subject NP as its value, just in case the sentence in question contains an NP bearing the relation [NP,VP,S] to the verb. The variable Y has the reading of the direct object NP as its value, provided the sentence contains an NP which bears the subject relation, [NP,S], to the verb. In the event that the sentence does not contain a direct object, the semantic marker containing the variable X (namely, X causes) is deleted, and the value of the variable Y is the reading of the subject NP. Our new notation thus provides that when open is used intransitively, it will be interpreted as a simple process verb, with no causal element in its meaning.

This notation is rather elaborate, and one is naturally interested to know what can be said for its inclusion in linguistic theory. This is clearly related to the question of what would be said by the inclusion of such a notation in the definition of possible grammar. The notation permits us to collapse two lexical entries only when the verbs involved are phonologically identical, and only when their senses are related in a certain way. Chomsky (1965, 42-46) points out that abbreviatory conventions of this sort are vehicles for claims about the way children learn languages. In this case, the claim seems to be that it is easier for a child to learn the two uses of open than it is for him to learn to use see and show, or black and blacken. Otherwise put, the claim shared by this new notation and the theory of case grammar is that it is more natural for a language to contain verbs like open than it is for a language to contain pairs of verbs like kill and die.

Is this claim about naturalness correct, or is it wasted effort to try to duplicate the simplification effected in the lexicon of a case grammar? Fillmore

(1968, 30n) believes that "it is a language-particular coincidence that English uses the same form" for both transitive and intransitive senses of <u>open</u>. I cannot think of any reason to dispute this view. By Fillmore's own lights, then, we must conclude that this bit of lexicon simplification embodies a false claim about what is natural in lexical structure.

In summary, none of Fillmore's syntactic arguments appear to favor case grammar over the standard theory. The attempt to define non-stativity in case terms failed, but success would not have supported the theory of case grammar, since a parallel definition is possible within the standard theory. Furthermore, the feature $[\pm\text{stative}]$ is somewhat problematical in itself. Fillmore's heirarchy hypothesis about subject selection can also be incorporated into the standard theory, in the form of universal constraints on the process of semantic interpretation. One is not inclined to carry out the importation, however, for the heirarchy hypothesis entails some unsupported and implausible phychological claims. In describing surface case systems, the theory of case grammar enjoys no clear advantage over the standard theory. Many of the lexical economies claimed for case grammar are illusory, and the rest carry with them a rather dubious claim about naturalness, which Fillmore himself rejects in another context.

III. A METHODOLOGICAL ARGUMENT

Fillmore (1966, 374) has argued that the theory of case grammar enjoys certain "general advantages", in addition to its alleged semantic and syntactic virtues. These more general advantages "relate to the interpretation of historical changes and cross-language differences in lexical structure". In the history of English, for example, _like_ was once "an exception to the rule that fronted actants are neutralized to the so-called nominative form" (Fillmore 1966, 374). The verb _like_ could appear in sentences like

(68) Him like oysters.

(69) He likes oysters.

(68), at that period of its history, but a change took place, and (69) superseded (68). This change, Fillmore claims, is "purely syntactic ; ... _like_ did not change in its meaning" (Fillmore 1966, 374). Case grammar has the advantage, according to Fillmore, of recognizing the purely syntactic nature of this change.

A synchronic comparison of English _kill_ and Japanese _korosu_ reveals that these verbs are exact synonyms, with only syntactic differences. The Japanese verb requires an animate subject, while _kill_ may have an inanimate Instrument phrase as its subject. Again, Fillmore (1966, 374) suggests that it is a virtue of case grammar to recognize the superficial character of the differences between these verbs, and their underlying semantic identity.

More recently, Fillmore (1970, 258-259) has used data taken entirely from present-day English to support this kind of argument. The adoption of a case theory, he claims, "makes it possible to provide a sharp separation between ... purely syntactic phenomena ... and facts about semantic interpretation". For example, the differences between *buy* and *sell*, *kill* and *die*, and the two uses of *open*, are synonymous.

Advocates of generative semantics have argued that it is misleading to draw any such sharp distinction between syntax and semantics as Fillmore proposes. These linguists are not likely, then, to view Fillmore's observations as an argument in favor of case grammar. Defenders of the standard theory, however, will not be at all inclined to disagree with Fillmore's claim that syntax and semantics are distinguishable enterprises. Katz (1970) has in fact recently presented empirical arguments on behalf of this position.

It may be, of course, that while Fillmore agrees that a line between syntax and semantics must be drawn somewhere, he is finding fault with the place at which the standard theory draws it. A closer look reveals, however, that there is considerable agreement between Fillmore's lexicon and Katz's. On Katz's analysis of converse pairs of verbs, like lend and borrow, these verbs get identical lexical entries except for the markers above categorized variables. These markers refer to grammatical relations, and indicate the positions at which to insert the readings for subject, direct object, etc. into the lexical reading for the verb. The vocabulary of these markers is exclusively syntactic. It seems fair, then, to say that a Katzian lexicon describes lend and borrow as synonyms, with only syntactic differences. Similarly, Fillmore's analysis of the historical change in like, and of the difference between Japanese korosu and English kill, is not in disagreement with the treatment these matters would receive under the standard theory.

There is disagreement, however, about kill and die, and about the two uses of open. The lexical entries for kill and for transitive open both contain the semantic marker (Cause) in a standard grammar of English, whereas the entries for die and for intransitive open do not. To this extent, these words are represented as semantically distinct. Fillmore maintains that these verbs are really synonymous, and he claims this fact as evidence in favor of case grammar, which recognizes their synonymy.

But is it really a fact that kill and die are synonyms? How does one find out about such things? The intuitions of any speaker of English can tell him that there is a close relationship between these two words, but on that the standard theory agrees with Fillmore. We want to know more specifically whether the close semantic relationship between these two verbs is one of synonymy, and on this question, native speakers cannot be expected to have meaningful intuitions, simply in their capacity as native speakers. The question is a technical one, having to do with the form of lexical entries, and to understand the question, one must have a certain amount of training in linguistic theory.

? compare p 128!

Fillmore is claiming to have carved Nature at the joints, specifically at the joints between syntax and semantics. The joints of Nature are not directly visible, however, to the inner eye of linguistic intuition. To decide which of two theories has a better grasp on Nature's joints, we first confront these theories with the facts. Are there facts which refute one theory, while supporting the other? Does one theory explain a wider range of facts than the other? If both theories remain unrefuted, and explain the same phenomena, we turn to considerations of theoretical simplicity to decide the issue. To say that _kill_ and _die_ are synonymous is to say that they get identical lexical readings in the best supported, or simplest, theory of English.

The methodological argument for case grammar is not a marshalling of facts to settle the disagreement between Fillmore's lexicon and Katz's. Nor is it an appeal to theoretical simplicity. Rather, it is a promissory note for an argument, a parochial description of the disagreement between case grammar and the standard theory over the proper lexical treatment of _kill_ and _die_. This argument attempts to make virtue of necessity, by taking a claim which follows from the theory of case grammar, and suggesting that this claim is somehow obvious to all English speakers. On examination, then, the methodological argument for case grammar is not an independent argument at all. The case for case must rest on the semantic and syntactic arguments discussed in sections I and II.

Fillmore's semantic argument did not actually establish the need for semantic roles in deep structure, because the semantic representations (though not the deep structures) provided by the standard theory contain complete information on the semantic roles of constituents. On the debit side, a case grammar has difficulty capturing the semantic distinction between holistic and partitive Locatives, and the theory of case grammar needs additional apparatus to exhibit the entailments of sentences containing Agents, Instruments, and Goals. Syntactically, the effects of adopting the theory of case grammar do not include any significant simplification in the lexicon, nor in the rules for surface case marking. The hypothesis about a universal subject selection heirarchy has some implausible psycholinguistic consequences. Fillmore's attempt to define the feature $\begin{bmatrix}\pm\text{stative}\end{bmatrix}$ in terms of the Agentive case fails. It appears that the case for case can be dismissed.

Footnotes

* I am grateful to Dr. Donald Frantz, of the Summer Institute of Linguistics, and to Greg Thomson, one of our students at the Institute, for a sympathetic introduction to the theory of case grammar. Prof. Jerrold Katz encouraged me to examine Fillmore's work in a more critical light, and his guidance in the revision of earlier drafts of this paper has been most valuable. Prof. Noam Chomsky and Prof. Steve Anderson read the penultimate draft, and several changes have resulted from their comments. This work was supported in part by NIH grant 5 T01 HD00111.

1. I have changed the numbering on Fillmore's examples, here and throughout this paper.

2. The bracketed remark is Fillmore's.

3. Steve Anderson has pointed out to me that Fillmore's one-occurrence-per-proposition constraint on case nodes amounts to a reformulation of Chomsky's uniqueness requirements.

Recently Fillmore (1971, 54-55) has expressed dissatisfaction with the practice of representing semantic roles by means of case nodes in tree diagrams, but so far as I can see, the arguments to be presented here do not depend on the continuation of this practice. Indeed, it is not clear to me, from Fillmore's brief sketch of his reform proposal, whether the formulation he now prefers is different in any substantive respect from his 1968 formulation.

4. Fillmore's argument may also be construed in the following way: "Grammatical relations in deep structure were intended solely to carry information about semantic roles. The relations defined by Chomsky do not, however, serve this purpose at all adequately. Therefore, no motivation remains for having deep structures to which Chomsky's definitions can be applied." For a rebuttal to this argument, cf. Katz 1972, 111-113.

5. My argument (pp. 13-22) derives in part from some insights by Chomsky (1972, 98-99).

6. Nouns in Algonkian languages are syntactically either animate or inanimate in gender, but as in Indo-European languages, syntactic and semantic gender do not always agree. Many nouns are syntactically animate but semantically inanimate, e.g. Cheyenne mhaa 'melon'.

7. This lexical entry is a simplified version of the entry proposed by Katz (1972, 358). For a discussion of process verbs and temporal variables, cf. Katz 1972, Chapter 7.

8. This generalization is meant to combine the virtues of two statements by Anderson (1971, 390).

9. Anderson (1971, 388) numbered this example (5) d.

10. Some but not all of Anderson's examples are discussed, and similar conclusions are reached, in Chomsky (1972, 101-102).

11. The subject selection rule will already have applied at this point (Fillmore 1968, 35-36).

Fillmore's recent work requires a slight change in the formulation of this object-selecting transformation. Since the cases now appear in deep structures in a uniform linear order, dictated by the subject choice heirarchy (Fillmore 1971, 37), the object selection rule will be an optional transformation that moves the Goal phrase *on the wall* to the object position from its normal deep-structure position to the right of *paint*.

Anderson's argument against the object selection rule assumes that sentences like (22), (30), and (31) arise out of simple, one-clause deep structures in which Instrument and Goal are co-constituents. Fillmore (1971, 45-46) has recently posited complex, multi-clause deep structures for sentences like (22), (30), and (31), and this deprives Anderson of the particular verbs he used in his argument. However, there are other verbs which behave in the requisite ways vis-à-vis the object selection rule. In place of *smear*, we can use the verb *hit*, which takes both Instrument and Goal arguments; either of these arguments may appear as the direct object in surface structure (cf. *John hit the fence with his cane*, to which the object selection rule has applied, versus *John hit his cane against the fence*, to which the rule has not applied). The verb *open* resembles Anderson's *cover*, in that the object selection rule *must* apply to sentences containing *open* (cf. **John opened his credit card of the door*, in which the order of post-verb constituents follows the deep-structure order imposed by the subject selection heirarchy, versus *John opened the door with his credit card*, in which the constituents have been rearranged by the object selection rule).

Finally, _throw_ in Anderson's argument may be replaced by _frighten_, to which object selection _cannot_ apply (cf. *_Lucy frightened a snake to Schroeder_, to which the rule has applied, versus _Lucy frightened Schroeder with a snake_, in which the constituents occur in their "natural" order).

12. Fillmore (1968, 48-49n) cites _spray_, _blame_, _open_, and _break_ as "examples of homonymy".

13. A conversation with Ned Block helped to clarify my understanding of the psychological issues between Fillmore and Anderson. For some further remarks on the psychological implications of case grammar, pertaining to subject selection rather than object selection, see pp. 50-52 of this paper.

14. It would be a mistake to give (32) a-b identical semantic representations, however. Cf. Katz 1970, 253n for discussion of a similar example. Neither Katz, Fillmore, nor I have any suggestions about how to represent the semantic difference between (32) a and (32) b.

15. For simplicity, I have eliminated the Mod and Prop nodes from these structures.

16. To be sure, Fillmore does not absolutely have to predict entailments in terms of case structures. He could, for example, predict entailments in terms of semantic representations, i.e. semantically interpreted case structures. But a great deal of semantic information is already available in case structures, in the form of case labels on NP's; the process of semantic

interpretation can only add semantic information that depends on the meanings of particular lexical items in the sentence. Back in the days of the Factitive case, Fillmore quite clearly regarded the entailment from (34) a to (34) b as an entailment which held, not by virtue of the lexical meaning of _write_, but by virtue of the case properties of this verb. I am assuming that he still holds this view, and that Figures 5 and 6 provide all the semantic information we need to predict the entailment under discussion.

17. If Fillmore's continuing analysis of agency uncovers other aspects of the meaning of the Agentive case, then presumably it will be possible to devise some new, _ad hoc_ semantic marker to distinguish Agents within the standard theory. To call such a marker _ad hoc_ is simply to give credit to case grammarians for the insights they may develop into the nature of agency. To call the _ad hoc_ device a semantic marker, rather than a case label, is to point out that such insights can be stated within the format of a Katzian lexicon. These insights would not provide evidence of the formal inadequacy of the standard theory.

18. Susumu Kuno makes a somewhat similar point in "Some Properties of Non-referential Noun Phrases", _Studies in Oriental and General Linguistics_, R. Jakobson and S. Kawamoto, eds., Tokyo, The TEC Company, 1970.

19. Until we discover a clear definition for the Agentive case, it will be difficult to tell Agents from Instruments. Otherwise, however, the cases should be recognizable in lexical entries for verbs, on the basis of the criteria sketched in section I.

20. Fillmore (1970, 257) describes rob and steal as converses. These verbs differ in their object selection properties, though both take Agentive subjects, and are thus regular with respect to the subject selection process. Whenever a pair of converses differ in their subject selection processes, however, one member of the pair must obviously be irregular, and both may be. In this connection, see also the discussion of the object-selection properties of smear, cover, and throw, pp. 23-31 of this paper.

21. Given Fillmore's present views, he might want to regard blame as a complex predicate, with an Experiencer subject plus an embedded S containing an Agent (or Instrument) and a Goal. Lexical entries (57)-(58), and rules (55)-(56), could be changed to reflect this analysis, but so far as I can see, these changes would not alter the conclusion of this discussion, but only its complexity.

22. I am indebted to Janet Fodor for calling my attention to this fact and its implications.

23. Ray Dougherty makes the same point in "Recent Studies on Language Universals," Foundations of Language 6, 520-525. There are of course uses of by which are not accounted for either by the standard passive rule or by Fillmore's amended rule. Consider, for example, He saw the UFO by the light of the moon, or The UFO landed by the outhouse.

References

Anderson, S. (1971) "On the Role of Deep Structure in Semantic Interpretation," _Foundations of Language_ 7, 387-396.

Bach, E. and R. T. Harms, eds. (1968) _Universals in Linguistic Theory_, Holt, Rinehart and Winston, New York.

Chomsky, N. (1965) _Aspects of the Theory of Syntax_, MIT Press, Cambridge, Mass.

Chomsky, N. (1972) "Some Empirical Issues in the Theory of Transformational Grammar," in Peters (1972).

Emonds, J. E. (1970) _Root and Structure-preserving Transformations_, Unpublished Ph.D. dissertation, MIT.

Fillmore, C. J. (1966) "Toward a Modern Theory of Case," in Reibel and Schane (1969).

Fillmore, C. J. (1968) "The Case for Case," in Bach and Harms (1968).

Fillmore, C. J. (1970) "Subjects, Speakers, and Roles," _Synthese_ 21, 251-274.

Fillmore, C. J. (1971) "Some Problems for Case Grammar," in O'Brien (1971).

Fodor, Jerry A. (1971) "Current Approaches to Syntax Recognition," in Horton and Jenkins (1971).

Horton, D. and J. Jenkins, eds. (1971) _The Perception of Language_, E. Merrill, Columbus, Ohio.

Katz, J. J. (1967) "Recent Issues in Semantic Theory," _Foundations of Language_ 3, 124-194.

Katz, J. J. (1970) "Interpretative Semantics vs. Generative Semantics," Foundations of Language 6, 220-259.

Katz, J. J. (1972) Semantic Theory, Harper and Row, New York.

Lakoff, G. (1966) "Stative Adjectives and Verbs in English," in Mathematical Linguistics and Automatic Translation, Report No. NSF-16, Harvard University Computation Laboratory, Cambridge, Mass.

O'Brien, R. J., ed. (1971) Report of the Twenty-Second Annual Round Table Meeting on Linguistics and Language Studies, Georgetown University Monograph Series on Languages and Linguistics, No. 24, Georgetown University Press, Washington, D. C.

Peters, S., ed. (1972) Goals of Linguistic Theory, Prentice-Hall, Englewood Cliffs.

Chapter 2

Linguistics and Epistemology*

Leibniz, in his New Essays on Human Understanding, maintains that there are certain "truths" or "principles" which "everybody knows" innately, rather than on the basis of "the external doctrine" of sensory experience (Leibniz, 1704, 77). Noam Chomsky has argued that the findings of transformational linguistics "are fully in accord with the doctrine of innate ideas", as that doctrine was defended by Leibniz and other classical rationalists. Indeed, Chomsky suggests that the theory of generative grammar "can be regarded as providing a kind of substantiation and further development of this doctrine" (Chomsky 1967, 10). It is apparently Chomsky's view that a child learning a language has innate "propositional knowledge" of " the linguistic universals" (Stich 1971, 480); that is, the language learner knows (innately) the propositions which make up general linguistic theory, or the theory of universal grammar.[1]

Not surprisingly, Chomsky's claims about innate linguistic knowledge have generated widespread controversy. Another, quite distinct (and equally controversial) claim made by rationalistic linguists is that "speakers [tacitly] know the grammatical rules of their language", as those rules are represented in a transformational grammar of their language (Graves, Katz, et al. in press). That is, a native speaker of English has tacit propositional knowledge of principles belonging to particular grammar, as well as those belonging to universal grammar. These controversial knowledge claims are based on a much less controversial one: that every English speaker knows (or can easily determine, in an indefinitely large number of cases) that "particular expressions have certain grammatical properties", eg. that ""The cat is on the mat" is grammatical" (Graves, Katz, et al. in press). This relatively uncontroversial knowledge claim might be described as the claim that "speakers have propositional knowledge of the consequences of the rules ... of their grammar" (Stich 1971, 481).

This paper examines arguments by three critics of linguistic rationalism. For convenience I have structured the discussion in terms of the traditional analysis of knowledge as justified true belief. First we shall consider a paper by R. Edgley, in which he argues that our innate linguistic beliefs (if any) cannot be justified beliefs, and hence cannot be instances (at least, not paradigmatic instances) of knowledge. Edgley also objects, like many others, to the concepts of implicit belief and inference. Next we shall examine Quine's thesis of the indeterminacy of translation. If that thesis is correct, then certain of our linguistic beliefs are not true (nor false), and for this reason fail to qualify as knowledge. Finally, we shall look at Stephen Stich's contention that we have no beliefs, let alone any knowledge, about universal or particular grammar.

One might suppose that such objections pose no threat to the rationalist's knowledge claims, because of Edmund Gettier's (1963) well-known attack on the traditional analysis of knowledge. This supposition would be mistaken, however. To see that Gettier has not provided rationalists with a defense against Edgley, Quine, and Stich, let us look briefly at one of Gettier's counterexamples.

> Suppose that Smith and Jones have applied for a certain job. And suppose that Smith has strong evidence for the following conjunctive proposition:
>
> (d) Jones is the man who will get the job, and Jones has ten coins in his pocket.
>
> ... Proposition (d) entails:
>
> (e) The man who will get the job has ten coins in his pocket.
>
> Let us suppose that Smith sees the entailment from (d) to (e), and accepts (e) on the grounds of (d), for which he has strong evidence. In this case, Smith is clearly justified in believing that (e) is true.
>
> But imagine, further, that unknown to Smith, he himself, not Jones, will get the job. And, also, unknown to Smith, he himself has ten coins in his pocket. Proposition (e) is then true, though proposition (d), from which Smith inferred (e), is false. In our example, then, all of the following are true: (i) (e) is true, (ii) Smith Believes that (e) is true, and (iii) Smith is justified in believing that (e) is true. But it is equally clear that Smith does not know that (e) is true ... (Gettier 1963, 122; emphasis Gettier's)

This example shows, as Gettier points out, that justification, truth, and belief do not together constitute "a sufficient condition for someone's knowing a given proposition" (Gettier 1963, 123; emphasis Gettier's). But the objections of Edgley, Quine, and Stich do not depend on the assumption that justification, truth, and belief are jointly sufficient to guarantee knowledge. These objectors require only the assumption that these conditions are severally necessary for knowledge.[2] The view Gettier has refuted, then, is not a view that Edgley, Quine, and Stich have to take in order to support their objections against tacit and innate linguistic knowledge. Thus Gettier has not provided rationalists with a ready-made defense against these critics.

I. EDGLEY: JUSTIFICATION AND BELIEF

Let us turn, then to Edgley's paper. First, we shall examine his critique of the rationalist argument for innate knowledge of linguistic universals. On the basis of this examination, I shall sketch what I take to be Edgley's views (never explicitly stated) on how one goes about justifying a knowledge claim. In order to provide further substantiation for that sketch, we shall then look at the way Edgley tries to show that we know (the grammar of) English. Finally, I shall reply to Edgley's critique of the rationalist argument, and to the views which appear to underlie that critique.

Edgley maintains that there can be only one argument to support "the strong claim that what is innate [in the language learning process] is not merely mental structure, and not merely Humean natural belief, but knowledge" (Edgley 1970, 30). That argument, which Edgley attributes to Chomsky, rests on the following principle, which he attributes to Hume. (I shall accordingly refer to it as Hume's principle.)

> ... if someone knows that something is the case on the basis of, or in consequence of, certain evidence or data, it follows that he must already know, explicitly or implicitly, whatever else (if anything) is necessary for what he consequentially knows to be logically derivable from what he already knows in conjunction with the data. (Edgley 1970, 30)

By Chomsky's account, a child acquires knowledge of the grammar of his language on the basis of certain primary linguistic data (Chomsky 1965, 30-37). But "since [this data] is inadequate, he must [according to Hume's principle] know some suppressed major premise from which, in conjunction with the evidence, his consequential knowledge logically follows" (Edgley 1970, 30). We must describe the child's innate language-learning endowment as knowledge, because "anything less would not account for the fact that the output of the language-learning situation is itself knowledge" (Edgley 1970, 30). The premisses of the argument, then,

(1) A native speaker of any language knows the grammar of that language on the basis of primary linguistic data drawn from that language.

(2) The grammar of a language is not logically derivable from the primary linguistic data available to language learners.

(3) The grammar of a language is logically derivable from the principles of universal grammar, in conjunction with the sort of primary linguistic data available to language learners.

apart from Hume's principle, are (1), (2), and (3) above.

As I remarked in the preceding paragraph, Chomsky's account of language learning seems clearly to commit him to (1). He also argues (Chomsky 1965, 47-59) for claims very much like (2) and (3). To this extent, Edgley has portrayed Chomsky's position faithfully. But notice that Hume's principle requires something slightly stronger than (3). Hume's principle only allows us to infer that a language learner knows the theory of universal grammar if knowledge of that theory is "necessary for what he consequentially knows to be logically derivable from [that theory plus] the data". That is, we are only allowed to supplement the data with the weakest theory sufficient to permit inference from the data to the acquired grammar.

It might appear that Chomsky holds this stronger version of (3). In one passage, for example, he writes" "A consideration of the character of the grammar that is acquired, [and of] the degenerate quality and narrowly limited extent of the available data ... leave[s] little hope that ... the language can be learned by an organism initially uninformed as to its general character" (Chomsky 1965, 58). Universal grammar is of course intended to specify the child's initial information about the general character of human language. But a child might be able to learn English with less (or less general) information than the entire theory of universal grammar. Chomsky postulates knowledge of universal grammar in order to explain not merely the child's learning of English, but his ability to learn any human language (cf. Chomsky 1969a, 61-64). Otherwise put, Chomsky wishes to explain language-learning competence, not merely language-learning performance. Hume's principle seems concerned only with the learning performance of a single individual, and in this respect the argument sketched by Edgley is hardly a Chomskian one.

But the argument based on Hume's principle departs from the rationalist position in another respect, and the departure

I have in mind is, I believe, important to an understanding of Edgley's views on the verification of knowledge claims. The most succinct presentation of the rationalist argument that I know of was written about three years after Edgley's paper, but that presentation differs from those available to Edgley, it seems to me, only in its succinctness. For convenience, then, I propose to refer to the recent, brief version of the argument. Here is the heart of it:

> If the best account of how speakers learn grammatical rules says that they learn them on the basis of certain innate linguistic principles $P_1, \ldots, P_n$ that determine a grammar G with respect to some class D of primary linguistic data, then there is a deduction of G from D and $P_1, \ldots, P_n$. (Graves, Katz, *et al*. in press)

The argument assumes that a native speaker knows the rules of G, the grammar of his language. This much follows from premise (1) of Edgley's argument. Also, the argument includes among its premises something amounting to (3): there is a deduction of G from D and $P_1, \ldots, P_n$. But there is another premise in the Graves-Katz argument[3] which has no counterpart in Edgley's version. That is the claim that the best explanation of the way the speaker came to know G is that he deduced it from other things he knew, viz. D and $P_1, \ldots, P_n$.

It seems clear that Edgley intends his argument to be a deductive one; he means for the conclusion (4) to follow

(4) The native speaker of any language knows the principles of universal grammar.

deductively from Hume's principle, together with (1), (2), and (3). The argument given by Graves, Katz, *et al.*, however, does not appear to be a deductive one, and it does not appear to require Hume's principle. The rationalist argument is rather a sort of inductive argument, specifically an argument to the best explanation. They argue that if the best explanation for an established fact (namely, our knowledge of particular grammar) says that we know the theory of universal grammar, then we probably do know that theory. What renders the ascription of innate knowledge plausible, according to the rationalists, is the explanatory value of that ascription. But of course false theories may enjoy a degree of explanatory power, so even inclusion in the best available explanatory theory provides no deductive guarantee of truth for the ascription of innate knowledge.

The Graves-Katz argument, I have suggested, can get on nicely without Hume's principle. But in order to establish the explanatory value of the rationalist account of language learning, one might wish to appeal to something like the converse of Hume's

(5) If someone knows *p*, and validly deduces *q* from *p*, then that person also knows *q*.

principle, namely (5).[4] On a deductivist theory of explanation, at any rate, one could make a *prima facie* case for the rationalist explanation by pointing out that the *explanandum*, (6),

(6) Subject A knows grammar G.

(7) Subject A knows data D and principles $P_1, \ldots, P_n$.

(8) Subject A deduced G from D and $P_1, \ldots, P_n$.

is entailed by the conjunction of (5) with the rationalist claims (7) and (8). I shall have more to say about Hume's principle, and the deductive character of Edgley's argument, later on.

Now, however, it is time to look at Edgley's criticisms of the rationalist argument. He has two criticisms. First, Hume's condition, he argues, imposes a condition on knowledge that is "too stringent even in normal cases"; and second, the process of learning a language is "far from being a normal case of the acquisition of knowledge" (Edgley 1970, 30-31).

On the first point, he writes:

> ... it is odd, on the face of it, that Chomsky should insist on such stringent conditions for consequential knowledge and allow such lax conditions for innate knowledge: to know a language a child needs not only the data but also knowledge of the general form of any human grammar; but he can apparently know the general form of any human grammar without so much as a scrap of evidence. ... If we try to account for knowledge by tightening up the standards to an almost impossible pitch in one place, we shall have to relax them beyond reason in another; as in other contexts, the price of a Puritanical exterior is an interior where anything goes. (Edgley 1970, 31)

In this passage, Edgley points out that the rationalist argument does nothing to show that our innate beliefs about human language are _justified_ beliefs. But his point is not merely that _this_ argument fails to show justification for our innate beliefs. If that were his only objection, we might hope to find some other argument that would satisfy the demand for justification. In another passage, however, Edgley has argued that the search for such an auxiliary argument is bound to fail. He points out that in classical rationalism, the principles alleged to be known innately were "so basic or rudimentary that it was from a logical point of view unnecessary for [them] to be justified". These principles were either "involved in the very notion of justification", or else they "could not be understood without [their] truth being accepted". But "the very contingency" of Chomsky's universal grammar makes empirical justification necessary. On the other hand, its very innateness makes empirical justification impossible, for surely no prenatal experience the child may have could provide him with any information relevant to the truth or falsity of universal grammar. (Edgley 1970, 29)

Edgley's first objection, then, is that the rationalists have not shown, and cannot show, that our innate linguistic beliefs are justified; to call such unjustified beliefs knowledge is to relax the standards "beyond reason". Edgley explains his second criticism in the following way:

> ... until the process of language acquisition has gone far enough to give the child a fair mastery of the language, he occupies, with the birds, bees and fishes, that area where the application of the concepts of belief and knowledge are essentially disputable. Even more disputable are the concepts of reasoning and inferring involved in my paradigm example.[5] Positively inapplicable are the concepts in terms of which Chomsky models the child's acquisition of language-- the concepts of forming hypotheses about linguistic structure and testing them against the evidence. Chomsky is, of course, fully aware that this picture of the child as a scientific genius is

> only a formal reconstruction of the language-learning situation. But it is only in the context of this picture that ... the input [can be plausibly] represented as data or evidence, ... and the consequential knowledge of the language represented as inferred from, or logically based on, the data, which then become the child's reasons justifying his claim to know the language. (Edgley 1970, 31)

In the passage just quoted, we may (although Edgley did not) distinguish two separate objections to the rationalist argument. The first of these objections consists in the remark that the application of the concept of belief to non-verbal or pre-verbal organisms is "essentially disputable". Three pages back, Edgley has suggested that "when someone is able to use and understand words and so make statements, the concepts of belief and knowledge become clearly applicable, because what a person believes or knows can be identified as what he says". Dogs, birds, and fish, however, "whose behaviour ... does not include their use of ... words", fall within "an essentially disputable area of application" of the concepts of belief and knowledge. (Edgley 1970, 28)

According to this objection, it is necessary condition on someone's knowing the principles P_1, ..., P_n that he believe those principles. Now "correct application [of psychological concepts such as belief] is determined by a range of criteria", and the standard criterion for belief directs us to a person's verbal behavior: a person believes proposition p if he utters some sentence expressing p. (Edgley 1970, 28) Knowledge of linguistic theory is supposed to be a precondition for language learning, but we can hardly hope to catch children asserting principles of linguistic theory, either at the beginning of the language acquisition process or (in most cases) at any other time in their lives. Indeed, as Stich points out, most people "are incapable of recognizing [principles of universal grammar] when presented"; many are even "incapable of understanding them" (Stich 1971, 486). On our principal criterion of belief, Edgley argues, linguistic theory fails dismally.

It is well worth noting that this same objection applies against the claim that native speakers know the rules of the grammar of their language. Again, most native speakers do not assert any of those rules at any time in their lives; even when presented with rules of their grammar, they do not recognize them as such. If the claim that we believe principles of universal grammar is "essentially disputable", then the claim that we know principles of particular grammar is disputable on the very same grounds.

But Edgley has a further objection to the rationalist argument. That objection is that premise (1) is not literally true, so our knowledge of English grammar does not satisfy the antecedent of Hume's principle. The grammar of English is not something we know to be the case "on the basis of ... certain evidence or data". A child learning English benefits from various inputs (parental corrections, for example), but these inputs are not literally evidence on the basis of which the child makes inferences about English grammar. Only in the context of a formal reconstruction of language learning does it make sense to describe these inputs as evidence, or data. This may be enough to give (1) a kind of metaphorical truth, but Hume's principle requires the literal truth of (1).

I wish to defer my comments on these objections until later. For the present, we shall only use these objections as a basis for reconstructing Edgley's conception of the way knowledge claims are to be established. First, I think these objections arise out of a belief that arguments in support of knowledge claims must be deductive arguments; inductive arguments, or anyway arguments to the best explanation, are clearly not sufficient, in Edgley's view, to ground knowledge claims. This claim about Edgley's views is further borne out, I think, by the way he misread the rationalist argument: he took it to be a deductive argument, something it was surely never intended to be.

Second, it would appear that the major premise in such an argument must state a set of criteria for applying the concept of knowledge. Edgley actually states such a list of criteria in his paper. In order for someone to know that it will snow tomorrow, it is necessary (and sufficient), Edgley maintains, that three conditions be satisfied:

> First, it must be true that it will snow tomorrow. Second, he must not only believe that it will snow tomorrow but believe with that degree of conviction that amounts to being sure or certain. Third, his certainty must be justified not simply by its being true that it will snow, nor even by there being good reasons for thinking that it will snow, but by his having good reasons for being certain -- reasons good enough, that is, to exclude reasonable doubt

> that it will snow tomorrow. Fulfillment of these conditions would make this a paradigm case of knowledge. (Edgley 1970, 28-29)

By a "paradigm case of knowledge", Edgley means a case in which our application of the concept of knowledge is minimally "faltering and uncertain" (Edgley 1970, 28). Any case which satisfies these three criteria, then, will be a case, Edgley predicts, to which we will apply the concept of knowledge without hesitation or uncertainty. I take Edgley to hold, further, that where it cannot be shown that these three conditions are satisfied, we find it difficult to decide whether or not to apply the concept of knowledge. Here the attribution of knowledge may be altogether false, or it may be that we know something, but in a non-paradigmatic sense of _know_. Even such non-paradigmatic knowledge claims must apparently be justified by appeal to criteria, however, and it seems that the criteria appealed to must always be derived from those which define the paradigmatic concept of knowledge. Edgley, at any rate, follows this practice. In two arguments we shall look at presently, he claims that we know (the grammar of) English in a non-paradigmatic sense, because our knowledge of English grammar fulfills "laxer conditions, which are, so to speak, natural analogues of the stricter logical conditions" on knowledge (Edgley 1970, 33).[6]

Third, and finally, Edgley seems to require more or less direct verification, by standard methods, of the minor premisses in arguments for knowledge claims--the premisses which state that each of the relevant criteria has been satisfied. Thus, for example, unless we can show that someone has uttered a sentence that expresses P_1, which would be the usual or direct way to establish that the person believes P_1, the claim that he believes P_1 has not been established beyond reasonable doubt. Until that kind of verification can be found, the belief claim is "essentially disputable". I suspect that this attitude underlies Edgley's third objection as well. In order to determine what evidence a person has for a given belief, we generally ask him. But obviously we cannot expect very satisfactory results if we ask a naive speaker of English to state his reasons for believing some principle of English grammar. Perhaps this is Edgley's reason for maintaining that the grammar underlying one's competence as a native speaker is not literally inferred from the primary linguistic data one encounters as a child.

We have seen that Edgley rejects Chomsky's claim of innate linguistic knowledge, and the rationalist argument in support of that claim. I have suggested that Edgley places three requirements on arguments in support of knowledge claims, and that he rejects the rationalist argument because it fails to satisfy these requirements. I wish now to examine two arguments which Edgley offers in support of the claim that we have intuitive (tacit) knowledge of (the grammar of) our language. This examination will, I believe, furnish additional evidence that Edgley does indeed hold the views I have suggested concerning arguments for knowledge claims; it will also lay the foundation for some critical comments on Edgley's views, and for a defense of the rationalist claims about linguistic knowledge.

For expository convenience, I shall discuss Edgley's second argument first. Here is the central premise of that argument, in Edgley's own words:

> The claim that [a certain skill or competence] is knowledge is further strengthened precisely if the causal or quasi-causal explanation of how this competence was acquired appeals to facts that, if they were attributable to the person concerned as his reasons justifying the claim to knowledge, would logically support that claim. (Edgley 1970, 32)

An example will help to clarify this new criterion of knowledge. (This example also bears on another criterion, to be discussed later.)

> In this way, for example, a woman may be said to know intuitively or instinctively that her husband is worried or anxious. The force of these adverbs ' intuitively' and 'instinctively' is to imply on the one hand that the worry or anxiety were not directly observed in any publicly checkable sense, and on the other hand that she did not reason it out, come to the conclusion, or infer, that her husband was worried or anxious. We ascribe to her an intuition or instinct for things of this sort if she tends to be right about them on different occasions; and <u>this 'faculty of knowledge' can be explained, and confirmed as knowledge</u>, if on this occasion, for instance, she thought that her husband was worried or anxious because of the way he looked at her or because of the set of his shoulders as he walked or sat. The word 'because' here introduces a fact that is neither simply an explanatory cause of her thinking that he was worried or anxious, nor her reason for thinking that he was worried or anxious. But in being a fact that would have been her reason, and in her circumstances of close acquaintance a good one, if she had not only noticed it but had drawn her conclusion from it, the fact, in explaining her thought, also helps to justify it as knowledge. This is not a central case of knowledge,

> but it is sufficiently like a central case to make the application of the concept reasonable. (Edgley 1970, 32; emphasis added)

Edgley's new criterion of knowledge, expressed in the shorter of these two quotations, may be crudely restated as follows:

(9) _A_ knows that _p_ if there is some fact _q_, such that (i) _q_ is appealed to in the causal explanation of how _A_ came to believe _p_, and (ii) _q_ (perhaps in conjunction with other relevant facts which do not by themselves support _p_) logically supports _p_.

Despite its crudity, principle (9) is close enough to Edgley's intent for our purposes. If we had a better paraphrase than (9), that paraphrase would still succumb, I think, to my argument against (9) (cf. pp. 102-103 below). Moreover, the whole of Edgley's discussion seems to accord fairly well with (9).

Edgley's example, from the underlined clause on, is meant to shed light on the criterion formulated in (9). Suppose we call the woman Brenda, and her husband Harry. Brenda is able to determine Harry's moods. According to the underlined clause, this ability is an instance of knowledge. But further along in the passage, it becomes clear that it is her _thought_, namely

(10) Harry is worried.

(10), which is to be "confirmed as knowledge". The fact _q_

(11) Harry's shoulders have set S.

which confirms (10) as knowledge, by appearing in the causal explanation for Brenda's belief in (10) and by logically supporting (10), is (11). But (11) does not support (10) all by itself. It is only because of Brenda's "close acquaintance"

(12) Whenever Harry's shoulders have set S, Harry is (almost) always worried.

with Harry that (11) provides a good reason for believing (10). Her experience with Harry has made Brenda highly competent at determining his moods, and I suggest that (12) expresses part of the intuitive knowledge underlying her competence. And of course (11), when conjoined with (12), logically supports (10).

The principle (9) seems to be derived from the justification condition, the third of the conditions which (according to Edgley) define the paradigmatic sense of _know_. In order for someone to know _p_ in the paradigmatic sense, that condition required that the person in question _have_ "good reasons for being certain" that _p_ (Edgley 1970, 29); the individual must know facts which support _p_, and his belief in _p_ must have arisen from an inference to _p_ from those facts. The second clause of (9) requires that there _be_ good reasons for believing _p_, but (9) does not require that the knower have those (or any other) reasons for believing _p_. All (9) requires is that some of the facts that _could_ serve as the knower's reasons, should play an essential role in the causal explanation for the knower's belief in _p_.

Edgley appeals to the criterion (9) in giving a deductive argument for the claim that a native speaker knows (the grammar of) his language. Before we look at that argument, let me explain why, in alluding to this argument, I have heretofore enclosed the words <u>the grammar of</u> in parentheses. Edgley (1970, 32-33), in describing the thesis to be argued for, only calls it a thesis about "knowledge of the language", not about knowledge of a grammar. There are two reasons why I take him to be arguing for a thesis about propositional knowledge (of a grammar). First, if the thesis were that native speakers know their language, i.e. that they know how to speak and understand it, the thesis would be too trivial to warrant an argument. Second, Edgley appeals to a principle that is formulated in terms of logical support, and logical support is a relation that holds between statements or propositions. If certain facts provide logical support for what the speaker knows, then it must be a proposition (or something very much like one) that he knows. But a proposition that formulates a speaker's knowledge of his native language is a grammar. Edgley's usage to the contrary notwithstanding, then, I shall henceforth abandon cautionary parentheses.

Edgley's argument from principle (9) runs as follows:

> ... the child has been taught the use of words, i.e. ... he has been exposed to the input data, and ... these data constitute positive evidence that he could, if he were a scientist, cite in justification of his claim to knowledge. If this evidence is not only not evidence to him, in the sense in which he literally uses it to test hypotheses, but is moreover from the point of view of a formal reconstruction logically inadequate to generate what he consequentailly knows, an explanation of the child's competence that links input to output through a mental structure that the child shares with all other language users will confirm rather than undermine the characterisation of the output as knowledge; for it will confirm the important practical

> implication of this characterisation, that the child's ability to use words correctly was not just a fluke, but will survive in unforeseen circumstances in the future. This will not be a central case of knowledge; but ... it will be considerably closer to the centre than Chomsky's alternative of innate knowledge. (Edgley 1970, 33)

In this argument, the "fact" q which logically supports the speaker's belief (his grammar) is the body of primary linguistic data to which he was exposed as a child (D in the Graves-Katz argument). These data are not sufficient to support the grammar on their own, however; they support it only when conjoined with principles of universal grammar. Edgley presumably has such principles in mind when he speaks of a "mental structure that the child shares with all other language users". The data thus satisfy clause (ii) of (9). But these data play an essential role in determining the specific character of the acquired grammar (within the limits imposed by linguistic theory), so any adequate explanation of the acquisition of that grammar must take those data into account. Hence the data also satisfy clause (i) of (9).

One might think that (9) could also be used to argue that we know at least some principles of universal grammar. For many of those principles can be expressed in the form of

(13) Every psychologically possible human language has a grammar with property P.

statements like (13). We do not as yet have a physiological explanation for the origin of our belief in such principles,

(14) Every normal child is born with genetic structure _g_.

but when such an explanation becomes available, it will presumably appeal to facts about genetic structure, like (14). In conjunction with information about effects of structure _g_

(15) Normal children are capable of learning only those languages whose grammars have property P.[8]

on learning and maturation, these genetic facts will presumably entail something like (15). But (15) is equivalent to (13). Hence (14) satisfies clause (ii) of (9), for it (in conjunction with other relevant facts) logically supports (13), a principle of universal grammar. But (14) also satisfies clause (i), since (14) is appealed to in a causal explanation of how we come to believe (13). Principle (9), as it is now formulated, can be used to argue that we know universal as well as particular grammar.

There is a fairly natural way to amend (9) so as to exclude the argument just given. In the case of Brenda and Harry, the fact _q_ (that Harry's shoulders had set S) was a fact that Brenda was at least tacitly aware of. The set of Harry's shoulders registered somehow in her nervous system; it was the sort of thing psychologists refer to, I believe, as a perceptual cue. Similarly, in the language acquisition case, the primary linguistic data were facts that the child was aware of. In both cases, the causal explanation of the resultant belief cites not merely the truth of a certain fact (Harry's shoulders have set S; goed is not a correct past-tense verb form), but the subject's awareness of that fact. I propose, therefore, to add a third clause to (9), requiring that the subject, A, be aware of the fact q. The requisite sense of aware may be tricky to spell out, but let us suppose that the spelling can be done. Such an amended version of (9) would, I think, exclude the argument of the previous paragraph, for there seems to be no reason whatever to suppose that children are aware of their own or anyone else's genetic structure. The explanation of how we come to have beliefs like (13) can proceed perfectly well without any such awareness assumption.

I think Edgley would approve the suggested amendment of (9), motivated as it is by his own examples. But there is another reason he might be expected to welcome the amendment. For in the amended version of (9), we have a criterion of knowledge under which we know particular, but not universal, grammar. Edgley needs such a criterion, for we found that his arguments against innate knowledge work just as well against the claim that we know the grammar of our native language. (That was perhaps to be expected, for he was arguing that we have no innate linguistic knowledge of the paradigmatic kind, and he holds that our knowledge of particular grammar is not paradigmatic either). As amended, (9) provides the first motivation we have seen for Edgley's discrimination against innate linguistic beliefs.

But that motivation is not adequate, for (9) is false, even in its amended version. For a counterexample, let us revisit Brenda and Harry. Let us suppose that they have

(16) There is a burglar in our room.

taken a hotel room, and are now out on the town, trying to enjoy themselves. Harry, however, is worried. If asked, he might express his worry by asserting (16). Harry has not come to believe (16) through any process of reasoning; he merely happens

(17) Harry and Brenda's room is number 413.

to be a latently superstitious person, and as he and his wife checked in, he became aware (perhaps only vaguely) of the fact

(18) There is a burglar in every odd-numbered room on the fourth floor.

expressed in (17). Somehow, his awareness of that fact caused him to believe (16). But (17) logically supports (16), when conjoined with another fact, expressed in (18). (Perhaps the leader of a gang of thieves, himself superstitious, saw an oddly-shaped numeral four in a dream the previous night.) Thus (17) figures in the causal explanation of Harry's belief in (16); (17) logically supports (16) when conjoined with another relevant fact, (18); and Harry is aware of the fact expressed by (17). Clearly, though, Harry does not know (16) under these circumstances, though (9) (as amended) implies that he does.[9]

This counterexample shows that Edgley's argument for tacit knowledge of particular grammar is unsound, because it refutes (9), a key premise in that argument. The counterexample does not, however, refute the converse of (9). Though the antecedent of (9) does not state a sufficient condition for ascribing knowledge, it might (for all my counterexample shows) state a necessary condition on knowledge ascriptions. If that were the case, then Edgley would still be able to argue, by appeal to the converse of (9), that we have no innate knowledge of universal grammar.[10] And since the antecedent of (9) is weaker (easier to satisfy) than the justification condition, it would be plausible to argue, on the basis of (9)'s converse, that our innate linguistic beliefs do not even constitute a non-paradigmatic instance of knowledge. I shall comment on this possibility later, when I reply to Edgley's argument, based on the justification condition, that we have no innate knowledge. As we saw earlier, (9) is a weakened version of the justification condition, and it seems to me that (9) and the justification condition are subject to the same difficulty.

Before we turn to Edgley's other criterion of non-paradigmatic knowledge, I wish to point out an additional moral of the counterexample just adduced. As noted earlier, it seems to be Edgley's view that knowledge claims may be supported by appeal to criteria other than the usual ones (truth, certainty or belief, and justification), provided that the criteria appealed to are derived in some way from the standard criteria. Our discussion of (9) has shown that a criterion of knowledge must have more going for it than its relationship to one of the conditions that in Edgley's view define the paradigmatic sense of <u>know</u>. If Edgley's other criterion is more plausible than (9), its plausibility will rest on something other than simply its resemblance to one of Edgley's conditions on central cases of knowledge.

Edgley states his other criterion of non-paradigmatic knowledge in the following passage:

> ... as we relax the third condition in my paradigm, the condition requiring that the person to whom knowlegde is ascribed should have reasons justifying his assurance, the application of the concept demands, in compensation, more stringent conditions of other kinds. The first condition, or some analogue, is tightened up, and we require not simply that the person concerned should believe what is true, or more generally get something right, on one occasion, but that he should continue to get things of that sort right on other occasions: one and the same skill or competence is then attributed to him as being exercised on these different occasions, i.e. his getting these things right is an exercise of knowledge. (Edgley 1970, 32)

I shall call this the reliability criterion. I shall not argue that the reliability criterion is false. Rather, I shall suggest that this criterion is a plausible one, and that Edgley is mistaken about the <u>source</u> of its plausibility.

Recalling Edgley's example, Brenda had reliable beliefs about Harry's moods. Because of this reliability, "we ascribe to her an intuition or instinct for things of this sort" (Edgley 1970, 32). Now just what is the knowledge that Brenda exercises, on the various occasions when she assesses Harry's mood correctly? I suggest that part at least of her knowledge is expressed by sentence (12), which states that Harry is always worried when his shoulders have set S. When conjoined with a relevant datum, to the effect that Harry's shoulders now have set S, (12) entails the proposition Brenda believes on this occasion; conjoined with similar data on other occasions, (12) (or some similar statement) will similarly entail the other statements about Harry's moods which Brenda knows on those occasions.

Concerning our knowledge of particular grammar, Edgley writes: "The chief criterion for the application of the concept of knowledge in this case is the fact that the child is able to use words correctly, i.e. that he can get things right in a variety of situations" (Edgley 1970, 32-33). Now the reliability criterion requires that there be a class of statements concerning which the knower must exhibit reliably correct judgment. In Brenda's case, these were statements which, like (10), concerned Harry's mood. In the linguistic case, I suggest that the appropriate class of statements includes statements about the grammaticality and meaning of English sentences.

(19) _The men went home_ is grammatical.

(20) _Please pass the salt_ is a request that someone pass the salt.

Statements (19) and (20) are two examples. When a child uses the sentence _The men went home_, he gives evidence that he believes the proposition expressed by (20); when, in response to an utterance of _Please_ _pass_ _the_ _salt_, he passes the salt to the speaker, he gives evidence of a belief in the proposition expressed by (20). Gradually, a child's beliefs about meaning and grammaticality become highly reliable. This reliability, according to Edgley, constitutes grounds for maintaining that the child knows some proposition which entails all his correct judgments about meaning and grammaticality, and the grammar of English expresses just such a proposition.

The reliability criterion can also be used to argue that we know the theory of universal grammar, as it happens. For there is a class of statements (namely, the class of grammars of natural languages) such that, whenever we believe a statement in that class, it is true; and the theory of universal grammar, in conjunction with relevant data, entails each of the particular grammars. To be sure, most people only know the grammar of one language; they exercise only once in their lives the competence they have in virtue of their knowledge of universal grammar. But the scarcity of language-learning performances in most of our lives should not be taken to show that no competence underlies that performance. To adapt a remark of Edgley's, the child's learning to use words correctly was not just a fluke; it was a highly reliable performance. This reliability is, I think, generally recognized; few would

(21) For any human language L, if _A_ had been reared by speakers of L, _A_ would have learned the grammar of L.

dispute the truth of (21), but (21) expresses a very strong claim about the reliability of children as language learners.

Edgley's reliability criterion is quite plausible, I think. Now Edgley seems to think his criterion inherits its plausibility from the truth condition, but I think this plausibility has a different source: The reliability criterion describes conditions under which an ascription of knowledge has certain *prima facie* explanatory power. If someone's beliefs about matters of a certain kind (say, Harry's moods, or the meanings of English sentences) are very often correct, this reliability calls for an explanation. One natural explanation is that the person in question has some general knowledge about such matters, and that his judgments in individual cases are derived from that general knowledge. It is also natural to suggest (though Edgley does not) that the derivation of those highly reliable individual judgments takes the form of an inference or a deduction.

There may be other grounds for rejecting such an explanation, in terms of deduction from background knowledge. We may imagine that a certain person's winnings at the racetrack are based on an unusually good knowledge of horses, but we may later discover that he has been doping several horses in each race he bets on. But if no such evidence turns up, and if we are unable to discover any better explanation for the reliability of a person's beliefs on a certain subject, we do well to accept the attribution of knowledge. But it is misleading to suggest, as

Edgley does, that the attribution is justified by the satisfaction of some criterion peculiar to the concept of knowledge. The "criterion" involved is merely that if some collection of statements provides the best available explanation for some phenomenon or other, we ought to accept that collection of statements, simply on the basis of their explanatory value. This principle bears no special relationship to knowledge claims; it is surely not a "criterion for the application of the concept of knowledge".[11]

It may be worth pointing out, before leaving this discussion of Edgley's criteria of non-paradigmatic knowledge, that the certainty or belief condition has dropped out of sight altogether. We have seen a weakened version of the justification condition, and a strengthened version of the truth condition, but no version at all of the certainty condition. Why didn't Edgley feel the need for some such requirement, in the course of his discussion of non-central knowledge? The reason, I suggest, is that where a knowledge claim has explanatory value, it is simply unnecessary to provide direct evidence, showing that each of the defining conditions of knowledge (or "natural analogues" of those conditions) is satisfied. This would be true even if we knew what those defining conditions were, though as Gettier has shown, we do not know this.

Some pages back, we saw that Edgley has three objections to the rationalist argument for innate knowledge: first, any innate linguistic beliefs we have cannot be <u>justified</u> beliefs, even if they happen to be true; second, an examination of what language learners say provides no direct evidence for the claim that they believe (let alone know) the principles of universal grammar; and third, though we know the grammar of English, we do not literally know it "on the basis of ... evidence or data", so our knowledge of particular grammar does not satisfy the antecedent of Hume's principle. It is now time to take a critical look at these objections.

Edgley introduces his first objection by remarking that Hume's condition is "too stringent" a condition on knowledge. In support of this remark, he points out (correctly) that one's innate linguistic beliefs do not satisfy the justification condition. But the argument from Hume's principle can be seen as an attack on the justification condition; that is in fact how Edgley seems to see it (cf. Edgley 1970, 29-30). Edgley thus appeals to the justification condition in his objection to Hume's principle. But this appeal begs the very question at issue: must all (contingent) knowledge be justified?

The argument from Hume's principle is a _reductio ad absurdum_. To avoid the absurdity, we must reject one of the premises from which it was deduced. What are those premises? First, there are two premises missing from the list given above (p. 83). One of these is the justification condition, a controversial premise. The other is accepted on both sides: the premise that our innate linguistic beliefs lack justification.

Second, there is premise (1) (p. 83), which states that a speaker knows a particular grammar, and knows it on the basis of primary linguistic data. Edgley accepts the first assumption (cf. pp. 97-98 above). He denies the second, but I shall argue below (pp. 116-123) that he should not do so.

Both sides in the dispute accept premises (2) and (3). This leaves only one further premise: Hume's principle, which Edgley would deny. I shall try to block this move, and force the rejection of the justification condition, by arguing for

> (22) *A* does not know *p* if (i) *A* came to believe *p* on the basis of a valid inference from a set S of premises; (ii) *A* does not know any other set S' of premises which validly support *p*; (iii) *q* is a member of S; (iv) elimination of *q* from S would render the inference to *p* invalid; and (v) *A* does not know *q*.

(22), a somewhat revised version of Hume's principle.

To refute (22), we would have to find a proposition *p* that we know, which also satisfied conditions (i)-(v). Is there any such proposition? Well, no directly evident proposition can fill the bill (cf. p. 168 below, and the references cited there); condition (i) is meant to say that *p* must be a piece of "consequential knowledge" (Edgley 1970, 30). (Conditions (i), (ii), and (iv) have inductive as well as deductive validity in view, by the way.) Nor can any necessary proposition be a counterexample to (22). Any set of premises entails a necessary proposition, so if *p* is necessarily true, it will violate conditions (ii) and (iv).

Any counterexample to (22), then, must be indirectly evident, and logically contingent. Here is a possible counterexample, which I owe to Ned Block: Suppose that I have seen the Loch Ness monster swimming. I know, then, that Nessy can swim (this is _p_). However, I believed in Nessy's swimming ability long before seeing her. Someone told me that she was a huge eel (call this _q_). Knowing that all eels can swim, I inferred that Nessy could swim. But suppose that Nessy is really an elongated whale. Then _q_ is false, and I do not know _q_. I do, however, know _p_.

Condition (ii) is intended to block counterexamples of this sort. My seeing Nessy swim gave me additional (and adequate) reason to believe _p_; hence _Nessy can swim_ does not satisfy condition (ii). This condition imposes a very severe restriction on _p_, and makes (22) quite weak. It would be surprising if such a weak principle turned out to be false. But we have some linguistic knowledge that satisfies condition (ii), despite its stringency, and that is all we need for our _reductio_ against the justification condition.

Suppose a foreigner asks me whether sentence S is grammatical in English, and I have never before heard sentence S. As a native speaker of English, I know that S is grammatical. By the rationalist acocunt (cf. pp. 116-123 below) I know this on the basis of a deductive inference. My internal grammar of English is an intermediate step in that deduction. The ultimate premises are my primary linguistic data, and the theory of universal grammar. Without the latter premise, I cannot validly infer that S is grammatical; nor do I have any <u>other</u> reason to believe that S is grammatical, since I have never heard any English speaker use S. Now if I do not know the theory of universal grammar, then by (22), I do not know the grammar of English; nor do I know that S is grammatical. But as Edgley admits, this is absurd. We have some independent basis for accepting (22). The source of our absurd conclusion, then, must lie in the justification condition, and that is the premise we must reject.

We turn now to the second of Edgley's three objections to the rationalist argument for innate knowledge. According to this objection, it is "essentially disputable" whether we have any beliefs (and hence whether we have any knowledge) about universal grammar, because the rationalists are unable to show that naive subjects ever utter sentences expressing those beliefs. Otherwise put, the objection is that the rationalist argument fails to provide direct evidence that the belief condition is satisfied by our alleged innate knowledge. Against this objection, I shall argue that direct verbal evidence of the kind Edgley has in mind is not necessary for the establishment of belief claims, and that in any case, establishment of a belief claim is not a necessary step in every adequate argument for a knowledge claim.

On the first point, there are circumstances in which verbal evidence is not necessary to justify belief claims. A dog, for instance, may be said to believe (and indeed to know) that his supper is under the basement stairs, though of course he never utters sentences that express his belief. Our evidence for a belief claim in this case consists of the dog's behavior on returning from a hard day in Harvard Yard without food. He heads directly down the basement and under the stairs, salivating heavily, immediately on arriving home.[12] Or consider another example of "irremediably nonverbal knowledge", discussed by David Lewis (1969). Suppose you and I are rowing together in rhythm.

> ... I cannot describe ... how we are rowing ... but I can keep on rowing that way; I can tell whether you keep on rowing that way; later, I could probably demonstrate to somebody what the rhythm was; ... and so on. Now there _is_ a description that can identify the way we are rowing. We take 1.4 $\pm$ .05 seconds for the stroke and .9 $\pm$ .1 for the return, exerting a peak force of 70 $\pm$ 10 pounds near the beginning of the stroke, ... and so on, in as much detail as you please. But, as we row, we have no use for this sort of description. We can neither give it nor tell whether it is true if somehow it is given. (Lewis 1969, 63-64)

Nonetheless, it is reasonable under the circumstances to say that we know what rhythm we are rowing in. The examples of the dog and the rowers indicate that the availability of direct verbal evidence is not a necessary condition for belief ascriptions.

Edgley's second objection, and the concept of essential disputability on which it is based, seem to me to grow out of a latent operationism. To support a claim that someone believes a proposition _p_, we very often (perhaps even usually) try to show that on some occasion he has uttered a sentence that expresses _p_. Observation of verbal behavior is thus the "operation" we perform when we verify belief claims in the usual, standard, or direct way. Equating the meaning of belief claims with the method usually used to verify them, Edgley concludes that any belief claim not supportable by this method must be either false, or true only on some non-standard reading of _believe_. But since verbal expression of a proposition is neither a sufficient nor a necessary condition on belief in that proposition, surely it is a mistake to identify the meaning of a belief claim with the operation of observing the verbal behavior of the putative believer.

But there is another difficulty with Edgley's second objection. Let us suppose that the concept of belief is related to that of knowledge in just the way Edgley says it is. Edgley apparently thinks that this fact about the meaning of knowledge claims places a constraint on the methods or strategies available for substantiating knowledge claims. He seems to think that a legitimate argument for a knowledge claim must contain a separate subargument showing that the belief condition is met; and similarly for the truth and justification conditions. But of course it is not in general true that the analysis of a concept states the only set of conditions sufficient for the application of that concept.

Consider, for example, the concept of ellipticity. The analysis of this concept may tell us that a figure is an ellipse just in case it is the locus of all points, the sum of whose distances from two given points is a constant. One way to show that a given figure is an ellipse, then, is to find a pair of points within the figure, and show that many randomly chosen points on the figure have distances from these two interior points which sum to the very same distance. But there is another way to argue for the ellipticity of a given figure. We can also choose an appropriate system of Cartesian coordinates, and show that many points on the given figure satisfy an equation of the form $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$. The latter method is certainly an adequate way of arguing that the given figure is elliptical. But unlike the method suggested by the analysis of the concept of ellipticity, the latter method does not require that we determine distances from points on the figure to a pair of fixed points; rather, it involves finding distances from points on the figure to a pair of fixed lines.

As with ellipticity, so with knowledge: The conditions stated in the analysis of the concept of knowledge may not be the only set of conditions jointly sufficient for application of the concept. It may be that Hume's principle (equivalently, (22)) states such a set of conditions, for example, for all Edgley has shown. But Hume's principle does not require that we pay special, separate attention to the question of belief. A fortiori, it does not require that we present "direct" or "observational" evidence of belief. Edgley complains, in his second objection to the argument from Hume's principle, that the argument is insufficient because it adduces no direct evidence of belief in the proposition alleged to be known. But if the argument has true premisses (cf. Edgley's third objection, which is directed against premise (1)), and is deductively valid, it is hard to see why this objection need be taken seriously.

For several paragraphs, I have been talking about sets of conditions logically sufficient to warrant knowledge claims. I do not wish to fall in with what appears to be Edgley's view, however, that any adequate argument for a knowledge claim must show that some such set of conditions is satisfied. This view amounts to the requirement that knowledge claims can only be supported by deductive, and never inductive, arguments. For most empirical claims, we are willing to accept either deductive or inductive arguments, and I can see no reason whatever for treating knowledge claims differently.

Now let us consider Edgley's third and final objection to the rationalist argument. According to that objection, we do not literally know the grammar of English "on the basis of ... evidence or data"; that is, we do not literally infer our grammar from primary linguistic data. Therefore our knowledge of particular grammar does not satisfy the antecedent of Hume's condition, and the rationalist argument is unsound.

Edgley's third objection, like his second, reflects his operationist assumptions (and indeed Edgley did not distinguish these two objections himself). Usually, a process of inference leaves certain traces which we may bring to light by asking the subject such questions as, "How did you find out that p?" Clearly, few if any adult speakers of English are able to verbalize the inference which led to their knowledge of English grammar. But failure to observe the usual traces of an inference process does not show that the process is fictitious, as Edgley apparently believes.

J.J.C. Smart has discussed a somewhat similar situation in the philosophy of physics. There, the fictionalist (or as Smart calls him, the phenomenalist) may argue against the reality of electrons as follows: If someone claims that a certain physical object is present in a certain region, we can ordinarily check his claim by looking in that region, perhaps using an optical or an electron microscope. We conclude that the object is really there if we can see it, and that it is not really there if we cannot see it. But physicists freely admit that electrons are in principle invisible; the "look-and-see" test can never turn up evidence that an electron is present, however much we refine the art of microscopy. From this, the fictionalist concludes that electrons are not real physical objects.

Against this argument, Smart writes:

> One can readily admit, then, that there are theoretical reasons why however small we were we could not see the theoretical entities of physics. In this respect these entities do differ from bricks, microscopic crystals, and bacteria, and even perhaps from protein molecules. But is this a good reason for putting them into a different

> *ontological* category? The mere fact that there are *theoretical* reasons why they cannot be seen gives no ground for saying that they are in any sense fictions. Theory asserts the existence of the elementary particles and *also* explains their non-visibility. Surely we need not *fall* back to Berkeley and suppose that *esse* is *percipi*. (Smart 1963, 38)

There are two ways in which the situation in psycholinguistics differs from that in physics, and these two differences cancel each other out. First, there is to the best of my knowledge no explanation available for the adult's inability to recall and state the inference which led to his knowledge of particular grammar, though in physics there of course is an explanation of why electrons are invisible. On the other hand, the inference from universal grammar plus data to particular grammar is not altogether "unutterable", in the way in which electrons are altogether invisible. It is conceivable that a person might be able to reconstruct and state the argument which led him to accept the grammar of English, if he had access to an adequate theory of language learning, and to sufficiently detailed records of his linguistic experience in childhood.[13] According to present-day physics, however, nothing will ever make electrons visible.

These two features of the language-learning case seem to me not to affect Smart's point. The fact that an inference process was subconscious, and did not leave the usual sort of memory traces, does not show that it never took place. The comparison to Berkeley is an apt one. But Smart has an additional argument:

> ... if it be granted that we need not fall back into phenomenalism it may be replied that there is nevertheless no reason why we should not. I wish to argue, on the contrary, that there is a most telling theoretical reason why we should not adopt a phenomenalist interpretation. ... If the phenomenalist about theoretical entities is correct we must believe in a cosmic coincidence. That is, if this is so, statements about electrons, etc., are of only instrumental value: they simply enable us to predict phenomena on the level of galvanometers and cloud chambers. They do nothing to remove the surprising character of these phenomena. Admittedly the physicist will not be surprised in the sense that he will find these phenomena arising in unexpected ways: his theory will have instrumental value in preventing this sort of surprise. But he ought still, if he is reflective, find it surprising that the world should be such as to contain these odd and ontologically disconnected phenomena: i.e. the phenomena are connected only by means of a purely instrumental theory. Is it not odd that the phenomena of the world should be such as to make a purely instrumental theory true? On the other hand, if we interpret a theory in a realist way, then we have no need for such a cosmic coincidence: it is not surprising that galvanometers and cloud chambers behave in the sort of way they do, for if there really are electrons, etc., this is just what we should expect. A lot of surprising facts no longer seem surprising. ... On theoretical grounds, then, we should regard phenomenalism as both unproven and unplausible. (Smart 1963, 39; emphasis Smart's)

Smart takes his earlier argument to show only that "we need not fall back into phenomenalism". The "cosmic coincidence" argument just quoted is meant to establish the stronger conclusion that "we should not" adopt a fictionalist attitude toward theoretical entities. I am not sure that the difference in strength between the first and the second arguments is as great as Smart takes it to be, but the point of the "cosmic coincidence" argument is nonetheless well worth making.

The fictionalist must admit that current physical theory has a great deal of predictive (instrumental) value. He contends, however, that that theory is false when taken literally. Now a false theory cannot explain phenomena; it can at best only predict them. The phenomena remain unexplained, or as Smart says, "surprising". But there is another surprise, in addition to the phenomena themselves. How are we to explain the predictive success of a false theory? To dispel our surprise at that success, we need a theory that is true, when taken literally; such a theory will hopefully explain not only the phenomena, but the predictive capabilities of our current, literally false theory. For example, Newtonian mechanics was a literally false theory, as was known from the rotation of the axis of Mercury's orbit. Still, Newtonian mechanics had considerable instrumental value. The general theory of relativity explained the phenomena predicted by Newton's physics, and also explained how a literally false theory enjoyed such great predictive success. (Hempel 1966, 54).

Actually, no new theory is required to show that a current theory is literally false; one contrary-to-fact prediction is enough. Against the rationalist theory of language learning, however, Edgley adduces no such recalcitrant facts. Lacking such a factual refutation, he might be able to justify his fictionalist attitude toward tacit linguistic inference by putting forward an alternative theory of language learning, and providing a compelling methodological argument for its superiority over the rationalist account. This alternative theory might then be used to explain the ability of the false theory to save the appearances. Lacking such a theory, supported by methodological argument, Edgley's fictionalism is a tune whistled in the dark. It is *prima facie* implausible that a theory with great instrumental value should be a false, and hence a non-explanatory, theory. Smart's "cosmic coincidence" argument points out that implausibility, but *prima facie* implausibility may be dispelled by appropriate arguments, empirical or methodological. For the present, however, a realist interpretation of rationalist language-learning theory seems as reasonable as a realist interpretation of physics.

There is one other consideration affecting the prospects for a vindication of Edgley's fictionalism via the theoretical route. In attempting to formulate a true theory of language learning, Edgley will presumably be bound by operationist strictures. Hempel, speaking primarily of the physical sciences, has argued that "laws ... formulated at the observational level generally turn out to hold only approximately and within a limited range" (Hempel 1966, 77; cf. also Hempel 1952, 20-50). If operationally defined vocabulary is inadequate for the formulation of explanatory theories in physics, we may expect that psychological phenomena will prove even less amenable to explanation in operationally definable terms. For as Edgley himself points out, "behavioural, and in general observable, characteristics ... seem logically less than adequate" for the expression of criteria for the application of psychological terms (Edgley 1970, 28).

Edgley's third objection was directed specifically against the inference to a particular grammar from universal grammar plus data. But a similar operationist objection could be raised against the claim that native speakers tacitly infer facts about the grammaticality and meaning of particular sentences from the grammar of their language. After all, conscious formulation of inference <u>from</u> a particular grammar (of English, say) is no easier than conscious expression of the inference <u>to</u> such a grammar. Operationism implies that neither sort of inference really takes place.

This conclusion ought, I think, to embarrass someone who holds that we literally know the grammar of English, and particularly someone who holds this view on the basis of the kind of argument offered by Edgley. For according to this view, we know the grammar of English, and we know various facts about the grammatical and semantic properties of various English sentences; but the latter knowledge is not inferred from the former, although the appropriate entailment relations hold. What, then, is the psychological relationship between these two kinds of knowledge? In view of their close logical relationship, it seems grossly implausible that they should not be psychologically related in some way or other.

Moreover, Edgley holds that all paradigmatic instances of knowledge must be justified, generally by a process of inference. If we do not infer our knowledge about particular sentences from our knowledge of English grammar, then our knowledge about sentences must either be inferred from something else we know (but what?), or else our knowledge about sentences is not a paradigmatic case of knowledge (surely a counterintuitive supposition). Actually, Edgley's fictionalism with respect to tacit inference processes is inimical to his argument that we know the grammar of English. That argument appealed to a weakened criterion of knowledge, the reliability principle. The plausibility of that principle, I argued, rests on the potential explanatory value of the knowledge claims licensed by the principle. If someone's beliefs on a certain topic (e.g. Harry's moods) are highly reliable, then the explanation of this reliability may refer to some general background knowledge the person has. But such an ascription of general background knowledge has no explanatory value all by itself; it must be supplemented by the supposition that the person in question can and does use his background knowledge as a premise in inferences to his (highly reliable) particular beliefs. Without an accompanying inference claim, a knowledge claim licensed by the reliability condition loses its explanatory value, and thus its motivation. The moral of this argument is that one must either hold that we know both universal and particular grammar, or that we know neither. Edgley's fictionalism seems to commit him to the latter course.

To summarize this discussion of Edgley's three objections to the rationalist argument for innate knowledge: First, Edgley is quite correct in pointing out that our innate linguistic knowledge does not satisfy the justification condition, but that is no reason to hold that our innate linguistic beliefs are less than knowledge. Our knowledge of universal grammar must be exempted from the justification condition on pain of absurdity. Second, direct verbal evidence is not necessary to establish that an individual believes a certain proposition. In any case, it is possible to establish a knowledge claim without first establishing a related belief claim, and not all arguments that establish knowledge claims are deductive arguments. Third, Edgley's operationism and his fictionalism with respect to tacit inference processes are no more reasonable than are similar attitudes toward theoretical entities in physics. Thus none of these objections seriously impeaches the rationalist argument for innate knowledge of universal grammar.

Our discussion of Edgley's paper sheds some light, I think, on the relationship between the analysis of the concept of knowledge, on the one hand, and the explanation of various instances of knowledge, on the other. First, an adequate analysis of a concept does not and need not automatically provide explanations for all known exemplifications of that concept. My reply to Edgley's first objection might seem to rest on a denial of this principle, since that reply bears some resemblance to the following argument, which does assume the denial of the principle just enunciated: "If we incorporate the justification condition into our analysis of the concept of knowledge, then certain paradigm cases of linguistic knowledge will be unexplained. Therefore, the justification condition cannot be part of the analysis of knowledge."

This argument, like my reply to Edgley, has the form of a _reductio_; but this argument, unlike my reply, represents as absurd a situation that is entirely possible. There is nothing absurd in supposing that a particular instance of knowledge is _unexplained_, even if we had an adequate analysis of the concept of knowledge (and Gettier has shown that we do not). On the other hand, it is absurd to suppose that a certain paradigm case of knowledge is not _knowledge_, and that is the absurdity I pointed out in my reply to Edgley. (That reply assumed, of course, that we literally deduce those items of paradigmatic knowledge from tacit linguistic beliefs.)

We have also seen that an analysis of a concept does not establish a unique method for determining when it is correct to apply the concept. To analyze a concept is not to specify a set of operations that must be performed, and a set of results that must be obtained, in order to justify application of the concept. Application of a concept in a particular instance may be justified, for example, by showing that such application of the concept plays a role in the best available explanation of some phenomenon or other. But any number of "operations" may be involved in assessing the relative merits of competing explanations of a given phenomenon, and surely it is not necessary to spell out all these operations in an analysis of the concept of _knowledge_.

Finally, an explanation for some exemplification of a concept may be available before we have an analysis of the concept. Thus Gettier has shown that we do not as yet have an adequate analysis of knowledge, but the rationalist account seems to provide an explanation for our linguistic knowledge. Nonetheless, explanation requires some insight into the concept exemplified by the explanandum. On a deductivist theory of explanation, of course, we must know of at least one condition or set of conditions sufficient for application of the concept. Premise (5), on which the rationalist explanation of linguistic knowledge is based, meets this deductivist requirement, though of course (5) offers a great deal less than a full analysis of the concept of knowledge.

II. QUINE: TRUTH

We turn now to a thesis of Quine's: The thesis that radical translation is subject to a radical indeterminacy. Ultimately, I shall argue that Quine's thesis amounts to an empirical hypothesis about language learning, and that the available evidence weighs against this hypothesis.[14] ~~First~~, however, we shall examine Quine's argument for indeterminacy, and two unsuccessful counterarguments that fail to deal with Quine's assumptions about language learning. Before we undertake any of this, however, let us see what Quine's thesis is, and what it implies about linguistic knowledge.

Quine sums up the thesis as follows:

> ... the analytical hypotheses, and the grand synthetic one that they add up to, are only in an incomplete sense hypotheses. ... The point is not that we cannot be sure whether the analytical hypotheses is right, but that there is not even an objective matter to be right or wrong about. (Quine 1960, 73)

An analytical hypothesis is an hypothesis to the effect that a given word or phrase is synonymous with some other word or phrase, generally in some other language. One such hypothesis, then, is that the English phrase _prickly pear_ means the same thing as the Spanish word _nopal_. Given a system of such hypotheses about English and Spanish words, together perhaps with "supplementary semantic instructions" and syntactic explanations (Quine 1960, 70), we can translate English sentences into Spanish sentences, and _vice versa_. The trouble with such translations, according to Quine's thesis, is that some of them will inevitably be neither correct nor incorrect, but unverifiable.

It is not really necessary to "resort to a remote language", such as Spanish, in order to find examples of indeterminate synonymy claims, for according to Quine, "radical translation begins at home" (Quine 1969, 46). If the equation of Spanish _nopal_ and English _prickly pear_ is neither objectively true nor false, the same can be said about the equation, within English, of _hare_ and _rabbit_.[15] The indeterminancy of synonymy between words infects synonymy relations between sentences which contain

(23) _Polly saw a rabbit_ means the same thing as _Polly saw a hare_.

the words in question. If *rabbit* and *hare* are not *objectively synonymous*, then proposition (23) is neither true nor false, and hence cannot be known. But when the rationalist argues that English speakers have tacit knowledge of the grammar of English, one of the premises in his argument is that English speakers know propositions like (23). Tacit knowledge of particular grammar is postulated to explain the latter sort of knowledge. Innate knowledge of universal grammar is then postulated in order to explain our knowledge of particular grammar. But the argument for innate knowledge cannot get started unless it is admitted that we know some particular grammar, and Quine's indeterminacy thesis threatens a crucial premise in the argument that we know a particular grammar.

If Quine's thesis is correct, then we cannot know the *whole* transformational grammar of English. We still might know part of that grammar, however. Quine's thesis only impeaches certain semantic consequences of generative grammars, and it is an easy matter to rid a grammar of the disputed consequences: we need only cut all semantic information out of the lexicon.[16] The grammar would retain its syntactic and phonological components in their original form, and the lexicon would still describe the syntactic and phonological peculiarities of every word in the language. The truncated grammar would still have consequences

(24) *Polly wants a cracker* is grammatical in English.

(25) The first syllable of *telegraph* bears primary stress.

(24) and (25). Quine does not dispute the determinacy of syntactic and phonological judgments. It would be consistent with the indeterminacy thesis, then, to claim that we know propositions like (24) and (25), and to explain this knowledge by postulating tacit knowledge of a truncated generative grammar. Similarly, we could argue for innate knowledge of a truncated theory of universal grammar, by excising all constraints on the semantic part of lexicons in particular grammars.[17]

We could save many of the rationalist's claims about linguistic knowledge, even if we granted the truth of Quine's indeterminacy thesis. But with such an admission, we would lose some of our most important and philosophically interesting linguistic knowledge: knowledge about meaning and synonymy of sentences, and the general grammatical knowledge from which this knowledge is derived. A refutation of the indeterminacy thesis is therefore well worth attempting. Let us now take a closer look at the thesis itself, and at Quine's argument in support of the thesis.

Suppose that we have just been transported to the village of some tribe which speaks a language never before encountered by the outside world. None of the natives of this village speaks any language but the tribal one. Under these circumstances, Quine asks, how can we learn to translate native sentences correctly into English? Otherwise put, how can we acquire knowledge about the meanings of native sentences? Quine's answer is that we must take note of what stimulations prompt the natives to utter various sentences. Take, for example, the native sentence Gavagai. If natives utter this sentence only on espying rabbits, pictures of rabbits, toy stuffed rabbits, etc., we may suppose that Gavagai means something like Lo, a rabbit. (Quine 1960, 29)

Once we have amassed a small stock of such sentences, we can use them to elicit the natice words for Yes and No. For example, we might take advantage of a moment when our informant's attention is focused on a tapir, and utter the sentence Gavagai. If our original translation of Gavagai was correct, this remark is sure to elicit dissent. To learn the word for No, we need only note how the native expresses his disagreement, on this and similar occasions. The word for Yes can be elicited, hopefully, by uttering Gavagai in the presence of rabbits. (Quine 1960, 29-30)

With native Yes and No safely in our vocabulary, we can put our studies on a somewhat more systematic footing. We define the affirmative stimulus meaning of a sentence as the set of all stimulations that would prompt our informant's assent to that sentence. Similarly, the negative stimulus meaning of a sentence is the set of all stimulations that would prompt dissent from that sentence. The stimulus meaning of a sentence, finally, is the ordered pair of these two sets. To determine the stimulus meaning of a sentence, we use what I shall call the method of prompted assent and dissent: a selective "querying of native sentences for assent and dissent under varying circumstances"

(Quine 1960, 68). For simplicity, we may take stimulations to be "ocular irradiation patterns" (Quine 1960, 32). Also, it may be impossible for the native to distinguish between a rain dance and a war dance, if we cut the stimulations too short. Consequently, we must choose some appropriate standard duration for stimuli, say n seconds. This standard time period is called the modulus of stimulation. (Quine 1960, 32-33)

As our mastery of the language improves, we will no doubt be able to distinguish various kinds of native sentences, using the modest technical apparatus just described. Some sentences, Gavagai among them, will elicit assent (or dissent) from an informant only after he has been subjected to relevant stimuli. (If we blindfold him, and then ask, Gavagai?, he will presumably not know what to say.) Other sentences will not exhibit this dependence on recent stimuli. Consider, for example, the native sentence which translates as The coconuts are ready to harvest. The first time we think to ask whether the coconuts are ready to harvest, we may have to let our informant examine the appropriate trees. Once he has checked the trees, however, he may continue to assent throughout the day, the week, or even the fortnight (for all I know about coconuts), without each time having to revisit the village palm grove. Sentences like Gavagai, whose truth value[18] may change within a period less than or equal to the modulus of stimulation, are called occasion sentences. Those whose truth values persist for longer than the modulus, Quine terms standing sentences. (Quine 1960, 35-36)

Even within the class of occasion sentences, we will be able to distinguish two types of sentences. Some, like Gavagai, will have the same stimulus meaning for each member of the tribe. Each native will assent to this sentence on being shown a rabbit, and each will dissent on being shown a non-rabbit. The native version of He's a bachelor may not elicit such uniform response, however. If the jungle tribe is anything like our own, each native will know the marital status of only a few men. He will be ignorant of the marital status of all others. Thus it will not be sufficient to show our informant just any bechelor, in order to elicit his assent to the relevant sentence. We must show him a person whom he knows to be a bachelor. Another informant, from the other side of town, might have a very different circle of acquaintances, and consequently a very different stimulus meaning for the translation of He's a bachelor, i.e. a different set of assent-prompting (and dissent-prompting) stimuli. Where the stimulus meaning of a sentence is uniform throughout the community, as with Gavagai, Quine calls the sentence observational. Where stimulus meaning differs from one informant to another, the sentences may still be an occasion sentence, but a non-observational one. (Quine 1960, 43)

It is not just in the native language, of course, that we are able to distinguish standing sentences from occasion sentences, and observational from non-observational occasion sentences. The same distinctions can also be applied to English. The job of translating native sentences into English can now be described as follows. We must pair each native observation sentence with an English observation sentence that has the same meaning; and similarly for each non-observational occasion sentence, and for each standing sentence, of the jungle language. Briefly, we must pair each native sentence with a synonymous English sentence. It is initially plausible to suppose that two sentences are synonymous just in case they have the same stimulus meaning.

Like English, the jungle language presumably contains an infinite number of sentences. We want our theory of Junglese-English translation to be complete: it must give us the correct English translation for <u>every</u> Junglese sentence, and vice versa. Moreover, we want a "manageably limited" theory (Quine 1969, 3). But the method of prompted assent and dissent can never yield a theory that satisfies both these requirments. After a finite number of stimulus-and-query experiments, we have translations for at most some finite set of Junglese sentences, and our theory is incomplete. If it were possible to carry out an infinite number of experiments, we might get a complete theory, but not a manageable one: it would consist of an infinite list of sentence-to-sentence equivalences. On the practical side, the sentence-by-sentence approach to translation is grossly inefficient, since the meaning of one sentence often gives clues to the meanings of other sentences. Part of one sentence (a word or phrase) may recur in another sentence. If we can determine the meaning that sentence-part has in the first sentence, we will have a head start on the translation of the second sentence.

Both practical and theoretical problems thus attend the prompted-assent method of translation. To circumvent these problems, the linguist adopts the method of analytical hypotheses. But if Quine is right, this method of translation has a problem all its own: logical indeterminacy. Let us turn now to Quine's argument for the indeterminacy thesis.

Earlier on, we met the Junglese observation sentence _Gavagai_. On syntactic analysis, this sentence may turn out to consist of a single word, which also occurs in longer, more complex sentences. Given the stimulus meaning of the sentence _Gavagai_, the most natural analytical hypothesis about the word _gavagai_ is that it means _rabbit_. Quine argues that this analytical hypothesis has some highly counterintuitive rivals, and that no experiment can ever verify our intuitions in the matter.

Note, for example, that whenever we present our informant with a view of a rabbit, we also present him with a view of a part of the rabbit (his head, say, or his right side), still attached to the rabbit. Similarly, if we think of rabbits as objects with a temporal dimension as well as spatial ones, then every view of a rabbit is also a view of a time-slice of a rabbit, a temporal stage of the whole spatio-temporal rabbit. Also, when we point to a rabbit, we point to the rabbit fusion in Goodman's sense, and to where rabbithood is manifested. (Quine 1960, 52-53) Thus we could, consistently with the stimulus meaning of _Gavagai_, render _gavagai_ into English as _rabbit stage_, _integral rabbit part_, _rabbithood_, or _rabbit fusion_.

With a slight distortion of English grammar, we may grant the status of sentencehood to Rabbit, Rabbit stage, and the rest. Each of these laconic occasion sentences has the same stimulus meaning as Gavagai. The assumption of Quine's indeterminacy argument can now be put very simply. If two terms (such as gavagai and rabbit stage) are stimulus synonymous when used as occasion sentences, then there is no way to refute the claim that these two terms are synonymous in the strict, intuitive sense. Otherwise put, if the meanings of two terms cannot be experimentally distinguished when the terms are presented without preceding or following context, then no distinction will emerge in experiments with sentences that contain other words before and/or after the terms in question.

Quine indicates that he is arguing here for something more than "normal inductive" uncertainty (Quine 1960, 68). To illustrate, let us suppose that there are two kinds of rabbits in the tribal territory, the common furry variety, and a very rare hairless species. At first, we might encounter only furry rabbits, and conclude (wrongly) that Gavagai means There's a rabbit. (Actually, Gavagai means There's a furry rabbit.) Clearly it is possible that we shall discover and correct our error on some future occasion. Our mistake results from a deficient sample. All empirical hypotheses are subject to this kind of inductive uncertainty. The indeterminacy of radical translation arises when errors of translation cannot (even in principle) be experimentally detected, when natural-sounding translations and deviant translations alike "accord perfectly not only with behavior actually observed, but with all dispositions to behavior on the part of all the speakers concerned" (Quine 1969, 29).

In some passages, Quine writes as though indeterminacy of translation were a problem peculiar to the translation of a certain class of standing sentences.[19] Thus he writes that "observation sentences can be translated", subject only to "normal inductive" uncertainty (Quine 1960,68.). And it is not just observation sentences that "peel nicely", he writes. The same is true of "occasion sentences more generally, since the linguist can go native" (Quine 1960, 76). If Quine's argument about the word _gavagai_ is correct, however, then radical indeterminacy must attend the translation of more than just the standing sentences which contain occurrences of _gavagai_. The observation sentence _Gavagai_, for example, will have no determinate translation: _There's a rabbit stage_ will do as well as _There's a rabbit_.

What could have led Quine to suppose that indeterminacy occurs only in the translation of standing sentences? On this we can only guess, but there are some remarks scattered throughout Chapter 2 of _Word and Object_ which add up to a half-way plausible argument for this view. First, Quine points out that the affirmative stimulus meaning of a sentence should include only those stimuli which _prompt_ the native to assent to that sentence, not all stimuli that are _followed by_ assent. Suppose, for example, that our sentence is the native equivalent of Sam is away tracking a giraffe. Quine writes:

> All day long the native will assent to [this sentence] whenever asked, under all manner of irrelevant stimulations; and on another day he will dissent from it under the same irrelevant stimulations. ... there are formal considerations which, under favorable circumstances, can assure [the linguist] of the prompting relation. If, just after the native has been asked S and has assented or dissented, the linguist springs stimulation σ on him, asks S again, and gets the opposite verdict, then he may conclude that σ did the prompting. (Quine 1960, 30) [20]

To determine whether a given stimulus prompts assent, or merely happens to precede assent, the linguist needs "favorable circumstances": a change in the truth value (cf. note 17) of the sentence in question. Now occasion sentences change truth value more frequently than standing sentences, and some standing sentences retain their truth values longer than others. The truth value of an occasion sentence, such as He's a bachelor, may persist for a second or less. It is easy to utter this sentence five times in five seconds, while pointing consecutively to five different men: first a bachelor, then a married man, then a bachelor, etc. A standing sentence like The Times has come retains its truth value much longer. This sentence is false all day, until the Times arrives; then it remains true until the following day. The truth value of The crocuses are out stands even longer, changing only twice a year. (Quine 1960, 35-36)

Now after five years, a foreign linguist studying English would have (at most) ten observations bearing on the stimulus meaning of _The crocuses are out_.[21] For _He's a bachelor_, or _That's a rabbit_, the linguist may have hundreds or even thousands of relevant observations, carried out under "favorable circumstances". At the end of five years, he may formulate an hypothesis about the kind of stimulus that prompts assent to _The crocuses are out_. This hypothesis will have weaker inductive support, all other things being equal,[22] than a similar hypothesis about _That's a rabbit_. This is because our linguist has a larger number of evidential instances for his hypothesis about the occasion sentence. To this extent, translation of standing sentences may be more uncertain than translation of occasion sentences.

But Quine holds a stronger view than this; he also maintains that standing sentences have "sparser" stimulus meanings than occasion sentences (Quine 1960, 63). I take it he means that the affirmative stimulus meaning of a standing sentence contains fewer stimuli than the affirmative stimulus meaning of an occasion sentence. Now we have seen that our linguist, in the course of five years, observes only five stimuli that prompt assent to _The crocuses are out_. But this tells us nothing about the number of stimuli that _would_ prompt assent to this sentence. Unless I misunderstand Quine's claim about sparseness, that claim does not follow from his earlier remarks about standing sentences and occasion sentences.

In the translation of a standing sentence, there is usually a greater degree of inductive risk than in the translation of an occasion sentence. This much does follow from the remarks cited above. Those remarks do not point to any difference in the kind of risk involved, however. In particular, they do not show that bad translations are in principle undetectable, where standing sentences are concerned. To consider an extreme example, let us suppose that the Junglese tribe observes a year of jubilee every fifty years, and a centennial feast every other jubilee, when jubilee falls on a leap year. On the basis of indirect evidence, we have mistakenly translated the standing sentence S as _This is a leap year_. Actually, S means _This is the year of the centennial feast_. Is it possible in principle for us to detect our error, and correct it? Certainly. After a wait of at most four years, we will discover that S is not true in all leap years. Perhaps at this point we will decide that S means _This is the year of jubilee_. In principle, we can rectify even this mistake, though we may not be willing or able to wait forty-six more years to do so. It seems, then, that we can correct bad translations even of extremely long-standing sentences. All we need is patience and long life.[23]

Let me recall the point I made at the beginning of this digression: If Quine's argument about _gavagai_ is correct, then translations even of observation sentences (e.g., _Gavagai_)are indeterminate. Robert Kirk (1969) has attempted to refute Quine's thesis, and his refutation has drawn comment from Quine. Here is Quine's summary of Kirk's paper:

> ... my indeterminacy thesis was that two translators could disagree on a translation and still agree in all speech dispositions, in both languages, except translation. Kirk's reflection, to the contrary, is that the conflicting translations would entail conflicting speech dispositions also within the home language, at the level of indirect quotation. ... I grant Kirk his critical point: the phrase "except translation" in my statement of indeterminacy of translation needs to be elaborated so as to except also indirect quotation and related idioms of propositional attitude. All these devices reflect interlinguistic correlations intralinguistically. Niceties of formulation aside, however, Kirk's observation can be seen as challenging not the indeterminacy of translation but the determinacy of indirect quotation. (Quine 1968-69, 267)

According to Quine, Kirk's paper shows that intralinguistic indirect quotation and interlinguistic translation are in the same boat epistemically: either both are determinate or both are indeterminate. If this is correct, then by Quine's argument we should expect indeterminacy "within the home language, at the level of indirect quotation". Kirk does not (and does not mean to) establish this conclusion, but if he had, he would have brought a refutation of Quine's argument within easy reach.

If we were to apply Quine's argument to synonymy in the home language, we would argue that <u>hare</u> is not objectively synonymous with <u>rabbit</u>, and objectively heteronymous with <u>rabbit stage</u>, because the three terms are all stimulus-synonymous as occasion sentences. But there is an easy way to test the intersubstitutability of these terms in indirect-discourse contexts.

(26) I just saw a hare.

(27) She said that she just saw a rabbit.

(28) She said that she just saw a rabbit stage.

We procure a tape recording of a woman's voice, uttering sentence (26). We play this recording to a large number of English speakers, privately in each case to avoid interactions between subjects. After this stimulus, we query either sentence (27) or sentence (28), choosing the queried sentence at random. I feel safe in predicting that our subjects will overwhelmingly dissent from (28), and that they will overwhelmingly assent to (27).[24]

Surely these results would provide an objective basis for claiming that <u>hare</u> and <u>rabbit</u> are synonymous, whereas <u>hare</u> and <u>rabbit stage</u> are not. If Quine's reading of Kirk were correct, then we should have refuted the assumption underlying Quine's second argument: that if two terms are intersubstitutable in the null context, then they are intersubstitutable everywhere.[25] Below, however, I shall suggest why, on Quine's view, we might expect a difference between the intralinguistic and the interlinguistic cases. (Cf. pp. 162-163) For now, I wish to examine Kirk's argument itself. That examination will show, I think, that Kirk provides no grounds for linking radical translation to indirect quotation "within the home language".

Kirk (1969, 334-335) argues that it is "objectively discoverable" whether a given sentence or sentence sequence is intelligible to native speakers. In the same passage, he lays down the assumption that a system of analytical hypotheses is objectively incorrect unless it preserves intelligibility. Then Kirk (1969, 336) invites us to consider two individuals, Aman and Beeman, who are bilingual in Martian and English. Aman and Beeman disagree over the translation of a certain Martian sentence S.

(29) Aman's translation

$$S = sA$$

$$S \neq T; \quad sA \neq sB$$

(30) Beeman's translations

$$S = sB$$

$$T = sA$$

$$S \neq T; \quad sA \neq sB$$

Their views are summed up in (29) and (30).

Suppose now that Aman, in the course of a conversation in Martian, utters the sentence S. Beeman, in a subsequent English conversation, refers to this utterance as "your remark that sB". This is natural enough, since sB is Beeman's English translation for S. But Aman objects that he never made any such remark. Again this is quite natural, since Aman translates S, not as sB, but as sA. Aman and Beeman are agreed that sA and sB are in no sense equivalent English sentences. Now in expressing his objection to Beeman's description, Aman uses sentence (31).

(31) I said that sA; my actual words were 'S'.

This sentence (or sentence sequence, if you like) is intelligible to English speakers. Kirk does not argue the point, but surely it could be argued to Quine's satisfaction, if he were in any doubt. Sentence (31) has stimuli in both its negative and its affirmative stimulus meanings, and surely this indicates intelligibility. An English speaker would dissent from (31) if he had heard Aman utter some other sentence than S, or if a large number of Martian-English bilinguals assured him that sA was not a good translation for S; he would assent to (31) if he heard Aman utter S, and if a large number of bilinguals agreed that sA was synonymous with S.

But now consider how Aman and Beeman would translate (31)

(32) I said that ____; my actual words were ' ...'.

(33) P ___; Q ' ... '.

(34) PS; Q 'S'.

(35) PT; Q 'S'.

into Martian. Kirk (1969, 238-239) argues that both would have to translate the English context (32) in the same way, represented here in (33). Nor would Aman and Beeman disagree over how to treat the occurrence of the Martian sentence S in (31). Since S occurs between quotation marks, it need not (indeed, must not) be translated; but even if this were not so, S is already in Martian, so there would be no disagreement. The rub comes rather with sA, Aman's translation of S. Aman will of course translate (31) as (34). But Beeman cannot translate sA as S, nor as any sentence equivalent to S, for then he would have translated two non-equivalent English sentences (sA and sB) with equivalent Martian sentences, and his system of translation would be empirically refutable. Instead, Beeman translates sA as T, a Martian sentence that is in no way equivalent to S. Beeman's translation of (31) is (35).

We saw that (31) was intelligible in English. Are (34) and (35) intelligible to Martians? Well certainly sentence (34) is.

(36) I said that Brutus killed Caesar; by actual words were 'Brutus killed Caesar'.

(37) I said that Cicero was a Roman; my actual words were 'Brutus killed Caesar'.

It is the Martian equivalent of a sentence like (36), and obviously (36) is comprehensible. But (35), on the other hand, will not be intelligible, since it is like the English sentence sequence (37), which is objectively unintelligible. As Kirk says, (35) and (37) are alike in that "the sentence in quotation marks is in no way equivalent to the one introduced in the manner of a report" (Kirk 1969, 340).

The unintelligibility of (37), and derivatively that of (35), can be explained in either of two ways. We may regard (37) simply as a conjunction of two statements, both purporting to describe a single speech act. But both descriptions cannot be correct. If my actual words were <u>Brutus killed Caesar</u>, then I did <u>not</u> say that Cicero was a Roman. Thus sentence (37) can be regarded as a conjunction of incompatible statements, and this incompatibility cited as the source of the incomprehensibility of (37).

Alternatively, we may regard (37) as an argument with a suppressed major premise. The conclusion of the argument is the first sentence of (37), <u>I said that Cicero was a Roman</u>.

The second sentence of (37), _My actual words were 'Brutus killed_

> (38) If my actual words were 'Brutus killed Caesar', then I said that Cicero was a Roman.

Caesar', is one of the premises. The other, suppressed premise is (38), but this suppressed premise is false, as every English speaker knows. The argument as a whole is thus unsound, although the premise which makes it so is not explicitly present. On this view, the incomprehensibility of (37) is somewhat like that of a sentence with a presupposition known to be false. Both this and the explanation of the previous paragraph can be applied to (35) as well as to (37).

Either way, Beeman's system of translation gives an unintelligible Martian translation for a perfectly comprehensible English sentence, (31). This refutes Beeman. Aman, on the other hand, seems to have a system of translation that preserves intelligibility. His translation for (31), namely (34), is as intelligible as (31) itself. But notice that Aman and Beeman do not exhibit "conflicting speech dispositions ... within the home language, at the level of indirect quotation" (Quine 1968-69, 267). Both find the Martian sentence (34) intelligible, and (35) unintelligible. They agree that S and T are not equivalent Martian sentences, and that sA and sB are non-equivalent in English. They disagree over whether (34) is an accurate translation of (31), and over whether (31) is itself intelligible, but surely this latter disagreement is not a strictly intralinguistic reflection of their disagreement over translation.

For while (31) is technically an English sentence, it nonetheless presupposes a controversial claim about translation.

Kirk's refutation is not directed specifically against Quine's indeterminacy argument. In describing the speech dispositions of Aman and Beeman, Kirk has not mentioned terms corresponding to Quine's gavagai. He has only named certain Martian and English sentences (S, T, sA, sB), without making any assumptions about their internal structure. In dealing so abstractly with the sentences involved, Kirk has made it impossible to tell, simply by looking at Aman's and Beeman's translation schemes, which is the standard hypothesis and which is the counterintuitive challenge hypothesis (cf. pp. 147-148). This is a fatal flaw in Kirk's argument, for Aman's system of translation can be as easily refuted as Beeman's, and by an exactly parallel argument.

Suppose that Beeman, in a Martian conversation, uses the sentence T. Aman later alludes, in English, to "Beeman's remark that sB". Beeman denies having made such a remark. He insists

(39) I said that sA; my actual words were 'T'.

on (39). Now sentence (39) is intelligible to all English speakers with the exception of Aman and his kind. What about Aman's and Beeman's translations of (39)? Since Aman takes sA to be synonymous with S, he translates (39) as (40). Beeman,

(40) PS; Q 'T'.

(41) PT; Q 'T'.

who regards T as the proper translation for sA, renders (39) as (41). But this time it is Beeman's translation, (41), that shares the intelligibility of the original, and Aman's translation, (40), that is incomprehensible. Apparently, both Aman and Beeman give incorrect translations for the English sentence sA. What is more remarkable, we have been able to discover their errors without having any information on the structure of the sentences involved.

Such linguistic acumen on our part is really too good to be true. Neither Kirk's argument against Beeman, nor mine against Aman, suffices to refute either system of translation, for a correct system of translation would be refutable by a parallel argument, given only a single dissenting bilingual. A genuine refutation of Quine's indeterminacy thesis will have to make use of information about the internal structure of the sentences to be translated.[26] Specifically, these sentences will have to contain terms like gavagai. Let us shift our focus from Martian to Junglese, and see whether we can refute Quine's indeterminacy thesis by showing, contrary to Quine's argument, that gavagai cannot be translated as rabbit stage.

Let us suppose that there is a complete system of analytical hypotheses which accords perfectly with the speech dispositions of both language communities, and which translates gavagai as rabbit. We shall call this the standard hypothesis. Now consider the system of analytical hypotheses which results from the standard hypothesis when we replace the gavagai-rabbit equation with the analytical hypothesis that _gavagai_ means _rabbit stage_. Let us call this system of analytical hypotheses Quine's _challenge hypothesis_, or Q_O for short.

Now translation under the standard hypothesis always preserves stimulus meaning. This is what we mean in saying that the standard hypothesis accords perfectly with the speech dispositions of all concerned. Quine's challenge hypothesis also preserves the stimulus meanings of many Junglese sentences. All those which are free of the term _gavagai_ (and its derivatives) receive the same translations under Q_O as under the standard hypothesis, so of course Q_O's translations of those sentences are all right. If Q_O mistranslates any Junglese sentences, they will be sentences that contain _gavagai_. But we cannot use just any such sentence to refute Q_O. The Junglese occasion sentence _Gavagai_, for example, has the same stimulus meaning as its

(42) He gavagai kai ho gavagai.

(43) This rabbit is the same as that rabbit.

(44) This rabbit stage is the same as that rabbit stage.

Quinean translation, _Rabbit stage_.[27]

Suppose that the standard hypothesis translates the Junglese sentence (42) into English as (43). Quine's challenge hypothesis will then give (44) as the translation of (42). But this translation is objectively incorrect, for (44) does not have the same stimulus meaning as (42). To ascertain the stimulus meaning of (42), we may examine its standard translation, (43), with which it is _ex hypothesi_ stimulus-synonymous. Now an English speaker will assent to sentence (43) if, while uttering (43), we point twice to the same rabbit. But this same stimulus will elicit dissent if imposed during an utterance of (44). The stimulus meaning of (42) (which is the same as that of (43)) thus differs from the stimulus meaning of (44), and Q_0 is an objectively false hypothesis about translation.

We have refuted Q_0, but we have not refuted the indeterminacy thesis. Quine does not claim that the translation of gavagai as rabbit stage is immune against refutation if that is the only change we make in the standard hypothesis. Rather, he claims that we can save this counterintuitive analytical hypothesis from refutation "by compensatorily juggling the translation of numerical identity and associated particles" (Quine 1960, 54). Let us say that a system Q_i of analytical hypotheses is an extension of Q_0 if it includes all the analytical hypotheses of Q_0 which are not included in the standard hypothesis. In these terms, Quine is claiming that Q_0 has an empirically irrefutable extension.

Quine has in fact suggested a way to juggle the translation of identity so as to straighten out the problem with sentence (42). Instead of translating the Junglese relation term kai as is the same as, he suggests that we render kai as is a stage of the same object as (Quine 1960, 72). Let us make this amendment in Q_0, and call the result Q_1. Now Q_1's translation for (42)

(45) This rabbit stage is a stage of the same object as that rabbit stage.

is (45). Pointing twice to the same rabbit also counts as pointing to two stages of the same rabbit. Hence the stimuli which prompt assent to (42) will also prompt assent to (45). Unlike Q_O, the revised challenge hypothesis Q_1 preserves the stimulus meaning of (42).

(46) Yo gavagailo nah gavagai.

(47) All rabbit stages are rabbits.

(48) All stages of rabbit stages are rabbit stages.

But there is another sentence which Q_O translated incorrectly, and Q_1 inherits this defect. Sentence (46), translated as (47) under the standard hypothesis, gets paired with (48) under both Q_O and Q_1. Now (47) is stimulus contradictory, and hence its standard Junglese translation, (46) is also stimulus contradictory. But (48), the Quinean translation of (46), is stimulus analytic (on one natural reading of *stage*). It is hard to imagine a worse blunder in translation.

It will have been noticed that in concocting sentence (46), I made the Junglese term for rabbit stages a complex term, derived from *gavagai*. But this structural decision on my part is not essential to the argument. Thus suppose that Junglese also has

(49) Yo nematai nah gavagai.

(50) All rabbit stages are rabbit stages.

a syntactically simple term for rabbit stages, say *nematai*.[28] Then the standard translation for (49) is (47), but Q_O and Q_1

render (49) as (50). The Junglese sentence (49) is still stimulus-contradictory, and its Quinean translation, (50) is still stimulus-analytic.

Thanks to the existence of the Junglese term _nematai_, Q_0 and Q_1 are many-one mappings of Junglese terms into English ones. Under these challenge hypotheses, both _gavagai_ and _nematai_ get mapped to _rabbit stage_. The argument against the Quinean translation of (49) can be generalized, with an added restriction on systems of analytical hypotheses, to apply against all many-one challenge hypotheses.

We shall have to suppose that for any given class of synonymous English expressions, a given system of analytical hypotheses contains no more than one analytical hypothesis that mentions a member of that synonym class. The standard hypothesis already contains an analytical hypothesis which mentions the term _rabbit_, equating it with Junglese _gavagai_. Our restriction says that the standard hypothesis must not contain any analytical hypothesis equating _hare_ with some Junglese term, since _hare_ is synonymous with _rabbit_. Translation of sentences that contain _hare_ will thus be a two-stage process. First we shall have to replace _hare_ with _rabbit_, and in general to replace expressions not listed in the standard hypothesis with their listed synonyms. Once this step is complete, we can translate the resulting sentence by means of the standard hypothesis. The first step in this process requires a method for reducing any English sentence to a translatable normal form. Without such a method, the proposed constraint would destroy the ability to translate <u>any</u> English sentence into Junglese. We might accomplish the transformation to a normal form by means of a sort of intralinguistic system of analytical hypotheses not subject to our anti-redundancy constraint. Such intralinguistic systems of paraphrase will be needed in Junglese as well as English, since we intend to generalize our constraint to apply to both sides of a system of analytical hypotheses.

By thus generalizing our anti-redundancy constraint, we gain an important advantage. To see what that advantage is, let us take a second look at the way the sentence (49) caused the downfall of Q_0 and Q_1. The Junglese terms <u>gavagai</u> and

(51) Yo ____ nah gavagai.

(52) All ____s are rabbit stages.

<u>nematai</u> would still appear twice in the standard hypothesis, under the new constraint, since according to that hypothesis these terms are heteronymous. But their heteronymy implies that there is some sentence frame which takes on one stimulus meaning when <u>gavagai</u> fills its blank, and a different stimulus meaning when <u>nematai</u> fills its blank. One such sentence frame is (51). Its Quinean English equivalent is (52). The anti-redundancy constraint assures us that for <u>every</u> pair of Junglese terms t_1, t_2 mentioned in the standard hypothesis, there will be a sentence frame F ____, like (51), which takes on distinct stimulus meanings depending on whether t_1 or t_2 fills its blank.

Now Q_0 and Q_1 give precisely the same English translation for

(53) Yo nematai nah gavagai.

(54) Yo gavagai nah gavagai.

(55) All rabbit stages are rabbit stages.

nematai as for gavagai. For this reason, both (53) (previously (49)) and (54) are translated as (55), though (53) and (54) have distinct stimulus meanings. Thus one sentence that is incorrectly translated by Q_0 and Q_1 is (53). But there is another, more

(56) (Yo nematai nah gavagai) gesso (yo gavagai nah gavagai).

(57) (All rabbit stages are rabbit stages) iff (all rabbit stages are rabbit stages).

complex sentence we could have used to refute Q_0 and Q_1. That sentence is (56), which happens to be stimulus-contradictory. But (57), the Quinean translation of (56), is stimulus-analytic.

Every non-redundant system of analytical hypotheses translates some stimulus-contradictory sentence into a stimulus-analytic one, if (relative to the standard hypothesis) the system is a many-one mapping of Junglese terms into English. Suppose, for example, that the Junglese terms t_1 and t_2 are heteronymous according to the standard hypothesis, but that some challenge hypothesis translates them both into English as e.[29] There is at least one Junglese sentence frame, say F___, which is sensitive to the difference in meaning between t_1 and t_2. From this sentence frame, we form a new sentence frame with two blanks, using the Junglese biconditional operator gesso.[30] This new frame is F___ gesso F...; its English translation will be, say, G___ if and only if G... . Now the result of inserting t_1 and t_2 into the blanks of the Junglese sentence frame, will be a stimulus-contradictory sentence. But the Quinean translation of this stimulus-contradictory Junglese sentence will be stimulus-analytic, since e goes into both blanks of the English sentence frame.

We have just seen that all many-one systems of analytical hypotheses are refutable. It is easy to get rid of the many-one character of Q_0 and Q_1, however. In translating gavagai as rabbit stage rather than as rabbit, Q_0 and Q_1 have left rabbit without a Junglese equivalent. If we assign rabbit to nematai, the resulting system of analytical hypotheses (call it Q_2) is one-one, except perhaps for its treatment of the Junglese identity predicate kai.

Unfortunately, Q_2 is no better off than its predecessors,

(58) Yo nematai nah gavagai.

(59) All rabbits are rabbit stages.

again because of the way it treats sentence (49) (here repeated as(58)). For remember, (49), i.e. (58), is stimulus-contradictory. But its translation under Q_2 is (59) and (59) is stimulus-analytic. (I take it that rabbit stages come in various lengths, and that the longest possible stage of a given rabbit is identical with the whole spatio-temporal rabbit himself.)

There are other, more complex ways to derive a one-one hypothesis from Q_1. Rather than giving non-standard translations for just gavagai and nematai, we could switch the translations of a larger class of Junglese terms. Suppose, for example, that the standard hypothesis includes the five

(60)	t_1 (gavagai)	=	e_1 (rabbit)
	t_2 (nematai)	=	e_2 (rabbit stage)
	t_3	=	e_3 (rabbit part)
	t_4	=	e_4 (rabbithood)
	t_5	=	e_5 (rabbit fusion)

analytical hypotheses listed in (60). Clearly, this is only one of 5! distinct one-one mappings from $\{t_1, t_2, t_3, t_4, t_5\}$ to $\{e_1, e_2, e_3, e_4, e_5\}$. Another such mapping is shown in (61),

(61) $t_1 = e_2$

$t_2 = e_3$

$t_3 = e_4$

$t_4 = e_5$

$t_5 = e_1$

and there are plenty of others. Perhaps one of these will provide an irrefutable translation for (49), and for all other categorical Junglese sentences involving t_1-t_5.

Such widespread tampering with the translations of Junglese terms is liable to have adverse effects on the translations of sentences other than (49). To counter these effects, it would be necessary to adjust the translations of other Junglese expressions, and who can tell in advance whether it will be possible to make all the necessary adjustments? Also, notice that no such amendments as those listed in (61) would give us a satisfactory translation for sentence (46), our earlier counterexample.[31] It is not a very promising strategy, then, to amend Q_1 by altering only the translations of terms that are stimulus-synonymous (as occasion sentences) with gavagai.

Nor is it likely that Quine would resort to this strategy, when we consider that both counterexamples to Q_1, sentences (46) and (47), were (according to the standard hypothesis) categorical statements of the form A. In defending the translation of gavagai as rabbit stage, Quine's usual move is to juggle the translations of "numerical identity [cf. his dodge with sentence (42), pp. 148-150above] and associated particles" (Quine 1960, 54; emphasis added). Among these associated particles are quantifiers (Quine 1969, 2-3) and "categorical copula[s]" (Quine 1960, 70). Both our troublesome sentences, (46) and (49), contain the quantifier yo and the categorical copula nah. To give stimulus-contradictory translations for (46) and (49), Quine would most likely change the English glosses fo; nah and/or yo.

As we saw in connection with sentence (42), Quine does offer an example of how the identity predicate can be reconstrued to avoid counterexamples. Unfortunately, however, he never indicates what sort of adjustments he has in mind for quantifiers and categorical copulas. Nor have I been able to come up with any adjustments that would deal in a satisfactory way with (46) and (49). It seems, then, that we have no compelling reason to believe that we can render Q_1 irrefutable by altering its treatment of yo or nah.

On the other hand, I have clearly given no argument to show that all extensions of Q_1 are refutable. I have failed, as Kirk failed before me, to refute Quine's thesis. There is a common explanation for both failures: neither Kirk nor I have challenged Quine's assumptions about language learning. Quine assumes that people are psychologically able to acquire a translation scheme like the Q_i. Kirk and I have accepted this assumption, and we have tried (without success) to show that people with such translation schemes are objectively mistaken.

By accepting Quine's assumption, Kirk and I have given the entire game away, for this assumption is the heart of Quine's thesis.[32] In one passage, Quine sums up his thesis in the following way. Suppose we have found a bilingual whose translations are exactly those of our favorite system of analytical hypotheses. Now, Quine writes,

> My point remains; for my point is then that another bilingual could have a semantic correlation incompatible with the first bilingual's without deviating from the first bilingual in his speech dispositions within either language, except in his dispositions to translate. (Quine 1960, 74; emphasis added)

What is meant here by could? Quine can hardly be claiming logical possibility for the described situation, given his views about modality. Rather, I think he has psychological possibility in mind. His very next paragraph confirms this reading. There he hints that a linguist could translate gavagai as rabbit part, without doing violence to any "substantive law of speech behavior" (Quine 1960, 74). But the linguist's translation would violate a psychological law (in the relevant sense), if that law prevented natural bilinguals from translating gavagai as rabbit part. (The linguist aims to translate in just the way a natural bilingual would.) Quine's thesis, then, denies the existence of any such law.

A number of questions arise when we view Quine's thesis in this light: What bearing could a psychological hypothesis have on the logical determinacy of synonymy claims? What could have led Quine to accept this psychological hypothesis? Is the hypothesis true? We shall take up each of these questions in turn.

On the first point, I shall assume that two expressions are objectively synonymous (or heteronymous) in L just in case they are synonymous (heteronymous) in the idiolects of (almost) all speakers of L. To allow for questions of interlinguistic synonymy, we may consider the two languages spoken by a bilingual to be a single "union" language.[33] Thus an expression e of L is synonymous (heteronymous) with an expression e' of L' just in case e and e' are synonymous (heteronymous) in the idiolects of (almost) all speakers of L U L', i.e. in the idiolects of (almost) all bilinguals in L and L'.[34]

This criterion for the truth of synonymy claims is, I think, in accord with Quine's views on the subject. In one place, he writes:

> ... the question whether two expressions are alike or unlike in meaning has no determinate answer, known or unknown, except insofar as the answer is settled in principle by people's speech dispositions, known or unknown. (Quine 1969, 29)

Section 11 in _Word and Object_ provides an especially revealing case study in how people's speech dispositions are to settle questions of synonymy. There Quine points out a problem concerning non-observational occasion sentences, such as _Bachelor_ and _Unmarried man_. Presumably when I call someone a bachelor, I mean just what you would mean in calling him an unmarried man. But my _Bachelor_ and your _Unmarried man_ do not have the same stimulus meaning. Many stimuli that would prompt me to assent to (or dissent from) _Bachelor_ would move you neither to assent nor to dissent. These stimuli consist in exposure to men whose marital status is known to me, but not to you.

Despite all this, Quine points out, "the stimulus meanings of _Bachelor_ and _Unmarried man_ are ... identical for any one speaker ... sameness of stimulus meaning is as good a standard of synonymy for non-observational occasion sentences as for observation sentences as long as we stick to one speaker" (Quine 1960, 46). He proposes to count two sentences as synonymous in L, then, if and only if they are stimulus-synonymous for each speaker of L. This criterion is simply a more specific version of the criterion I proposed three paragraphs back.

I have been unable to find an explicit endorsement of such a criterion in Chomsky's writings, but I nonetheless suspect that he and most other proponents of linguistic rationalism accept some such criterion. For instance, Fodor and Katz (1963) argue that it is an empirical question whether _inadvertently_ and _automatically_ are synonymous in English, and that their interchangeability in the speech of a single individual (an eccentric baker) does not settle the question. They write:

> ... anything we learn about ourselves when we describe the language we speak is also something we learn about every other speaker of standard English _qua_ speaker of standard English. Conversely, anything we can learn about English by studying our own speech, we can in principle learn by studying the speech of speakers other than ourselves. ... Put it another way: any facet of a speaker's use of English that is not shared by other speakers is _ipso facto_ not relevant to a description of English. (Fodor and Katz 1963, 70)

In short, Fodor and Katz argue here that two expressions are synonymous in English just in case they are synonymous in the idiolects of a substantial majority of English speakers. Dissenting idiolects do not overthrow a synonymy claim if they form only a small minority.

There is, I think, good reason for rationalists and empiricists alike to accept this view about the truth conditions of synonymy statements. For suppose, first, that expressions e and e' are synonymous in a substantial majority of English idiolects. It is hard to imagine what grounds there could possibly be, under these circumstances, for claiming that e and e' really have distinct meanings nonetheless. These expressions may be non-synonymous technical terms in a jargon spoken by specialists in some field or other, but such a technical dialect surely enjoys no privileged status. Nor do the protests of the prescriptivist carry weight. He can at best show that e and e' used to be heteronymous, or that they remain so in some restricted population.

If, on the other hand, e and e' are heteronymous in almost all idiolects, what could justify the claim that there is really no difference in meaning between them? Here we are dealing with cases like that of Fodor and Katz's eccentric baker, and to such objectors, we say: sorry, but in describing e and e' as heteronymous, we weren't talking about *your* dialect.

Finally, suppose that the idiolects of English are more or less equally divided between those in which e and e' are synonymous, and those in which e and e' are heteronymous. In this case, we would be inclined to suspect that at least one of the expressions has no clear and established meaning in English.[35] We would be unwilling to describe e and e' as synonymous, but then we would also hesitate to describe them unequivocally as heteronymous. Where the idiolects show no unanimity, the question of synonymy has, as Quine says, "no determinate answer, known or unknown" (Quine 1969, 29). Perhaps, indeed, it is analytic that two expressions are synonymous in L iff they are synonymous in almost all idiolects of L.

We can now see how the psychology of language-learning affects the logical determinacy of synonymy claims. Synonymy claims are objectively true or false (in a speech community) only if idiolect variation is highly restricted. If the laws of language acquisition permitted different children to acquire widely different idiolects, then synonymy claims might gradually lose their determinacy.

Now, Quine holds that indeterminacy is consistent with the laws of grammar acquisition. What led him to this view? In seeking the answer, we must remember that Quine's indeterminacy thesis is a limited one. He does not hold that <u>all</u> synonymy claims lack truth value. Quine even concedes the truth of certain statements about synonymy between <u>terms</u>, at least within English, or in "languages whose translations of 'all', 'are' and '=' are somehow settled in advance" (Quine 1960, 55). All English speakers have substantially the same speech dispositions toward <u>bachelor</u> as toward <u>unmarried</u> <u>man</u>. And it is not just that we all more or less unconsciously use these terms interchangeably in all but a few contexts. We also feel intuitively that these words are synonymous.

Now what could be the source of our intuitions that <u>bachelor</u> and <u>unmarried</u> <u>man</u> are synonymous? One possibility is that our intuition is based on a belief about the speech dispositions of the rest of our community. Each of us could have made "an implicit sociological guess that [even] under extraordinary stimulation, most people would hold <u>bachelor</u> and <u>unmarried</u> <u>man</u> coextensive" (Quine 1960, 56). But Quine does not think that our intuitions about synonymy arise in this way.

Rather, he suggests, it is a peculiarity in the way we learn the term bachelor that gives us our intuition. As an occasion sentence, Bachelor is non-observational; Indian nickel, by way of contrast, is an observational occasion sentence. Now when we learn the term Indian nickel, we can and do learn it by "learning directly to associate the term with sample objects" (Quine 1960, 56). This is possible, because being an Indian nickel is an observable property. But the property of being a bachelor is not an observable property. Hence we cannot very well learn the term bachelor simply by observing a collection of sample objects to which the term applies. Rather, we learn this term by "learning appropriate associations of words with words" (Quine 1960, 56). Specifically, we learn bachelor by associating it with the phrase unmarried man.

> One looks to 'unmarried man' as semantically anchoring 'bachelor' because there is no socially constant stimulus meaning to govern the use of the word; sever its tie with 'unmarried man' and you leave it no very evident social determination, hence no utility in communication. ... it is only the few verbal links that give the terms the fixity needed in communication (Quine 1960, 56)

Now perhaps we do learn bachelor through word-word, rather than word-object, associations, and perhaps this is why we have the intuition that bachelor and unmarried man are synonyms. Still, the content of this intuition might be a "sociological guess". Indeed, when people use the term bachelor, they do so in the conviction that their association of bachelor with unmarried man is not idiosyncratic, but conventional (cf. Lewis 1969), and widely shared in our society. Recall Fodor and

Katz's point: when we study English grammar, we learn something not just about ourselves, but about every other English speaker as well.

Strong intuitions about synonymy are not confined to terms that are non-observational when used as occasion sentences, however. The term _hare_ (and for that matter, _lapin_) is intuitively synonymous with _rabbit_, and emphatically _not_ synonymous with _rabbit stage_. By Quine's account, we learn to use these terms by associating them with ostended objects, rather than with other words. Ostension, however, is according to Quine, powerless to distinguish rabbits from their stages. Whence, then, the strong convictions about the semantic relationships between _hare_ (or interlinguistically, _lapin_) on the one hand, and _rabbit_ and _rabbit stage_ on the other? And more particularly, does a language learner encounter evidence that would _justify_ the convictions he acquires?

Quine and the rationalistic linguist give very different answers to these questions. Let us tend first to Quine's answers. He can, I think, provide a plausible explanation for the feeling we have about _hare_ (or _lapin_) and _rabbit stage_; or rather, I think I can supply such an explanation on his behalf. A child (or a Frenchman) learning English is almost certain to learn (quite possibly by ostension) _rabbit_ before he learns _rabbit stage_. We rarely talk about rabbit stages, and since _rabbit_ is syntactically a constituent of _rabbit stage_, it makes pedagogical sense to introduce the former term first. Perhaps, then, rabbit stage is normally learned, not by ostension, but by "learning appropriate associations of words with words", i.e. by definition in terms of the antecedently available word rabbit.[36] The definition will very likely make it clear that rabbit stages are not rabbits. Quine himself, for example, introduces the term by suggesting that gavagai may refer "not to rabbits after all, but mere stages, or brief temporal segments, of rabbits" (Quine 1960, 51). This is doubtless where many of us first encountered the term rabbit stage. Similarly, hare may be learned, not by ostension, but by definition in terms of the more commonly used rabbit.

On this account, the intuitive synonymy of hare (lapin) and rabbit, and the intuitive heteronymy of hare (lapin) and rabbit stage,[37] arises from causes similar to those which make bachelor and unmarried man intuitive synonyms. In both instances, we have learned "appropriate associations of words with words", rather than associations between words and sample objects. The claim is that although rabbit stage perhaps could be taught by ostension, it is normally introduced by definition. Where this term is taught ostensively, a verbal warning not to confuse this term with rabbit may accompany the ostension, of course.

To eliminate the possibility of a language learner being taught word-word associations, Quine concentrates on "radical translation, i.e., translation of the language of a hitherto untouched people". Thus "all help of interpreters is excluded" (Quine 1960, 28). The linguist (or the incipient "natural bilingual",(Quine 1969, 6) must base his interlinguistic "associations of words with words" solely on ostensive evidence.[38] There is no pre-existing population of bilunguals available to inculcate "a traditionally evolved dictionary" (Quine 1969, 5), a uniquely "correct" system of analytical hypotheses. Indeed, on Quine's criterion of synonymy, there are no correct analytical hypotheses. For "the question whether two expressions are alike or unlike in meaning" (Quine 1969, 29) is settled, if at all, by the speech dispositions of the relevant speech community, and we are being asked to consider a situation in which the relevant bilingual community does not as yet exist. When our linguist becomes the first (and only) bilingual in the two languages involved, his system of analytical hypotheses will on Quine's criterion be correct, whatever it is, provided only that it preserve stimulus meaning.

The rationalist may concede that where radical translation is concerned, ostension provides the only available evidence about the meanings of Junglese expressions. He may concede that ostensive evidence is not sufficient, all by itself, to settle questions of translation. He denies, however, that ostension provides the language learner's (or the linguist's) only information about the language he is learning. Besides the evidence he gathers from speakers of Junglese, the language learner (and derivatively, the linguist) has knowledge of the principles of universal grammar.[39] These principles help the child to make his "implicit sociological guess". Taken together with the evidence, the principles of universal grammar rule out many semantic hypotheses that are compatible with the evidence alone. In fact, the language learner is able to rule out all systems of analytical hypotheses but one, using the principles of universal grammar as well as the evidence.

Actually, the language learner does not directly acquire a system of analytical hypotheses, I suspect.[40] Rather, he acquires a grammar of Junglese, including a lexicon in which the meanings of Junglese terms are represented in the vocabulary of universal semantic theory (cf. Katz 1972, 116n.). The learner already has a grammar of English, in which the same vocabulary is used to describe the meanings of English terms. Junglese and English terms can thus be correlated via their representations in the vocabulary of a language-independent semantic theory: If a Junglese term has the same lexical reading as an English term, the child's two lexicons entail, in effect, an analytical hypothesis: the hypothesis that these two terms are synonymous. In this way, a pair of back-to-back lexicons yield a whole system of analytical hypotheses. Together with the associated grammars, this system yields translations for all sentences of both languages.

According to the rationalist hypothesis, each child is born with the same information about linguistic universals. All children who are exposed to a typical sample of Junglese usage will thus develop substantially the same Junglese grammar, up to idiolect variation. (For an argument that such variation need not jeopardize the determinacy of synonymy relations, cf. the Appendix.) Similarly, children who encounter distinct but typical corpora of English data will all come to have the same English grammar. The rationalist claims, then, that all bilinguals who have had access to a normal body of evidence about both English and Junglese, will exhibit the same set of dispositions to translate. Two bilinguals cannot have radically different translations of the same sentence, unless one of them has been exposed to a radically atypical body of evidence in the process of learning one of the languages in question.

This means, by the way, that the rationalistic linguist claims the support of an hypothetical population of bilinguals, even when he engages in radical translation. "True enough", he admits, "there does not now exist a community of Junglese-English bilinguals. But if there did exist such a community, I can predict that (almost) all its members would have such-and-such translation dispositions." (Cf. note 39) The rationalist, in justifying his radical translations, uses a subjunctive version of the kind of argument Quine would use to show the synonymy of bachelor and unmarried man. Similarly, the natural bilingual makes a sociological guess about the translation dispositions of an hypothetical society of bilinguals.

It seems fair enough to appeal to the speech dispositions of potential as well as actual bilingual populations. The point of contention between Quine and the rationalist is the empirical claim I made two paragraphs back: that two bilinguals, both exposed to typical bodies of English and Junglese sentences during the learning process, <u>cannot</u> have radically different translations of the same sentence. Quine denies this. Quine also holds that intralinguistic indeterminacy is psychologically possible. I think Quine may have come to this position by reasoning from the supposed limitations of ostensive learning (cf. pp. 162-163 above). This leads us to our third question: As an empirical hypothesis about language learning, how well does Quine's indeterminacy thesis accord with the available evidence?

Quine's position is difficult to refute, because it is comparatively weak. Quine claims that variation among bilinguals (or among monolinguals, with respect to indirect quotation) <u>can</u> develop, but he does not specify the conditions under which he expects that it <u>will</u> develop. His discussion of the synonymy between <u>bachelor</u> and <u>unmarried man</u> tells us something about the circumstances in which Quine expects variation <u>not</u> to develop, however. Words that are learned via "appropriate associations of words with words", and not merely by word-object associations, (ostension), will typically stand in determinate synonymy and heteronymy relations with other words and phrases.

In testing a given pair of expressions for synonymy, then, we must assure ourselves that the subjects have not learned one of the expressions by definition in terms of the other. Otherwise, any objective synonymies we uncover will be useless against Quine's indeterminacy thesis. Within the limits imposed by this restriction, we shall have to vary the conditions of learning in every way that could conceivably lead to the development of idiosyncratic patterns of translation behavior. If all bilinguals[41] translate in the same way, despite wide variations in learning experience, then the rationalist hypothesis will be confirmed, and Quine's denial of that hypothesis proportionately disconfirmed, though not decisively refuted. If, as I suspect, Quine's indeterminacy thesis is false, then we will be able to achieve any desired degree of disconfirmation of that thesis.

There is one other kind of experiment that might be useful in settling the issue. The rationalist predicts that among all the hypotheses compatible with the evidence alone, language learners will have a strong preference for one hypothesis in particular, and a strong prejudice against all others. Quine, on the other hand, seems committed to the view that there is no psychological basis for any such prejudice. To test these predictions out, we might depart from the policy of isolating our subjects from all other bilinguals. We might deliberately try to convince them, for example, that the proper translation for <u>lapin</u> is <u>rabbit stage</u>. (This one deviant analytical hypothesis will call for adjustments in other analytical hypotheses, as we have seen.) If this suggestion encountered severe resistance, the rationalist hypothesis would be confirmed. If our subjects accepted this hypothesis readily, and integrated it smoothly into their performance as translators, this would confirm Quine's views.

Experiments of the latter sort have never, to the best of my knowledge, been attempted. But many people _have_ learned a second language without being contaminated by contact with other bilinguals. The "direct method" of foreign language instruction is based on the supposition that explicit instruction in translation hinders, rather than facilitates, the process of second-language learning. For this reason, instructors at "direct method" or "total immersion" language institutes carefully avoid using English, and outside of class, students are fined for speaking English with faculty or fellow students. I have never heard it reported that students emerged from such an institute translating _lapin_ as _rabbit stage_.[42]

It may be objected that this is an artifact of the language-learning situations. In English, we talk about rabbits much more often than we discuss rabbit stages. Perhaps if French instructors were to use _phase d'un lapin_ frequently, and unadorned _lapin_ rarely, there would be less unanimity in their students' translations. Or perhaps English and French share some structural peculiarity which facilitates translation between these two languages.

Both objections are of course subject to empirical test. In principle, the French instructors can control the relative frequency of _lapin_ and _phase d'un lapin_ in their conversations with students, though they may find this difficult in practice.[43] We can search for a more suitable test language than French, once we have been told what structural feature of French is supposed to be responsible for the artifactual ease and uniformity of translation between French and English. If every language in the world turns out to have the structural feature in question, this will be *prima facie* evidence that all possible human languages have this feature. In principle, we could test *this* hypothesis by attempting to teach English speakers a concocted language (perhaps a structurally doctored-up French) that lacked the suspected feature.

But enough of these imagined objections, and imagined experiments to test them. The point is this. Quine's indeterminacy thesis is an empirical hypothesis about the kind of evidence (selective reinforcement) needed to establish synonymy relations within or between languages. For practical reasons, it is difficult to test the implications of this hypothesis with respect to intralinguistic synonymy. Our information in the interlinguistic case is scanty, anecdotal, and not altogether conclusive, but that information seems to suggest that Quine's thesis is more likely false than true. The semantic knowledge native speakers have about the sentences of their language(s) is in no great jeopardy from Quine, since there are no strong grounds to suppose that Quine's indeterminacy thesis is true, and some grounds to suppose that that thesis is false.

III. STICH: BELIEF

Edgley's argument against innate knowledge was based on the justification condition: the requirement that in order to know a proposition, one must have good reasons for believing that proposition. Quine's indeterminacy thesis threatened our semantic knowledge (and hence our semantic knowledge at the level of particular and universal grammar) because of the truth condition: the principle that we can only know propositions that are true. Neither Edgley's argument nor Quine's proved compelling, however. It remains to consider whether our alleged knowledge of particular and universal grammar satisfies the belief condition: the requirement that in order to know a proposition, one must believe that proposition. Relative to this question, we shall examine a paper by Stephen Stich, in which he argues that "the speaker has no such beliefs" as those represented in the grammar of his language, or in linguistic theory (Stich 1971, 494).

(62) John overestimated himself.

(63) *John overestimated themselves.

Graves, Katz, et al. (in press) point out that a native speaker of English is able to determine whether or not any given string of English words is grammatical. He can determine, for example, that (62) is grammatical and that (63) is not, even if he has never before encountered either (62) or (63). To explain this ability to determine the grammaticality of strings arbitrarily drawn from an indefinitely large class, Graves, Katz, et al. postulate that English speakers know the grammar of English. Stich suggests that a weaker hypothesis would suffice to explain the open-ended ability of English speakers to recognize grammaticality.

This suggestion arises in the course of an argument intended to refute a view not held by the people Stich means to refute. Stich writes:

> My opponent places the time of acquisition of knowledge about particular sentences in childhood; I would maintain we have no such knowledge until we hear the sentence in question. ...on the view I would disparage, the beliefs speakers express when we inquire after their linguistic intuitions are long-standing beliefs. The speaker acquired them on learning his language. The questioning we subject him to merely brings these beliefs to consciousness... (Stich 1971, 493,495)

Thus on the view Stich thinks he must refute, every English speaker knows that sentence (62) is grammatical, whether or not he has ever encountered sentence (62). Indeed, every native speaker knows, concerning each of the infinitely many sentences of English, that that sentence is grammatical.

I know of no evidence that Chomsky ever held this implausibly strong view. But however that may be, Graves and her colleagues repudiate this view early in their reply to Stich. They "agree that the knowledge that a particular expression has certain grammatical property is only acquired upon encountering the expression, not when the language is learned by the child" (Graves, Katz, et al. in press). This agreement deprives Stich's conclusion of its interest, but his argument for that conclusion nonetheless remains worthy of our attention.

When I hear a sentence for the first time, and come to believe that the sentence is grammatical, my belief cannot be a long-standing one, Stich argues, because such beliefs are "directly evident perceptual judgments", in Chisholm's sense (Stich 1971, 496; cf. Chisholm 1957, 1966). Now directly evident perceptual reports, Stich writes, "commonly express something known to the reporter", but they "need no further evidence or justification" (Stich 1971, 495). Moreover, such reports are not "open to further justifications" (Stich 1971, 496). If this is correct, then it is neither necessary nor

(64) _The cat is on the mat_ is grammatical in English.

possible for an English speaker to justify his belief in (64), in order for that belief to count as knowledge. _A fortiori_, it is not necessary for him to justify that belief by tacit deduction, nor to know (or believe) a grammar that would provide

(65) I seem to see a yellow object.

a basis for such a deduction. In Stich's view, our knowledge of (64) is analogous to our knowledge of propositions like (65). In both instances, we "just know" the proposition in question, directly and immediately. In neither case does our knowledge rest on a foundation of prior knowledge. And if knowledge of a grammar is not needed to explain our knowledge of (64), then there is no reason to postulate even so much as belief in a grammar.

Now Graves, Katz, _et al_. argue that propositions like (64) are not directly evident, as Stich claims, because "there is always evidence that bears on their truth or falsity, e.g., evidence regarding the intuitions and behavior of other speakers, comparison with other sentences, _etc_. Hence, judgments [like (64] are capable of justification", and this shows that they are not directly evident in Stich's sense (Graves, Katz, _et al_. in press). One might think that if (64) is not directly evident, then surely (66) must be. Not so, according to Graves and her

(66) _The cat is on the mat_ is grammatical in my idiolect.

associates. Again, "there is always evidence available about whether a sound sequence is well-formed in an idiolect", because "any sentence in a speaker's idiolect is related to infinitely many other sentences" in that idiolect (Graves, Katz, _et al_. in press).[44] These critics of Stich concede, however, that "judgments

(67) _The cat is on the mat_ seems to me to be grammatical in my idiolect.

about what sound sequences _seem_ well-formed in our own idiolect [for example, (67) above] _might_ be directly evident in Stich's sense" (Graves, Katz, _et al_. in press; emphasis added).

Now if any proposition is directly evident, (65) is. Unfortunately, we can argue that (65) is _not_ directly evident, using the same kind of argument that Graves, Katz, _et al_. direct against (64) and (66). There may be physiological evidence to show that I seem to see a yellow object. This means that propositions about visual appearances, such as (65), are "capable of justification" (Graves, Katz, _et al_. in press), and hence (according to the Graves-Katz argument) not directly evident.[45]

Perhaps, then, (64) and (66) really are directly evident, despite the Graves-Katz argument. But if judgments like (67) are directly evident, as Graves and her colleagues suggest, then it seems to me that knowledge of propositions like (64) is explicable without assuming knowledge of a particular grammar. The explanation I have in mind might still appeal to tacit inferences

(68) Usually if a sentence seems to me to be grammatical in my idiolect, then it really is grammatical in my idiolect.

(69) Usually if a sentence is grammatical in my idiolect, then it is grammatical in standard English.

from tacitly known premises, but the grammar of English would not be among those premises. Rather, the premises would be (68), (69), and propositions like (67). Propositions (67), (68), and (69) seem[46] to support (64), though without entailing it, as the grammar of English does.

The plausibility of this explanation depends on whether or not we can be shown to know (68) and (69), without assuming that we know a particular grammar. Presumably we could arrive at (68)

(70) Sentence S seems to be grammatical in my idiolect.

(71) Sentence S is really grammatical in my idiolect.

(72) Sentence S is grammatical in standard English.

and (69) by induction from appropriate evidence statements. In the case of (68), the relevant evidence statements would be conjunctions of statements like (70) and (71); to confirm (69) we would need many conjunctions of statements like (71) and (72). But do we ever know propositions like (70), (71), and (72) without deducing them from an explicitly or implicitly known grammar?

Earlier on, we saw the Graves, Katz, et al. tentatively concede that judgments like (67) are directly evident. But (70) is a judgment like (67), so (70) must be directly evident. This means that (70) is something we know, which is, however, not inferred from any prior knowledge. Now if (70) were deduced (even tacitly) from a grammar, it would be an instance of *in*directly evident knowledge. As Stich puts it, directly evident beliefs "stand at the *end* of chains of evidence we might marshal for other beliefs" (Stich 1971, 495; emphasis Stich's). If we attempt to follow a chain of evidence back beyond the level of directly evident beliefs, we find that these beliefs arise, not from a process of inference, but from "the mechanical operations of a piece of computing machinery ... a "wheels-and-pulleys" device" (Graves, Katz, *et al*. in process), namely the human nervous system. Directly evident perceptual beliefs, if such there be, stand at the boundary of inner space (Nagel 1969b), at the boundary between a person and his nervous system.

At one point, Graves, Katz, *et al*. claim that they "know of no case where it is *plausible* to suppose that beliefs and knowledge are the end-products of "wheels-and-pulleys" operations" (Graves, Katz, *et al*. in press; emphasis in original). But their own account of language acquisition seems to contain such a case, if directly evident perceptual beliefs were not enough of a counterexample. Presumably our alleged innate knowledge of linguistic universals does not come as the conclusion of an argument, but rather as the culmination of a genetically controlled process of maturation. But if, on the rationalist account, no inference (tacit or otherwise) underlies our innate knowledge, then the rationalists agree with Stich that knowledge and belief can arise out of "the mechanical operations of a piece of computing machinery". They only disagree with Stich over the precise point at which merely mechanical operations leave off, and knowledge (or belief) begins.

Leaving aside the question of just which beliefs are directly evident, and which inferred, there are two reasons why chains of evidence must <u>have</u> ends, somewhere or other. The first reason is an ancient one. If chains of justifying arguments are circular, or if they extend back indefinitely far, then we face Pyrrhonistic skepticism. The second reason is related to a point made by Thomas Nagel. As we analyze mental processes such as inference (Nagel wrote about actions, e.g. shoe-tying), we eventually penetrate

> to the level of changes in the permeability of cell walls and in the potential gradient at nerve synapses, down beyond that to alteration in the large molecules at the nucleus of the cell, or to the subatomic events on which that depends... [There is] no inclination to ascribe tacit knowledge at the level of atomic structure, because we cannot be said to move the atoms in our bodies... (Nagel 1969b, 453-4)

Nagel's point is that long before our regress of arguments and premises threatens to become infinite, we reach a point at which it is implausible to identify physical processes with inference processes or human actions.

It seems, then, that some beliefs must "stand at the _ends_ of chains of evidence", and Graves, Katz, _et al_. have given given no cogent argument against supposing that the beliefs expressed by statements like (70) have this status. Let us suppose, then, that a native speaker knows propositions like (70), without inferring them from any other propositions (_e.g._, a grammar) which he might know.[47] What, now, about proposition (72)? On the rationalist theory, could a person know that a given sentence is well-formed in standard English, without knowing a grammar from which he deduced this fact? If not, then the rationalist account of language learning is bankrupt. For according to that account, propositions like (72) are among the child's primary linguistic data. A child corrects his early grammatical hypotheses on the basis of primary linguistic data which _conflict_ with those crude early hypotheses. Clearly, a child's knowledge of recalcitrant data cannot be inferred from the grammar he has internalized. The rationalist theory of language learning is firmly committed to the view that a child knows propositions like (72), without deducing them from his knowledge of particular grammar. And surely this is plausible enough. The knowledge that S is grammatical in English need not come from an internalized grammar; it can also come from "evidence regarding the ... behavior of other speakers, comparison with other sentences, etc." (Graves, Katz, _et al_. in press).

Finally, does a child need to know a grammar in order to know propositions like (71)? To answer this question, let us turn back for a moment to the child's knowledge of propositions like (72). Presumably the child acquires this knowledge by noting "the ... behavior of other speakers, [by] comparison [of S] with other sentences, _etc_." (Graves, Katz, _et al_. in press). Now if, by observing the behavior of _other_ speakers, and by comparing the structure of S with that of other sentences in their speech, the child can determine that S is grammatical in standard English, then he should be able to determine that S is grammatical in his own idiolect, by applying the same methods to the study of his _own_ speech behavior. Indeed, Graves and her associates concede that "there is always evidence available about whether a sound sequence is well-formed in an idiolect" (Graves, Katz, _et al_. in press).

A grammar comes into play, to be sure, when the child compares the structure of S with that of other sentences in his idiolect. But as I read Stich's account, the grammar merely directs the "wheels-and-pulleys" operations of the child's perceptual equipment. The grammar does not function as a premise in a tacit deduction. Without knowing or believing a grammar, then, a child could (even on the rationalist theory) arrive at (possibly tacit) knowledge of propositions like (70), (71), and (80). From such knowledge, the child could arrive at (tacit) knowledge of principles (68) and (69), by means of a tacit inductive inference. Given a previously unencountered sentence T, the child uses (68) and (69), plus his directly evident perceptual belief about T, to infer (tacitly) that T is grammatical in standard English. Parallel accounts can be given for properties and relations other than grammaticality. On this view, linguistic universals are not premises in a deduction of a particular grammar, which we then know. Rather, linguistic universals are constraints on the development of a perceptual system, which ultimately comes to operate in conformity with the rules of English grammar. Stich agrees that there are internal (perceptual) structures which correspond to principles of universal and particular grammar, but not that this internal structure is knowledge.

(73) The cat is on the mat is grammatical in English.

The rationalists postulate knowledge of particular and universal grammar in order to explain our knowledge of propositions like (73). We now have an alternative explanation of this knowledge. How are we to adjudicate between the account just sketched, and that of Graves, Katz, *et al*.? Not, presumably, by means of empirical tests. Stich's internal perceptual structures reflect exactly the principles reckoned as tacit knowledge by the rationalist, so Stich's (= my) account and that of Graves, Katz, *et al*. make precisely the same predictions. One might, however, argue that one account or the other includes among its *explanantia* claims which are false or implausible on philosophical grounds. Against Stich, one might argue that we have no directly evident perceptual beliefs at all, or that linguistic propositions like (70) are not among them. One might argue that perceptual competences, unlike grammars, are innate and not acquired. One might argue that only an inferential, and not a perceptual process, could be as complex as a derivation generated by a transformational grammar. One might observe that sentences are a peculiarly abstract object for perceptual or directly evident knowledge, but on the latter point at any rate, the rationalist inhabits a glass house; _he_ claims that we have directly evident (though not perceptual) knowledge about the class of all possible human grammars, i.e. that we do not acquire that knowledge by any process of inference, explicit or tacit.

None of these lines of argument, with the possible exception of the first and second, seems very promising to me. Nor do Graves and her colleagues seem inclined to defend themselves by attacking Stich's account of linguistic knowledge. Instead, they suggest that Stich's account is not incompatible with their own, and that consequently there is no need to choose between th: two. They write that judgments like (73) (or (64)?)

> certainly may be perceptual judgments. But even if they are perceptual in some appropriate sense, still there is no reason to think that no tacit deductions underlie them. Until we are given some reason to think that perceptions do not involve tacit deduction of the kind postulated here, [it is] a _non sequitur_ [to argue from the perceptual character of (73) to the claim that no tacit deduction from a tacitly known grammar underlies our knowledge of propositions like (73)]. (Graves, Katz, _et al_. in press)

It surely does Stich no justice to suppose that his explanation merely claims that beliefs like (64) are perceptual. Rather, he claims that such beliefs, besides being perceptual, are not deduced from anything else we believe or know, least of all from a grammar. This is clearly incompatible with the rationalist explanation of our knowledge about individual sentences. If

Stich's account provides an adequate explanation for the _explanandum_ knowledge, then "good sense sharpened by Occam's Razor" (Stich 1971, 486) must reject the gratuitously strong _explanantia_ of the rationalist account. At the very least, I think we may conclude that the Graves-Katz account is no better that the account I have worked out on the basis of Stich's paper.

Actually, Occam's Razor as I understand it cautions against multiplying _entities_ beyond necessity, and the rationalist theory is gratuitously strong in its epistemic, rather than its existential, commitments. But this seems a mere quibble. I can think of no reason why needlessly strong knowledge claims in an explanation should be any more acceptable than needlessly strong existence claims. The only salvation for the rationalist theory of linguistic knowledge, so far as I can see, would have to come from an argument supporting one of the following three conclusions: that principles (68) and (69) do not have the reliability needed to support conclusions like (73) (cf. note 46); that Chisholm's theory of perception does not provide an adequate account of any sort of perceptual knowledge; or that some profound (and relevant) disanalogy between color judgments and grammaticality judgments excludes the latter from the scope of Chisholm's theory, or any reasonable extension of that theory. Unfortunately, I do not know how to provide such a defense of the rationalist theory. Propositional knowledge of particular and universal grammar survived Edgley's attack, via the justification condition, and Quine's, via the truth condition. But unless some defect can be found in Stich's account of linguistic knowledge, there would appear to be no explanatory value in the claim that we believe the principles of universal or particular grammar.

Footnotes

*I wish to express my appreciation to my advisers, Sylvain Bromberger and Ned Block, for their help in eliminating some gross defects in earlier drafts of this chapter. I, not they, am responsible for all remaining errors.

1 For a general discussion of the theory of universal grammar, and of its relationship to particular grammar (the grammars of particular languages), cf. Chomsky 1965, 24-59; Chomsky 1969a, 59-65.

2 As we shall see, Edgley does not regard justification as a necessary condition on knowledge _tout court_, but only on "central" or "paradigm" cases of knowledge (Edgley 1969, 28-29). In order to argue that we know (the grammar of) English, he suggests, we must "relax...the condition requiring that the person to whom knowledge is ascribed should have reasons justifying his assurance" (Edgley 1969, 32).

3 This designation of the argument is merely a convenience. The unnamed co-authors deserve equal credit for the argument.

4 Gettier's counterexample poses no problem for (5), for the following reason. Smith validly deduced (e) from (d), but since (d) was false, Smith did not _know_ (d). Therefore (5) does not imply that Smith knew (e).

In describing (5) as "something like the converse of Hume's principle", I mean to call attention to the following fact: that in (5), the clause about consequential (derived) knowledge forms the consequent of the conditional, whereas in Hume's principle, that clause occurs in the antecedent.

5 I shall discuss Edgley's "paradigm example" below, when I attempt to state his position on the substantiation of knowledge claims. Cf. pp. 92-93 of this paper.

6 Edley's remarks about criteria for applying psychological concepts, partially quoted above on p. 90, also suggests that he regards an appeal to criteria as an essential feature of any acceptable argument for a knowledge claim.

7 Graves, Katz, et al. (in press) deny that statements like (13) formulate principles of universal grammar. Rather, they maintain that such statements formulate claims to the effect that such and such a principle is a principle of universal grammar. In the Appendix, I shall argue that the principles of universal grammar must (with one exception) conform to schema (13). Otherwise the conjunction of linguistic theory with primary linguistic data ceases to entail particular grammars, and the ascription of innate knowledge loses its explanatory value.

8 Actually, children may be able to learn "impossible" grammars, using some learning mechanism other than that involved in normal first-language learning; cf. Chomsky 1965, 56.

9 One way to evade this counterexample would be to require that _A_ be aware not only of _q_, but of all the "relevant facts" (if any) that must be added to _q_ in order to yield a conjunction

that logically supports p. This move is not likely to appeal to Edgley, however, for in the language acquisition case, it would amount to the demand that language learners be aware of the principles of universal grammar. If we say that this demand is met, why not say that the child knows these principles, and uses them to infer his grammar from his data? If we say the new requirement is not met, on the other hand, we cannot use the newly amended (9) to argue that a speaker knows the grammar of his native language.

10 Such an argument could not be brought against our claim to know particular grammar, of course, for that knowledge satisfies the antecedent of (9).

11 Of course, idiosyncrasies of the concept of knowledge will determine what constitutes a possible explanation of any given instance of knowledge. For example, it is in virtue of (5), a principle which concerns the concept of knowledge, that the rationalist account is even so much as a possible explanation of our knowledge about individual English sentences.

12 I am indebted to Ned Block for suggesting this example.

13 But cf. Nagel 1969a, 175-178, and Stich 1971, 485-486, 488-490, where it is observed that such a reconstruction of the inference would not be accompanied by the ah-ha! Erlebnis of genuine recall. This observation does not seem to add much strength to the fictionalist account, however.

14 I am not the first person to take this view of Quine's indeterminacy thesis. Cf. Chomsky 1969b; Katz 1972, 286-292.

15 Biologists make a distinction between _rabbit_ and _hare_, but I suspect that most English speakers do not.

16 Here and throughout, I assume the "standard theory" (Chomsky 1965) about the structure of particular grammars.

17 Quine is prepared to admit that some claims about synonymy (_e.g._, that _bachelor_ is synonymous with _unmarried man_, and that _alligator_ is not synonymous with _cypress_) are objectively true. Cf. pp. 157-158 below. Consequently, it might be possible to retain some sort of lexicon in particular grammars, and in linguistic theory some constraints on such lexicons, without running afoul of Quine's thesis.

18 I use _truth value_ interchangeably with "assent value", assuming that our informant always knows and tells the truth when he assents to (or dissents from) a queried sentence.

19 Sylvain Bromberger has pointed out to me that a truth-functional compound of observation sentences may be a standing sentence. (One example is: _If that is a rabbit, then that is not a pterodactyl_.) If both observation sentences and truth-functional particles can be translated without indeterminacy (cf. Quine 1960, 68), then so can standing sentences of this sort.

20 Note that Quine's rule would not screen out _all_ irrelevant stimuli, however. An English speaker might dissent from

You have a toothache, see a rabbit, and then assent to You have a toothache.

21 I assume, unrealistically, that the linguist uses only one informant.

22 I do not think that this proviso threatens the weak indeterminacy thesis with vacuity. In isolated cases, the effects of sample size may be offset by lack of variety in the sample, or by the circumstances under which the sample was collected. A sample of 100 stimuli may not tell us much about the stimulus meaning of Gavagai if the sample consists of 100 distinct views of the same furry rabbit (identifiable, say, by a collar around his neck). Or our sample may have been gathered on the testimony of an unreliable informant, who tried to please us by dissenting before, and assenting after, every stimulus we tried. But I see no reason to suppose that such factors will systematically, or normally, offset the effects of sample size.

23 An exception arises in the limiting case of eternal sentences, whose truth value never changes. All spinsters are female and No bachelor is married are intuitively heteronymous, but these sentences have identical stimulus meaning. Despite this, we can give objective grounds for the claim that these sentences are not mutual paraphrases, using a natural extension of the method of prompted assent and dissent. Cf. pp. 140-141 below. This same extension provides an objective check on analytical hypotheses, Quine's second argument to the contrary notwithstanding.

24 A careless subject might mistake (35) as an attempt at direct quotation, and dissent for that reason. But presumably we can sensitize our subjects beforehand to the difference between direct and indirect quotation, without using sentences that would tend to inculcate any new beliefs about the meaning of hare. Also, as remarked in note 15, I assume that biologists are peculiar among English speakers in distinguishing between hare and rabbit.

25 Note, incidentally, that tinkering with the meaning of "numerical identity and associated particles" (Quine 1960, 54) is not likely to make room for the hypothesis that hare and rabbit stage are really synonyms. The identity predicate does not occur in (27) and (28). We might read the indefinite article as an existential quantifier, and make the meaning of this quantifier dependent on the noun it precedes. But in making the meaning of quantifiers sensitive to context, we would make the challenge hypothesis more complex, and thus less plausible, than the standard hypothesis.

26 One reason for the failure of Kirk's argument may be that he ignored the importance of internal structure in discussions of meaning and translation. Another cause for its failure may be Kirk's reliance on the judgement of undiscerning subjects: monolingual speakers of English. If the intelligibility of (31) were judged by true bilinguals, rather than by monolinguals, the argument might carry some weight. Any sentence like (31) will be intelligible to English monolinguals, however much the indirect

quotation may differ in meaning from the directly quoted sentence. It is precisely the English speaker's ignorance that guarantees the intelligibility of such sentences. The question is whether more discriminating subjects (true bilinguals) will always in principle be available. On this point, cf. pp. 167-169 below.

27 The plausibility of Quine's argument derives, it seems to me, from Quine's concentration on sentence contexts, such as _Gavagai_, in which _gavagai_ may be translated indifferently as _rabbit_ or as _rabbit_ _stage_. We shall see presently that there are other, more discriminating contexts.

28 In describing _nematai_, and earlier _gavagailo_, as terms for rabbit stages, I assume the truth of the standard hypothesis.

29 The challenge hypothesis would thus mention only one of the terms t_1 and t_2, equating (say) t_1 with e. The supplementary system for converting Junglese sentences to their translatable normal forms would state that t_2 is synonymous with t_1. Thus many-one challenge hypotheses reflect deviant views about synonymy within the home language.

30 Remember (Quine 1960, 57-60) that truth-functional connectives are not subject to the alleged indeterminacy of radical translation.

31 Of course Junglese might not have contained the derivational suffix -_lo_, meaning _stage_. But if Junglese- English translation escapes indeterminacy due to structural peculiarities of these two languages, then Quine will have to say in general

what structural features make for indeterminacy. Then it might turn out that no human languages can have the requisite structure.

32 It is logically possible that translation schemes like Q_1 might be learnable, but objectively wrong in a given bilingual population (cf. pp. 158-161). Circumstances might conspire to give "standard" bilinguals an imposing majority, while keeping Quinean bilinguals in the minority. There is, however, no evidence that Quinean translation hypotheses are accessible to the human language-learning apparatus, so this speculation seems quite academic.

33 This is merely a formal convenience. I do not mean to imply that such a "union language" can be generated by a single Chomskian grammar.

34 If this definition is to work, there must of course be at least one such bilingual. But if the rationalist theory of language acquisition is correct, we can always invoke a population of hypothetical bilinguals; cf. pp. 167-169.

35 We may find two homogeneous regional dialects, but I shall assume hence forth that idiolect differences are randomly distributed, and bear no systematic relationship to such factors as geography.

36 Another possibility is that stage may be learned, like rabbit, by ostension. Quine apparently assumes that children would be unable to learn stage by ostension, but this is an empirical assumption about the character of the child's innate language-learning equipment.

It is incidentally strange that Quine should admit that objective synonymy relations may arise out of "learning appropriate associations of words with words" (Quine 1960, 56). A child may determine inductively that *All bachelors are unmarried men, and vice versa* is stimulus-analytic (Quine 1960, 53-54; 68). But if the meanings of *all* and *are* are really indeterminate (Quine 1960, 74), then how does the child determine, from the stimulus-analyticity of this sentence, that *bachelor* and *unmarried man* are interchangeable in all but a few contexts? And if the child can recognize a test of synonymy in sentences of the form *All F's are G's, and vice versa*, why can't the foreign linguist?

37 Perhaps, in view of these heteronymy and synonymy relations, Quine's thesis is not that such relations *are* indeterminate, but only that they *could* be. For a quotation that supports this interpretation, cf. p. 157 above. Another possible interpretation of Quine's thesis, to much the same purpose, is that determinate synonymy and heteronymy relations never arise out of ostensive learning, but only out of "associations of words with ords".

38 In one place, Quine intimates that the learner's "past knowledge of languages" (Quine 1960, 71) may influence his choice of analytical hypotheses, but he has nothing to say about the nature of the hypothesized influence.

39 Of course linguists do not yet know very many of these principles, but such principles are accessibl to empirical investigation.

40 Little is in fact known about the process of second-language learning, or the simultaneous learning of two languages. My remarks on this subject are entirely speculative, but I think they constitute one natural extrapolation from the rationalist theory of first-language learning.

41 We could in principle use monolingual subjects, and study their paraphrase behavior but it would be more difficult to protect monolingual subjects from contamination, in the form of exposure to other people's paraphrase standards.

42 If all these students were trained linguists, Quine could blame their unanimity on an artificially inculcated set of "implicit supplementary canons" (Quine 1960, 74). But we could surely eliminate all such "contaminated" students from our sample. Quine would then have no explanation for the unanimity among the remaining "naive" subjects.

43 Also, students must have sufficient exposure to acquire the native speaker's intuition that _lapin_ and _phase d'un lapin_ are heteronymous.

44 To show that a sentence S is well-formed in an idiolect I, one argues that S is structurally similar to other sentences of I, or that S has the structure of a sentence of I. Such claims of course depend on the grammar of I, but that grammar can be justified on the basis of a corpus that excludes S, but includes sentences "related to" S in various ways.

45 I owe this point to Ned Block.

46 The conjunction of (67) with the evidence for (68) and (69) might not support (64) at all, however, if the usually in (68) and (69) turns out to mean in just over 50% of all cases. Chisholm (1966, 54n.) makes this point in the following way. Suppose that there are 100 Christians in Goleta, 51 of whom are Protestants. Moreover, 26 of the Protestants are Presbyterians. Under these

(a) Usually if a Goleteño is a Christian, then he is a Protestant.

(b) Usually if a Goleteño is a Protestant, then he is a Presbyterian.

conditions, hypotheses (a) and (b) are true. These hypotheses closely resemble (68) and (69).

Consider now Jones, an arbitrary Goletan Christian. One is tempted, on the strength of (a) and (b), to infer that Jones is a Presbyterian. Relative to our total information, however, this is improbable, since only 26 of the 100 Christians in town are Presbyterians.

47 Notice, by the way, that (70) would not follow from a grammar alone, but only from a grammar supplemented by a theory of linguistic performance. A grammar would at most entail that I am able to recognize sentence S as grammatical in my idiolect (if indeed a grammar says anything at all about me). Proposition (70) states that sentence S actually does strike me as grammatical in my idiolect, on one (or perhaps on every) occasion when I

encounter S. Graves and her associates emphasize in some passages that grammars are theories about languages, not about speakers and their reactions to sentences. Cf. Graves, Katz, et al. in press.

References

Chisholm, R. (1957) Perceiving: A Philosophical Study, Cornell University Press, Ithaca.

Chisholm, R. (1966) Theory of Knowledge, Prentice-Hall, Englewood Cliffs.

Chomsky, N. (1965) Aspects of the Theory of Syntax, MIT Press, Cambridge, Mass.

Chomsky, N. (1967) "Recent Contributions to the Theory of Innate Ideas," Synthèse 17, 2-11.

Chomsky, N. (1969a) "Linguistics and Philosophy," in Hook (1969), 51-94.

Chomsky, N. (1969b) "Some Empirical Assumptions in Modern Philosophy of Language," in Morgenbesser, Suppes, and White (1969), 260-285.

Edgley, R. (1970) "Innate Ideas," in Vesey (1970), 1-33.

Fodor, J. and Katz, J. (1963) "The Availability of What We Say," Philosophical Review 72, 57-71.

Gettier, E. (1963) "Is Justified True Belief Knowledge?" Analysis 23, 121-123.

Graves, C., J. Katz, Y. Nishiyama, S. Soames, R. Stecker, and P. Tovey (in press) "Tacit Knowledge," Journal of Philosophy.

Hempel, C. (1952) Fundamentals of Concept Formation in Empirical Science, University of Chicago Press, Chicago.

Hempel, C. (1966) Philosophy of Natural Science, Prentice-Hall, Englewood Cliffs.

Hook, S., ed. (1969) Language and Philosophy: A Symposium, New York University Press, New York.

Katz, J. (1962) The Problem of Induction and Its Solution, University of Chicago Press, Chicago.

Katz, J. (1966) The Philosophy of Language, Harper and Row, New York.

Katz, J. (1972) Semantic Theory, Harper and Row, New York.

Kirk, R. (1969) "Translation and Indeterminacy," Mind 78, 321-341.

Leibniz, G. (1704) New Essays on Human Understanding, trans. A. G. Langley (1949), Open Court, LaSalle, Ill.

Lewis, D. (1969) Convention: A Philosophical Study, Harvard University Press, Cambridge, Mass.

Morgenbesser, S., P. Suppes, and M. White, eds. (1969) Philosophy, Science, and Method: Essays in Honor of Ernest Nagel, St. Martin's Press, New York,

Nagel, T. (1969a) "Linguistics and Epistemology," in Hook (1969), 171-182.

Nagel, T. (1969b) "The Boundaries of Inner Space," Journal of Philosophy 66, 452-458.

Quine, W. (1960) Word and Object, MIT Press, Cambridge, Mass.

Quine, W. (1968-69) "Reply to Harman," Synthèse 19, 267-269.

Quine, W. (1969) Ontological Relativity and Other Essays, Columbia University Press, New York.

Smart, J. (1963) Philosophy and Scientific Realism, Routledge and Kegan Paul, London.

Stich, S. (1971) "What Every Speaker Knows," Philosophical Review 80, 476-496.

Strawson, P. (1952) An Introduction to Logical Theory, John Wiley and Sons, New York.

Vesey, G., ed. (1970) Knowledge and Necessity, Royal Institute of Philosophy Lectures, v. 3 (1968-69), St. Martin's Press, New York.

Chapter 3

Linguistics and Language Learning*

Noam Chomsky has argued that when we "investigate the conclusions that can be established concerning the nature of language, the ways in which language is used and understood, the basis for its acquisition", it turns out that "these conclusions have interesting consequences for psychological theory, in particular, that they strongly support an account of mental processes that is, in part, familiar, from rationalist speculation about these matters" (Chomsky 1969a, 65). Many writers have argued, contrary to Chomsky, that the findings of transformational linguistics do <u>not</u> support a rationalist approach to the psychology of language learning, and that those findings do not refute empiricist proposals in this area. In this paper, I shall examine three such arguments against Chomsky's thesis.

Before describing those arguments, however, I wish to indicate the nature of the disagreement between rationalist and empiricist theories of learning, and to emphasize my intention of concentrating on psychological rather than epistemological issues. On the first matter, Gilbert Harman writes:

> The "empiricist" believes that the linguistic-acquisition device can arrive at correct grammatical rules by inductive inference from the primary linguistic data. The "rationalist" denies this and holds that the linguistic-acquisition device requires further "information" in addition to the primary linguistic data. Such further information consists of what Chomsky calls "innate ideas and principles." ... (Harman 1967, 84)

This is not a complete description of the rationalist-empiricist dispute. I shall fill in some further details below (pp. 234-236). For present purposes, however, Harman's thumbnail sketch will suffice.

Many of Chomsky's critics object to the claim that children have innate *knowledge* of principles relevant to language learning, or to the claim that adults tacitly *know* grammatical principles relevant to language use. Even if these critics are right, even if we do not *know* principles of universal or particular grammar, these principles may still play an important role in describing the psychology of language acquisition and use. In a forthcoming paper, Graves, Katz, *et al*. emphasize that we must "distinguish between the claim that speakers know the grammatical rules of their language and the claim that these rules are internally represented. ... not all internalized structure need be counted as knowledge" (Graves, Katz, *et al*. in press). I shall in large part avoid the question whether language learners *know* these principles, since I have dealt with that question elsewhere (cf. Chapter 2 of this thesis). In this paper, I shall be concerned only with the weaker rationalist claim that universal grammar tells us something important about the child's innate language-learning equipment. The theory of universal grammar defines a class of grammars; the rationalist claims that a child's language-learning apparatus can help him to internalize any grammar in this class, but none outside it. The theory of universal grammar also defines a plausibility ordering on this class of possible grammars; the rationalist claims that this plausibility metric determines which of the possible grammars a child will internalize, on being exposed to a given corpus of language data.

First, we shall consider an argument by Sidney Morgenbesser (1969), which (if it were successful) would show that the rationalist approach to language learning is vitiated by an infinite regress. Secondly, we shall examine Gilbert Harman's attempt (1967) to show that there is no empirical difference between rationalist and empiricist theories of language learning. Finally, we shall look at an enriched theory of inductive reasoning, which in the view of L. Jonathan Cohen (1970b) improves the credibility of empiricism, and invalidates anti-empiricist arguments directed against more familiar (but weaker) theories of induction.

I. MORGENBESSER: DOES RATIONALISM GENERATE AN INFINITE REGRESS?

Morgenbesser puts forward his infinite-regress argument in response to Fodor's defense of "The Appeal to Tacit Knowledge in Psychological Explanation" (Fodor 1968). Morgenbesser is not concerned specifically with the rationalist account of language learning, but more generally with accounts of mental competences that postulate rule-governed structures within the nervous system. Now rationalist theories of language acquisition claim that the operation of our innate language-learning equipment is governed by principles of universal grammar. Such theories, then, are instances of the explanatory paradigm to which Morgenbesser objects.

In Fodor's paper, the discussion focusses on our acquired shoe-tying competence, rather than on our innate language-learning competence. Fodor suggests the following account of the process which underlies the act of tying one's shoes:

> There is a little man who lives in one's head. The little man keeps a library. When one acts upon the intention to tie one's shoes, the little man fetches down a volume entitled Tying One's Shoes. The volume says such things as: "Take the left free end of the shoelace in the left hand..."
>
> When the little man reads the instruction 'take the left free end of the shoelace in the left hand', he pushes a button on a control panel. The button is marked 'take the left free end of a shoelace in the left hand'. When depressed, it activates a series of wheels, cogs, levers, and hydraulic mechanisms. As a causal consequence of the functioning of these mechanisms, one's left hand comes to seize the appropriate end of the shoelace. Similarly, mutatis mutandis, for the rest of the instructions. (Fodor 1968, 627)

According to Morgenbesser, Fodor commits himself (not necessarily in the passage just quoted) to the following principle:

> (E) If any entity performs an act that manifests competence, then either that entity knows (actually or latently) relevant rules and is guided by them, or one of its sub-systems is in state S. (Morgenbesser 1969, 462)[1]

But principle E, coupled with Fodor's account of shoe-tying, lays on Fodor a responsibility which his paper fails to discharge, in Morgenbesser's view:

> ... it is, I think, up to Fodor to show that, when his little man knows, it is not the case that he knows that and knows how. ... without such showing he may not have escaped Ryle's infinite-regress hold...
>
> We attempt to describe the little man at work, and find it appropriate to say not only that he knows the rules but also that he knows how to apply them; for otherwise how could he successfully guide our actions? And, once we notice that he is exhibiting competence when he applies the rules, we are, given our commitment to Fodor's

> principle E, required to have the little man know rules for the application of rules and so into the night. (Morgenbesser 1969, 464-465)

In order to understand the import of Morgenbesser's argument, it will be helpful to see what Fodor means by his picturesque talk about the little man in one's hand. He writes that "the little man stands as a representative pro tem for psychological faculties which mediate the integration of shoe-tying behvaior by applying information about how shoes are tied... Assigning psychological functions to little men makes explicit our inability to provide an account of the mechanisms that mediate those functions" (Fodor 1968, 629). The little man, then, is the nervous system, or a subsystem whose function it is to execute the instructions in a single book (or library). These functional systems cannot as yet be identified anatomically, nor their operation described physiologically. The instructions in his books are information that is in some way stored in the nervous system. Now clearly the human nervous system has only finite storage capacity. But if Morgenbesser's argument is corect, then rationalists (or at least rationalists like Fodor) are committed to the absurd view that there is stored within the nervous system an infinite amount of information essential to the language-learning (shoe-tying, etc.) task. If this is really a consequence of the rationalist approach to language acquisition, then surely that approach does not deserve serious consideration.

How can Fodor (and other rationalists) escape Morgenbesser's "infinite-regress hold"? Morgenbesser himself suggests a way out: by showing that "when [the] little man knows, it is not the case that he [both] knows that and knows how". And as it happens, Fodor's article provides the basis for an argument to the effect that, when the little man applies rules, he does <u>not</u> exhibit "competence or know-how" (Morgenbesser 1969, 466) in the relevant sense.

At one point, Fodor contends that "although an organism can know how to <u>X</u> without knowing the answer to the question "How does one <u>X</u>?" it cannot know how to <u>X</u> unless there <u>is</u> and answer to the question "How does one <u>X</u>? (Fodor 1968, 637) According to Fodor, the act of <u>X</u>-ing exhibits competence (and thus satisfies the antecedent of principle E) only if there is an answer to the question, "How does one <u>X</u>?

Elsewhere, Fodor suggests that for some operations O of the nervous system, there is no answer to the question, "How does the nervous system carry out operation O?"

> Assume that there exists a class of <u>elementary</u> instructions which the nervous system is specifically wired to execute. Each elementary instruction specifies an <u>elementary operation</u>, and an elementary operation is one which the normal nervous system can perform but of which it cannot perform a proper part. Intuitively speaking, the elementary operations are those which have no theoretically relevant internal structure. Now to say an operation is elementary is to say that certain kinds of "how"-questions cannot arise about it. In particular, we cannot ask for instructions for performing it by performing some further sequences of operations. (It makes sense to ask how to spell 'add' but not to ask how to spell 'n'.)...
>
> The nervous system carries out its complex operations in some way or other (i.e., by performing one or another sequence of elementary operations). But the nervous system performs elementary operations in no <u>way</u> at all: it simply performs them... we [constrain] a completed psychological theory to be written in sequences of elementary instructions... (Fodor 1968, 629)

It is not the case, then, that the little man _knows how_ to perform elementary operations, or that he exhibits competence when he carries out elementary instructions. But Fodor proposes to restrict the vocabulary of the little man's library strictly to elementary instructions. This means that the little man needs only one set of how-to books. Contrary to Morgenbesser's argument, he does not need a second set of books containing meta-rules, a third set containing meta-meta-rules, and so on.

It may be objected that I have only succeeded in misinterpreting the term _competence_, and the principle E in which it occurs, because I have interpreted the term by appeal to Fodor rather than by appeal to Morgenbesser. Perhaps elementary operations _do_ exhibit competence, in _Morgenbesser's_ sense of the word. If so, then I have not really shown that elementary operations lie outside the scope of Morgenbesser's principle E. But no matter. For then E entails that further rule-knowledge underlies the little man's execution even of elementary operations. Clearly, Fodor would reject principle E, on this interpretation. Elementary operations are _defined_ as exceptions to (this reading of) principle E.

Of course, it is not a mere definition that gets Fodor off the hook. It is the claim that there _are_ some operations of the nervous system that are elementary, in the defined sense, and that "every operation of the nervous system is identical with some sequence of elementary operations" (Fodor 1968, 629).

Thomas Nagel (1969b) has challenged this claim. Let us see whether Nagel can restore Morgenbesser's infinite regress, by forcing the rationalist to give up his belief in elementary operations.

Nagel suggests various ways in which "operations that are not elementary for an organism" may fail to "consist of sequences or groups of operations that are elementary for that organism" (Nagel 1969b, 454). Of these various objections, Nagel attaches greatest importance to the following possibility:

> ... the way in which ... elementary operations are brought about as part of [a] larger action may have nothing to do with the way they are brought about when they occur in isolation. (Nagel 1969b, 454)

To lend plausibility to this description of neural functioning, Nagel introduces an example. The _abductor pollicis brevis_ muscle in the ball of the thumb plays a role in such actions as depressing the space bar on a typewriter. This muscle contains three hundred motor units, each consisting of a single motoneuron and a number of individual muscle fibers, all innervated by that particular motoneuron. Nagel cites an experiment in which human subjects were trained to "activate selected individual motor units in the abductor pollicis brevis at will... They were able to activate adjacent motor units separately, and even to beat out rapid and elaborate drum rhythms... When asked how they did this, the subjects could not say, just as we cannot say how we manage to move our thumbs" (Nagel 1969b, 455).

Since we _can_ (after suitable training) voluntarily activate individual motor units, presumably such performances count as elementary operations for the human organism. (The anatomy of the situation makes it unlikely that this operation will have any "theoretically relevant internal structure", since it results in the firing of only a single motoneuron.) It seems likely, moreover, that "when I depress the space bar, all or most of the motor units are activated at once" (Nagel 1969b, 455). But it does not follow that when I depress the space bar, I do so by performing the action of _activating_ all the motor units at once.

> I may be unable even to attempt that feat; or, if I do attempt it, nothing may happen, or I may produce a dreadful cramp in my thumb. Of course it is conceivable that I should succeed; but what reason is there to believe that the method by which the ordinary thumb movement is produced bears any interesting relation to the method by which it might be produced through the summation of three hundred simultaneous intentional motor-unit innervations? ... It is not _necessarily_ false that we tie our shoes or play musical instruments or speak by the composition, conscious or unconscious, of distinct individual finger and tongue motions (not to mention individual motor-unit innervations). It is just false. An answer to the question "What happens when one says 'Beethoven'?" is not ipso facto an answer to the question "How does one say 'Beethoven'?" (Nagel 1969b, 456)

Fodor's paper is about our tacit _knowledge_. I have no access to rules that I know tacitly, but the little man in my head does have access to such rules, via his library. The most straight-forward answer to Nagel's objection, I think, is to point out that Fodor is interested not only in tacit knowledge, but also in tacit _actions_ -- actions which my nervous system performs when I perform some competent, voluntary act.[2] I may never be able to achieve voluntary control over these actions, just as I may never be able to recognize the rules I know tacitly. It is the little man in my head that actually performs my tacit actions for me. Fodor's claim is not in the ordinary sense a claim about what I do when I say Beethoven, or tie my shoes. Rather, it is a claim about what the little man in my head does on these occasions. Nagel might prefer to describe this as a claim about "what happens [in one's nervous system] when one says 'Beethoven'". But however we describe Fodor's claim, Nagel's objection appears to rest on a misinterpretation of that claim.

No doubt Nagel would object to the concept of tacit action, just as he objects (Nagel 1969a) to the concept of tacit knowledge. The basis for that objection is Nagel's claim that if tacit knowledge were really knowledge, it would at least be "capable of reaching consciousness upon adequate reflection" (Nagel 1969b, 456). But Nagel's argument against knowledge seems no more compelling than an analogous argument against electrons: "Electrons are in principle incapable of being seen. Therefore electrons are not real."[3]

Unfortunately, our appeal to the concept of tacit performance or action does not altogether dispose of Nagel's objection. We have agreed that when I depress the space bar, I do not activate each of the relevant motor units individually. But,Nagel may go on to insist, no little man does this for me either. And Nagel may be right.[4]

To see what is at stake, we shall have to complicate our picture of the nervous system, regarding it not as a single little man, but as a whole heirarchy of little men. The complication would have been needed anyway, however:

> ... there can be no single button marked 'take the left free end of a shoelace in the left hand'... grasping a shoelace should be considered complex behavior, because doing so involves the production of motions that also play a role in other actions.
>
> We might thus consider expanding the population in one's head to include subordinate little men who superintend the execution of the "elementary" behaviors involved in complex sequences like grasping a shoelace. When the little man reads 'take the left free end of the shoelace in the left hand', we imagine him ringing up the shop foreman in charge of grasping shoelaces... the shop foreman might be imagined to superintend a detail of wage slaves, whose functions include: searching inputs for traces of shoelaces, flexing and contracting fingers on the left hand, etc. (Fodor 1968, 628)

In terms of this picture, the motoneurons (or perhaps the motor units) may be regarded as "wage slaves". They are under the "supervision" of a "shop foreman" -- a subsystem of the nervous system. Now this shop foreman can ring up individual wage slaves; Nagel cites experimental evidence for this. But he may also be able to call for the activation of all three hundred motor units at once, without calling individually on each of these "wage slaves". He may have a sort of limited "public address system", for example, as in Figure 1.[5] For the little man on the left in this diagram, there may be a physiological difference between simultaneous performance of elementary operations A, B, D, and E on the one hand, and

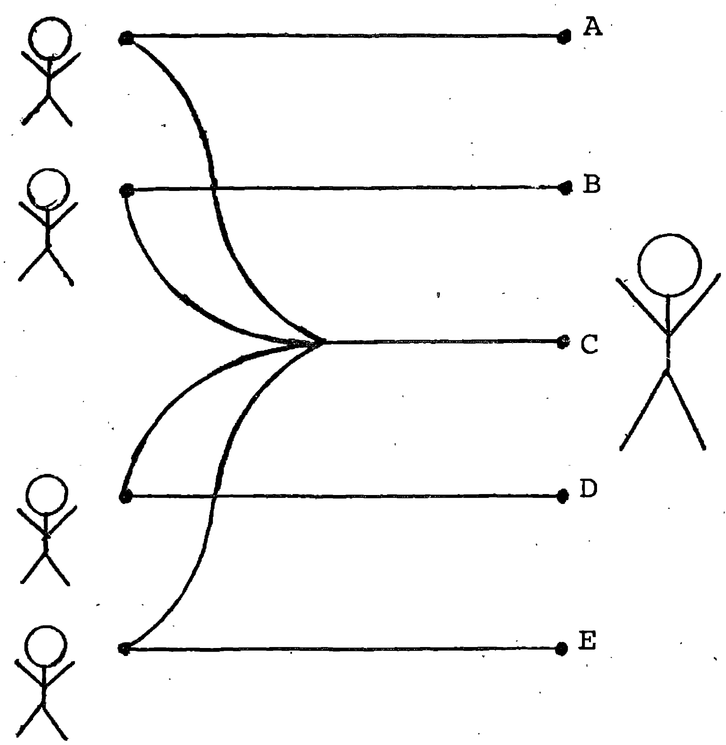

Figure 1

performance of operation C on the other. Still, what _happens_ among his subordinates (the way _they_ perform) may be the same in both cases. As I depress the space bar, then, the foreman in charge of my _abductor_ _pollicis_ _brevis_ may be doing something quite unlike performing three hundred elementary operations. Instead, he may be performing something like operation C in Figure 1.

Fodor (in conversation) has pointed out a way around this version of Nagel's objection. We need only relativize the concept of _elementary_ _operation_, so that an operation is elementary only relative to a specific task.[7] I assume, with Fodor, that in answer to the question, "How does one *X*?", it is appropriate (and possible) for the psychologist to describe the sequence of operations which the nervous system performs when one *X*'s. For example, when we make visual depth estimates, our nervous system monitors texture gradients, processes information about texture gradients, and computes the first derivatives of texture gradients (Fodor 1968, 631-632). Now suppose that when I *X*, my nervous system performs operation C in Figure 1. Thereupon, various things happen. One of these happenings is just like the event that would have occurred if my nervous system had performed operation A.

Now my nervous system can perform A in isolation, so A is an elementary operation in the original, absolute sense. But when I *X*, my nervous system performs C, not A. (Presumably this claim will eventually be subject to physiological test.) Consequently, operation A is not an elementary part of the task of *X*ing; A is not elementary *relative to* this task. Of course, A is still an elementary operation in the old sense, because it may be elementary with respect to other tasks. For example, operation A will be elementary relative to the task of *Y*ing if, whenever I *Y*, my nervous system performs (say) operations A, B, D, and E. Similarly, operation C is an elementary operation, absolutely and relative to the task of *X*ing, but not relative to the task of *Y*ing.

The point of this relativization move was to forestall a refutation of rationalism that appealed to physiological evidence. It is still possible, of course, for specific hypotheses about internal books to succumb to physiological objections. Now prior to this relativization, our foreman's vocabulary included only the following elementary instructions: A!, B!, D!, E!. We must now add C! to his vocabulary. But his vocabulary remains finite, and as long as he has a merely finite vocabulary, there will be no absurdity in supposing that the foreman "just knows" how to execute his elementary instructions, without knowing any rules for interpreting or executing those instructions (cf. the argument which begins in the next paragraph). And clearly, the concept of an elementary operation can be a relative concept without having an infinite domain. Fodor's relativization move thus allows him to escape Nagel's objection, without reinstating Morgenbesser's infinite regress. To be sure, it may not be possible to analyze all the behaviors of an organism into sequences of elementary operations performed by a single little man. But we have seen no reason to doubt the theoretical adequacy of a *finite* heirarchy of little men, each having a *finite* repertoire of operations that are elementary for *him.*

Morgenbesser's infinite-regress argument was intended as an objection to rationalist (or intellectualist) explanations in psychology. As we have seen, rationalist psychology is not the easy mark Morgenbesser took it to be. We shall come presently to an argument by Gilbert Harman, in which he contends that rationalist psychology is viable all right, but empirically indistinguishable from empiricist psychology. Before we take up that argument, however, it may be of interest to see whether Morgenbesser's argument can be made to work against the rationalist's _epistemic_ claims. Against those claims, Harman gives the following infinite-regress argument:

> Taken literally, Chomsky would be saying that we are to explain how it is that Smith knows how to speak and understand a language by citing his knowledge of another more basic language in which he has (unconsciously) "internally represented" the rules of the first language. (It does not seem to make sense to assume that Smith can represent rules without representing them is some language.) The main problem with such a literal interpretation of these remarks would be the implausibility of the resulting view. How, for example, would Smith understand the more basic language? In order to avoid infinite regress or a vicious circle, one would have to suppose that Smith can understand at least one language directly, without unconsciously knowing the rules for that language. But if this is admitted, there is no reason why Smith cannot know directly the language he speaks. (Harman 1967, 76)

Now the rationalist agrees that we know the language in which we represent the rules of English grammar. Indeed, he agrees that we know the _rules_ of this metalanguage. We know substantive linguistic universals, which specify the vocabulary used in formulating particular grammatical rules (_NP_, _VP_, + _vocalic_, (_Human_), etc.), and we know formal universals, which specify a (presumably infinite) class of "sentences" (rules) that may be constructed out of this vocabulary (cf. Katz 1972, 30-31). But the rationalist need not admit that we know the rules of the language in which linguistic universals are formulated.

To see why this is so, we must recall the rationalist argument for saying we know the rules of English. This argument begins by pointing out that when presented with a putative sentence of English, we know whether or not that sentence is grammatical in English. Now this knowledge might be stored piecemeal in our nervous system, in the form of "dossiers" on the individual sentences we recognize as grammatical or otherwise, or it might be derived from a stored set of recursive rules that generate the class of all English sentences. The rationalist rejects the former account, because on that account, we would have to postulate infinite memory capacity in order to explain human competence. English speakers can recognize grammaticality in any of an infinite class of sentences. A similar argument can be used to show that we must know the rules of the language in which we formulate the rules of our (English) grammar (cf. pp. 238-239 below).

The class of grammatical English sentences is infinite, and an English speaker is competent to recognize any sentence in this class. The set of possible grammars (sentences in the language of particular grammar) is infinite, and a normal child is competent to internalize any grammar in this set, on being exposed to appropriate data. Now perhaps the language of universal grammar is also an infinite language. Assuming this to be the case, the relevant question is: are all sentences in this language potential objects for some human competence, analogous to our competence at sentence recognition and grammar acquisition? I do not know of any such competence, and Harman does not attempt to establish the existence of any such competence. Unless and until such a competence is demonstrated, the rationalist theory can do all its explanatory work on the assumption that human competence in the language of universal grammar extends only to that finite set of sentences which constitutes the theory of universal grammar. Such an assumption implies no claim of further rule knowledge, and Harman's "infinite" regress of rule-sets terminates after two steps. The infinite-regress argument is no more of a threat to epistemic rationalism, than it is to psychological rationalism.

Morgenbesser's infinite-regress argument was intended to show that rationalist theories of language use and acquisition are incompatible with the finite character of the nervous sy-tem's storage capacity. That argument failed, however, because it ignored the concept of an elementary operation. We turn now to the argument by which Gilbert Harman attempts to show, not that rationalism is incompatible with some obvious fact, but rather that rationalism makes no claims that are incompatible with empiricism. If this is correct, then transformational grammarians cannot have found evidence that (as they claim) both refutes empiricism and supports rationalism.

II. HARMAN: ARE RATIONALISM AND EMPIRICISM COMPATIBLE?

At the beginning of this paper, I quoted Harman's characterization of an empiricist as someone who "believes that the linguistic-acquisition device can arrive at correct grammatical rules by inductive inference from the primary linguistic data" (Harman 1967, 84). In the same passage, he goes on to point out that the viability of empiricism, so defined,

> depends upon what is to count as inductive inference. On some interpretation[s] of "inductive inference," empiricism can be ruled out directly, without appeal to transformational linguistics. On other interpretations, empiricism would not be incompatible with transformational linguistics. However, there does not seem to be any reasonable interpretation on which empiricism survives direct and relatively a priori refutation but is refuted by the appeal to transformational linguistics. (Harman 1967, 84)

Harman complains that Chomsky, in his critique of empiricism, concentrates on a version of empiricism that "can be ruled out directly, without appeal to transformational linguistics". Specifically, Chomsky attends only to the brand of empiricism espoused by taxonomic linguists. But taxonomic linguists "represent an anti-theoretical position" among empiricists (Harman 1967, 86). Unlike Quine, for example, the taxonomic linguist does not, in Harman's view, see the need for a learning device that can "infer the truth of theories as well as generalizations" (Harman 1967, 85).

Harman also points out that there is another sort of empiricist theory that would be subject to "direct and relatively a priori refutation". Appealing to Goodman's "New Riddle of Induction" (Goodman 1965, 59-83), Harman notes that "any <u>consistent</u> set of inductive principles must favor certain [kinds of] generalizations over others" (Harman 1967, 85-86; emphasis added). Harman clarifies this remark in a later publication, writing that "if we are to learn from experience at all, then our initial assignment of probabilities must from a certain point of view arbitrarily favor certain hypotheses over others" (Harman 1969, 150). Consider, then, an empiricist proposal that fails to specify any initial ranking of hypotheses, with respect to plausibility. Such a proposal describes a device that would be unable to "learn from experience at all". Hence such a proposal can be refuted in a "direct and relatively a priori" way.

Finally, Harman sketches an (allegedly) empiricist theory that avoids the defects of taxonomic linguistics, and those of inductive procedures without initial biases. Harman suggests that a "resourceful empiricist ... no matter what the facts of language turn out to be" (Harman 1967, 87), will be able to devise an initial plausibility ranking that accounts for those facts without postulating any innate principles. He writes:

> ... one cannot support rationalism by showing that only languages with certain types of grammar (e.g., transformational grammar) are learnable, since an empiricist could reply that this shows only that the principles of induction used (which must be biased in favor of some hypotheses) are biased in favor of grammars of the designated types. (Harman 1967, 86)

Later on (pp. 230-242) I shall undertake a closer examination of resourceful empiricism, and Harman's claims about it. For now, I wish to return to Harman's complaint that taxonomic linguists hold a radically "anti-theoretical position" which makes them inadequate representatives of empiricism. I shall argue, first, that Harman is correct in suggesting that a device can "be allowed to infer the truth of theories" of the relevant sort (transformational grammars) without deserting empiricist assumptions in favor of rationalist ones. However, I shall argue (secondly) that taxonomic linguists are not more "anti-theoretical" than other empiricists, and that taxonomic linguists constrain their grammatical theories in ways that are quite typical of classical empiricism. Finally, I shall argue that such empiricist constraints make an empiricist learning device incapable, not of weighing the appropriate grammar against rivals and inferring its truth, but incapable of formulating such a grammar in the first place. A learning device can only infer the truth of a theory from relevant evidence if it has previously formulated that theory and decided to consider it as a serious candidate for adoption. An empiricist learning device, however, could not carry out this first step in the language-learning process, because no empiricist heuristic could formulate or propose a transformational grammar for testing. Empiricist learning devices have adequate resources for testing and inference; their failings lie rather in their heruistics.

On the first point, there is an argument by Katz (1966, 256-258) which is meant to show that no empiricist learning device could infer the truth of the theory (grammar) of English, but I believe that Katz is mistaken. His argument concerns unobservable _you_ subjects in the deep structure of normal English imperative sentences, such as (1) below. An empiricist might

(1) Help the man.

(2) You help the man.

suggest that we interpret sentence (1) by correlating it with sentence (2), a synonymous sentence in which the surface structure does contain the _you_ subject.[7] But Katz points out that _you_ is not the only NP that may occur as the surface subject of

(3) John Jones help the man.

(4) Everyone help the man.

an imperative. Besides sentence (2), we also have sentences (3) and (4) to reckon with. Katz maintains that an empiricist learning device would have no basis for correlating sentence (1) with sentence (2), and for rejecting the alternative hypothesis that (1) is correlated with (3), or with (4).

It seems to me, however, that an empiricist does have resources for eliminating these two incorrect hypotheses. First consider the hypothesis that (2) is always correlated (and synonymous) with (3). Surely it is possible that a child will observe utterances of (3) in circumstances that preclude this correlation. Suppose, for example, that on some occasion when (3) is uttered, there are only three people nearby: the child, the speaker, and the addressee. (The man in need of help is perhaps drowning some distance from shore.) The addressee is a lifeguard named Rebecca, and the child can tell that she (and not he) is the addressee, because the speaker is looking at her, not him, when he utters (1). If none of those present is named

John Jones, and if the child knows that the speaker knows this, the child has very strong grounds for rejecting the hypothesis that (1) is correlated with (3).

An alternative hypothesis, consistent with evidence of this sort, might be that (1) is <u>normally</u> correlated with (3), but that under circumstances precluding this interpretation, (1) is correlated with (2). But this hypothesis is clearly more complex than the hypothesis that (1) is <u>always</u> correlated with (2). An empiricist would thus prefer the latter hypothesis because of its greater simplicity. There is nothing peculiarly rationalistic about such an appeal to simplicity, so long as we invoke a <u>general</u> concept of simplicity, and not one that is applicable only to grammatical hypotheses (cf. pp. 234-235 below).

What, though, about correlating (1) always with (4)? Here I wish to illustrate a somewhat more complex form of reasoning that an empiricist might use. An empiricist learning device could, I think, rule out the correlation of (1) with (4), even if it never encountered sentence (1) itself in circumstances that precluded this correlation.[8] The class of subjectless imperatives, which includes (1), can I think be defined in terms of its surface-structure characteristics. To determine that sentence (1) cannot be correlated with (4), it is enough to

(5) Hide behind the easy chair.

hear <u>some</u> subjectless imperative (say, sentence (5)) under conditions which show that it cannot be correlated with an imperative having <u>everyone</u> as its surface subject.

(6) Everyone hide behind the easy chair.

Now suppose that two dozen people are in a room preparing a surprise party. Under these conditions, someone might utter (5), but hardly (6), since one easy chair could not provide concealment for twenty-four people.[9] Here, then, is a case where it is inappropriate to supply <u>everyone</u> as the subject for a subjectless imperative. But this refutes the correlation of (1) with (4), if we assume that there must be some uniform, <u>simple</u> way of interpreting all subjectless imperatives -- surely a reasonable assumption, on empiricist as well as rationalist principles. It seems reasonable to suppose that the correlation of (1) with (2) will be the only simple correlation to escape direct and indirect refutation.

Actually, of course, the empiricist argument I have been sketching is meant to support a general rule for correlating a subject-bearing imperative with each subjectless one, rather than merely to support a single such correlation. But even this rule for correlating pairs of sentences is not the same as the imperative rule in the transformational grammar of English. I have been discussing a rule that expresses a transformational relationship, in Harris' sense, between surface structures of two sorts (cf. Harris 1957). The imperative rule relates the surface structure of (1), not to another surface structure (e.g. that of (2)), but to a deep structure (which happens to underlie both (1) and (2)). Nonetheless, it seems to me that the considerations I have adduced in support of the Harris-style transformation rule, can also be used to support the standard (Chomskian) imperative rule.

Katz tried to argue that an empiricist device could not determine the correct deep structures for subjectless imperatives, nor the transformation which maps these deep structures into surface structures. This argument has failed, because Katz assumed that the correct hypothesis (or something very much like it) was among the hypotheses to be considered by the empiricist device. Once we have the correct rule (or grammar) in our field of candidates, the learning device only needs to weed out all the false hypotheses. It may do this on considerations of general simplicity, or it may reject grammars which turn out to have false empirical consequences, e.g. grammars which generate non-sentences, or fail to generate certain grammatical sentences. So far as I can see, no rationalist assumptions are needed in this weeding-out process, though rationalist assumptions may be very helpful.[10]

We come now to Harman's charge that taxonomic linguists occupy a radically "anti-theoretical position in linguistics, a position that would sharply distinguish inductive generalization from theoretical "speculation" so as to do without the latter" (Harman 1967, 86). Elsewhere, Harman indicates that the anti-theorist "permits only rather weak generalizations about unobserved speech" (Harman 1967, 84). His most specific remark is that

> Weak principles of generalization would not even permit the inference of a corpus of sentences from a record of natural speech. Given the "data" possessed by the language-learning device, the notion of sentence is a theoretical notion. (Harman 1967, 85)

Now it is just false that taxonomic linguists regard talk about sentences as methodologically treife. Harris, for example, thought he could "infer a corpus of sentences from a record of natural speech". He writes:

> The stock of recorded utterances constitutes the corpus of data ... An UTTERANCE is any stretch of talk, by one person, before and after which there is silence on the part of the person. (Harris 1963, 12, 14)

Harris goes on to outline a series of segmentation and classification operations, which he claims will eventually allow him to distinguish sentences from non-sentences in the language or corpus under investigation; cf. Harris 1963, 158n, 350 ff, 378.

Nor do taxonomic linguists confine themselves to "weak generalizations about unobserved speech", in any plausible sense. Gleason, for example, maintains that "there are several levels of structure in a language" (Gleason 1961, 65), and Longacre writes that "language is structured in three semi-autonomous modes, phonology, grammar, and lexicon" (Longacre 1964, 7). Gleason's remarks about strata, and Longacre's about modes, are put forward as empirically defensible hypotheses about the structure of all human languages. Such hypotheses are hardly "weak generalizations about unobserved speech".

Harman is thus mistaken when he describes taxonomic linguists as "anti-theoretical". Nonetheless, I think we can identify the feature of taxonomic grammars that led Harman to regard them as non-theories. Such grammars use many of the same theoretical terms as transformational grammars: <u>Noun</u>, <u>Verb</u>, <u>Sentence</u>, etc. But a taxonomic grammar applies these terms only to continuous stretches of observable acoustic signals, whereas a transformational grammar also uses these terms to refer to unobservable constituents, e.g. the deep-structure subjects of English imperatives. A taxonomic linguist insists that the theoretical concepts in grammars must be derived from experience with segments of observable speech events, and that theoretical terms must be applied only to such audible constituents. One might say somewhat paradoxically that taxonomic linguists posit theoretical entities, but only observable ones.

Taxonomic linguists, then, require a rather intimate connection between sense experience and grammatical concepts. But in this respect, they resemble other empiricists, classical and contemporary. Thus Locke writes:

> ... <u>even the most abstruse ideas</u>, how remote soever they may seem from sense, or from any operations of our own minds, are yet only such as the understanding frames to itself, by repeating and joining together ideas that it had either from objects of sense, or from its own operations about them... (Locke 1690, 217; emphasis Locke's)

Hume expresses the same sort of view:

> Nothing, at first view, may seem more unbounded than the thought of man... To form monsters, and join incongruous shapes and appearances, costs the imagination no more trouble than to conceive the most natural and familiar objects...But...our thought... is really confined within very narrow limits...all this creative power of the mind amounts to no more than the faculty of compounding, transposing, augmenting, or diminishing the materials afforded us by the sense and experience... all the materials of thinking are derived either from our outward or inward sentiment... all our ideas or more feeble perceptions are copies of our impressions or more lively ones. (Hume 1748, 28-29)

Similarly, twentieth-century learning theorists in the empiricist tradition assume that the peripheral physiology of an organism is of primary importance in constraining the class of concepts (and thus the class of hypotheses) accessible to the organism (Bever, Fodor, and Garrett, forthcoming, Chapter 8).

We have seen that taxonomic linguists place certain restrictions on the sorts of theoretical entities a grammar may refer to. But these restrictions do not reflect a peculiarly anti-theoretical orientation. Rather, they are quite typical of the empiricist tradition in philosophy and psychology. But now we must ask whether a language-learning device, working within these constraints, could ever succeed in discovering the transformational grammar of English. Quite clearly, the answer is that such a device could not be successful. The device might consider the hypothesis that every sentence consists of a noun phrase followed by a verb phrase, but it would irrevocably reject that hypothesis upon encountering "subjectless" imperatives such as (1) and (5) above.

The device would never attempt to save this hypothesis by "resort to auxiliary hypotheses and forced interpretations of various kinds", e.g. the supposition that the subject of an imperative is "left unexpressed" (Bloomfield 1961, 13). An empiricist might possibly claim that the child can form the complex concept of subject deletion, because of his experience with subjects of declaratives, and his experience of deletion of inscriptions (from walls or blackboards). Implausible as this is (especially in preliterate societies), let us for the sake of argument grant the imagined empiricist his claim: that children acquire the concept of subject deletion from experience. Even so, the empiricist cannot explain why the child should <u>apply</u> that concept, to form a subject-deletion hypothesis. For according to empiricism, experience "serves ... to provide the things to be copied by the mind" (Katz 1966, 278n.), thus "suggesting" the hypotheses to be tested. If the child had access to the minds of adult speakers, he might have some experience of the subject deletion process, and this would suggest an hypothesis about subject deletion. But the child has no experience of this sort, and consequently (on the empiricist theory) he never even considers a subject deletion hypothesis. In fact, however, a child learning English apparently considers and adopts not only this hypothesis, but many other hypotheses about equally unobseravable transformational processes.[11]

To summarize the argument thus far: Harman was quite possibly right when he suggested that inference to the truth of a correct grammar lies within the scope of an empiricist learning device. (Katz's argument, at any rate, did not show the inadequacy of an empiricist device, provided the correct hypothesis is among those to be tested by the device.) He was wrong, however, in claiming that taxonomic linguists are more anti-theoretical than other empiricists. Finally, Chomsky's critique of taxonomic linguistics can be extended to apply to any theory of learning which shares with taxonomic linguistics a certain empiricist assumption: the assumption that sensory experience, rather than some innate mechanism, performs the crucial heuristic role in generating the grammatical hypotheses to be tested against primary linguistic data.

We turn now to Harman's "resourceful empiricist". He tries to explain language learning without appeal to innate principles, by postulating an innate bias instead, that favors some kinds of grammars and frowns on others. He need not be resourceful in order to determine what sort of bias to postulate. He can simply wait for a rationalistic linguist to tell him what kinds of grammars are learnable, and which are unlearnable. He then arranges his bias to favor learnable grammars over unlearnable ones. Presumably this can be done mechanically, without any resourcefulness, if it can be done at all. It is somewhat ironic, then, that Harman's hypothetical champion of empiricism should be described as resourceful. It is also ironic that Harman should appeal to Goodman in support of an innate bias that discriminates against some hypotheses, and in favor of others. For in Goodman's theory of projection, the difference between a valid and an invalid conclusion from experience "is not attributed to anything inevitable or immutable in the nature of human cognition" (Goodman 1965, 96-97). Rather, Goodman "regard[s] the mind as in motion from the start, striking out with spontaneous predictions in dozens of directions, and [only] gradually rectifying and channeling its predictive processes" (Goodman 1965, 87).

But these are only ironies in the doctrine of resourceful empiricism, not arguments bearing on the truth or falsity of that doctrine. I wish first to argue that any adequate plausibility ordering on grammatical hypotheses will have to use devices more powerful than those of Goodman's theory of projectibility.[12] Second, I shall point out two fundamental tenets of empiricist learning theory which Harman's "empiricist" seems quite willing to repudiate. Finally, I shall argue (contrary to Chomsky 1969 a,b, and indeed contrary to Harman himself, 1969) that a rationalist theory of language learning cannot be stated within the framework available to a resourceful empiricist. If this is correct, then resourceful empiricism, like empiricism properly so called, is disconfirmed by the evidence favoring a rationalist theory of language acquisition.

First, then, let us recall that Goodman proposes to measure the lawlikeness (confirmability) of hypotheses primarily in terms of the degree of entrenchment of the predicates they contain. Complexity of logical form may also be a factor in determining the *prima facie* plausibility of an hypothesis. Now transformational theorists, besides distinguishing between learnable and unlearnable grammars, postulate an innate evaluation procedure that imposes a plausibility ordering on the class of learnable grammars. But many of their hypotheses about the character of this plausibility ordering seem quite unlike hypotheses about the entrenchment of predicates.

Chomsky and Halle (1968, 404-407), for example, propose a set of thirty-nine so-called marking conventions, in an effort to account for some of the empirically established facts about naturalness (a priori plausibility) in phonological hypotheses. In eleven of these marking conventions, only a single phonological feature (predicate) is mentioned. Clearly these conventions cannot be telling us which of two predicates is the better entrenched. But even those conventions that mention more than

$$(7) \qquad \begin{bmatrix} +\text{vocalic} \\ -\text{consonantal} \end{bmatrix} \longrightarrow [+\ \text{voice}]$$

one phonological feature can hardly be construed as statements about the relative entrenchment of the predicates involved. Consider, for example, Chomsky and Hallé's marking convention V, here presented in abbreviated form as (7). This convention tells us that vowels (i.e., segments that are vocalic and non-consonantal) tend to be voiced rather than voiceless. By implication, the convention tells us that vowel devoicing rules (such as that in Cheyenne) make a grammar less plausible. But convention V does not tell us that the predicate voice is better (or worse) entrenched than the predicates vocalic and consonantal. It comments only on the naturalness of vowel devoicing rules, telling us nothing about the inherent plausibility of other rules that make use of the predicates in question.

One might still suppose it possible to specify an entrenchment ranking that would duplicate the overall effect of all thirty-nine marking conventions. This supposition seems quite improbable, however, in the light of some of the phenomena that

(8) $k \rightarrow \check{c} / \underline{\quad} \begin{bmatrix} -\text{cons} \\ -\text{back} \end{bmatrix}$

(9) $\check{c} \rightarrow k / \underline{\quad} \begin{bmatrix} -\text{cons} \\ +\text{back} \end{bmatrix}$

motivate the marking conventions. Chomsky and Halle cite rule (8) as a highly plausible rule that is "observed in many languages", and rule (9) as unnatural and "quite extraordinary despite [its] "simplicity" "(Chomsky and Halle 1968, 401). Rules (8) and (9) are virtually indistinguishable in formal complexity, by any plausible measure, while on the substantive side, exactly the same predicates appear, with exactly the same number of occurrences, in both rules. Evidently, then, it cannot be differences of syntactic complexity or predicate entrenchment that make rule (9) so much less plausible than rule (8).[13] Rather, the plausibility of (8), and the implausibility of (9) is apparently related to what Chomsky and Halle call the "intrinsic content" of the phonological features involved. If this is correct, then perhaps no general measure of a priori plausibility can deal in a satisfactory way with the facts about (8) and (9).

This raises the question whether a "resourceful empiricist" deserves to be considered an empiricist at all. Bever, Fodor, and Garrett (forthcoming, Chapter 8) point out that empiricist students of learning have generally committed themselves to the view that a single mechanism underlies all human learning. On this view, an adequate theory about the way we learn visual discrimination tasks should also be able to account for the way we learn English. Empiricists see no need to postulate a <u>task-specific</u> language-learning mechanism, that is, a mechanism that can accomplish only a <u>language</u>-learning task. Cohen, for example, writes:

> Children's tacit reliance, in their language-learning, on a principle of linguistic uniformity ... may be compared with their tacit reliance, in judgments about their siblings, on a principle of fairness, impartiality, or universalisability: the same underlying principle is at work. An inductive language-learning device is thus structurally isomorphic with an inductive ethics-learning device, as with an inductive device for natural-scientific discovery. (Cohen 1970a, 179n).

Putnam, criticizing Chomsky's "innateness hypothesis", expresses a similar view in more polemical language:

> Just how impressed should we be by the failure of current learnings theories to account for complex learning processes such as those involved in the learning of language? If innateness were a <u>general</u> solution, perhaps we should be impressed. But the Innateness Hypothesis <u>cannot</u>, by its very nature, <u>be</u> generalized to handle all <u>complex</u> learning processes. ... In the absence of any knowledge of what <u>general multipurpose learning strategies</u> might even look like, the assertion that such strategies (which absolutely must exist and be employed by all humans) cannot account for this or that learning process, that the answer or an answer schema must be 'innate', is utterly unfounded. (Putnam 1967, 20, 21; emphasis Putnam's)

This empiricist belief in a single pattern for all learning is implicit, I think, in the following remark by Hume:

> From causes which appear similar, we expect similar effects. This is the sum of all our experimental conclusions. (Hume 1748, 50; cf. also p. 112)

On the whole, then, empiricists have sought a general theory of human learning, eschewing task-specific learning mechanisms. But Harman's "empiricist" apparently does not share these scruples. He is willing to postulate a bias favoring "certain types of grammar ... no matter what the facts about language turn out to be" (Hume 1967, 86-87). In the later article, Harman indicates that as far as he is concerned, "there is no real difference between Professor Chomsky's "rationalist" view and Professor Putnam's suggestion that we should attempt to account for language-acquisition in terms of general multipurpose learning strategies" (Harman 1969, 150). Harman exhibits no interest in the question whether the innate bias favoring certain types of grammars is a reflection of some more general bias, or whether this bias applies to grammatical hypotheses. In this respect, the views of Harman's "resourceful empiricist" are unlike the views of other empiricists.

There is another tenet of empiricist learning theory which the "resourceful empiricist" seems willing to give up. That is the view that language acquisition is subject to the same psychological laws which govern learning in non-human animals (e.g., rats and pigeons). Empiricists generally resist the suggestion that the mechanism of language learning is a species-specific, uniquely human mechanism (cf. Bever, Fodor, and Garrett forthcoming, Chapter 8). Nowhere does Harman build this typical empiricist attitude into his description of the resourceful empiricist. Resourceful empiricism contrasts sharply, in this respect, with the empiricism of Skinner's Verbal Behavior (1957), or of Hume's discussion "Of the Reason of Animals" in the Treatise (1739, 176-179).[14]

We have just seen two important doctrines of empiricism that a resourceful empiricist seems willing to repudiate. There is reason to doubt, then, whether a "resourceful empiricist" is a real empiricist, or merely a counterfeit one. I wish now to dispute Harman's claim (1969, 151) that resourceful empiricism "coincides with rationalism resourcefully defended". (Chomsky agrees with Harman that "resourceful empiricism" incorporates, as a special case, the "rationalist" approach that [he has] been trying to develop" (Chomsky 1969b, 158), and that resourceful empiricism is "immune to any factual discovery" (Chomsky 1969a, 91).)

Harman recognizes, of course, that there is a formal difference between the rationalist account of language acquisition and that proposed by the resourceful empiricist. According to the resourceful empiricist, only certain inductive principles, and a bias in favor of certain types of grammar, are innate. On the rationalist theory, there are innate linguistic principles, as well as an innate bias and innate inductive principles. But what are the linguistic principles which Harman regards as a

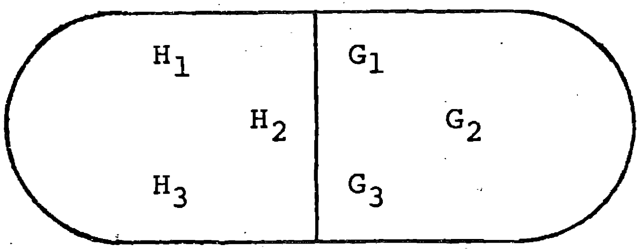

Figure 2

$G_1, G_2, G_3, \ldots, G_i, \ldots$

(Plausibility ranking of learnable grammars only)

Figure 3

$G_1, G_2, G_3, \ldots, G_i, \ldots, G_f, H_1, H_2, H_3, \ldots$

(Plausibility ranking of all logically possible grammars, learnable or not)

Figure 4

dispensable part of the rationalist theory? Apparently he means to do away with the principles used to express the finding that "only languages with certain types of grammar (e.g., transformational grammar) are learnable" (Harman 1967, 86). These are the principles which define the concept <u>possible generative grammar</u> (Chomsky 1965, 30-31), and thus implicitly partition the class of all logically possible grammars into those which are learnable (in the normal way) and those which are not. This partition is illustrated in Figure 2. In addition, the rationalist theory includes an evaluation procedure that established a plausibility ordering (Figure 3) on the learnable grammars. Harman apparently means to eliminate the partition of Figure 2 by specifying a comprehensive plausibility ranking that is biased in favor of the learnable grammars, and (by implication) against the unlearnable ones (cf. Figure 4). Intuitively speaking, Harman means to gloss the rationalist's <u>unlearnable</u> as <u>very difficult (but not impossible) to learn</u>. No one will in fact learn any of these "very diffficult" grammars, because long before he reaches H_1, the first such grammar in the list, he will find the grammar (say G_i) which governs the speech of his community. He will test G_i and, finding no counterexamples to G_i, he will adopt G_i as his grammar, thus successfully completing his learning task.

Unfortunately, there cannot be any such ranking as that of Figure 4, which "swallows up" the class of unlearnable grammars along with the learnable ones.[15] For in order to favor learnable over unlearnable grammars, the ranking would have to begin with a listing of <u>all</u> the learnable grammars. The first <u>unlearnable</u> grammar in the list, H_1, would have to follow the last learnable grammar, say G_f, because H_1 is less plausible than any learnable grammar. But there can be no least plausible learnable grammar, because the set of learnable grammars is infinite, for much the same reason that the class of English sentences is infinite: there is no upper bound on the length of grammars.

(10) Dillinger kicked the bucket.

In English, as in other languages, there are idiomatic expressions, i.e. expressions whose meaning cannot be discovered by the normal process of semantic interpretation. The verb phrase <u>kick the bucket</u> in sentence (10) is an example. The lexicon of English must contain a special entry, indicating that this VP, despite appearances, means <u>die</u>. Now suppose that G_f is some particular grammar, putatively the least plausible of all. It is easy to construct a learnable grammar that is less plausible than G_f: we simply select an arbitrary VP (there are infinitely many, since there is no upper bound on the length of VP's) and an arbitrary lexical reading (other than the normal one for the selected VP), put them together to form a lexical entry, and add this new lexical entry to the lexicon of G_f. The resulting grammar is learnable, but less plausible than G_f. This rebuts the hypothesis that G_f is the least plausible learnable grammar, and shows that the class of learnable grammars

$$\boxed{G_1,\ G_2,\ \ldots,\ G_{10^{50}},\ H_1,\ G_{10^{50}+1},\ H_2,\ G_{10^{50}+2},\ \ldots}$$

Figure 5

is infinite.

Of course there is a way of incorporating the unlearnable grammars into a single list with the learnable ones, despite the fact that both sets are infinite. We could begin such a list with a very long sequence of learnable grammars, say 10^{50} of them. The last grammar in this initial subsequence, $G_{10^{50}}$, will be a learnable grammar, but such a complex and implausible one as to rule out the possibility that any human could actually acquire $G_{10^{50}}$ in a lifetime. Immediately after $G_{10^{50}}$, we insert the first unlearnable grammar in the list, H_1. Next comes $G_{10^{50}+1}$, then H_2, then $G_{10^{50}+2}$, and we continue the list by interleafing unlearnable grammars with learnable ones (cf. Figure 5). Every grammar, learnable or not, will appear in the resulting list.

The difficulty with the list of Figure 5 is that this list, interpreted as an hypothesis about our innate capacity for language-learning, is palpably in conflict with the rationalist hypothesis (Figures 2 and 3). Figure 5 claims, for example, that H_1 is more plausible than $G_{10^{50}+1}$. The rationalist hypothesis, on the contrary, implies that $G_{10^{50}+1}$ is the more plausible grammar, since every learnable grammar is more plausible than any unlearnable one. Our intent in constructing the list of Figure 5 was to duplicate the claims of the rationalist hypothesis, without resort to a partitioning of the class of all possible grammars. Clearly, our efforts have not been successful.

Thus far, I have been assuming that the list representing the resourceful empiricist's innate bias must be of order type ω.

$G_1, G_2, G_3, \ldots, H_1, H_2, H_3, \ldots$

Figure 6

But suppose we allow a list of order type $\omega + \omega$, as in Figure 6. Surely this will do the trick for Harman?

Here the resourceful empiricist faces a dilemma. For each finite set of possible sentences, he must predict what grammar a child will acquire if he is exposed to just that corpus of possible sentences.[16] (The resourceful empiricist's theory will be incomplete if it does not meet this requirement.) Consequently, either there is or there is not some collection of primary linguistic data which would, according to the resourceful empiricist, refute every one of the G_i, and thus induce the child to adopt H_1 as his grammar. If the resourceful empiricist claims that there is such a corpus, he clearly is asserting something the rationalist denies. If, on the other hand, he denies that there is such a corpus, he is in effect saying that the language-learning mechanism is structurally incapable of ever considering any of the H_j (the unlearnable grammars). In this case, the resourceful empiricist is in perfect agreement with the rationalist, but he has admitted that the H_j are not merely very difficult, but in fact impossible to learn in the normal way. His plausibility ranking simply reproduces, in thinly disguised form, the very partition (Figure 2) it was meant to eliminate.

The first horn of this dilemma, by the way, suggests a new way of looking at the "interleafed" plausibility ranking of Figure 5. That ranking implicitly claims that there is some corpus C of linguistic data, such that presentation of C would lead the acquisition device to reject the first 10^{50} learnable grammars, and accept H_1 on a tentative basis. The rationalist predicts, on the other hand, that presentation of C would lead to the acceptance of $G_{10^{50}+1}$, not to the acceptance of H_1. While it is not possible to arrange a direct test to settle this disagreement, there will surely be evidence that bears on it indirectly, namely evidence about language learners' responses to corpora smaller than C. We can then ask: which of the two hypotheses projects the pattern of observed responses in the simpler, more natural way? Given the highly artificial construction of the resourceful empiricist's list, it seems likely that he will lose out when considerations of naturalness and simplicity are taken into account.

It seems, then, that the resourceful empiricist cannot make good on his offer to paraphrase away the difference between unlearnable and very-difficult-to-learn grammars. If I read Harman correctly, then resourceful empiricism does not include "any specific empirical proposal that anyone can formulate" (Chomsky 1969b, 158). To be sure, "it can accomodate the view that knowledge of English is innate", by placing the grammar of English first in its plausibility ranking. But it is specifically intended to rule out the claim that "knowledge of universal grammar is innate" (Chomsky 1969b, 158). Any evidence, then, which tends to show that certain types of grammar are not accessible to the human language-learning mechanism, is evidence against resourceful empiricism. The resourceful empiricist thus occupies a position distinct from that of the rationalist. But his position is also different from traditional empiricism, because of his willingness to countenance innate heuristics, and learning mechanisms that are task- as well as species-specific. Since resourceful empiricism is a much weaker position than genuine empiricism, even a full vindication of resourceful empiricism would not be enough to save empiricism. And I know of no reason to suppose that resourceful empiricism can be vindicated over against the rationalist theory of language acquisition.

III. COHEN: CAN ENRICHED EMPIRICISM EXPLAIN LANGUAGE LEARNING?

Harman tried to argue that rationalism and empiricism are not distinct positions, that any rationalist theory about language learning can be mechanically translated into an empirically equivalent empiricist hypothesis. L. J. Cohen, whose paper (1970b) we shall take up next, acknowledges that rationalism and empiricism are empirically distinguishable hypotheses about the internal structure of a language-learning device. He maintains (plausibly enough, I think) that if an empiricist theory can account for the facts about language acquisition, then we should reject the rationalist theory, on the grounds that it attributes an unnecessarily complex internal structure to the child's language acquisition device (Cohen 1970b, 305, 306). But as Katz has pointed out, "simplicity is a consideration in choosing between competing hypotheses only when the hypotheses can afford an equally plausible basis on which to explain the available evidence" (Katz 1966, 268-269). I shall argue that Cohen, despite the innovative aspects of his theory of induction, does not present a successful explanation of language learning.

Cohen is critical of recent attempts (especially that of Chomsky and Miller, 1963) to show the inadequacy of inductivist theories of language learning. These anti-empiricist arguments, he complains, have concentrated on an enumerative conception of induction, which "stems from ... the associationist tradition that goes back not to Bacon but to Hume". In enumerative induction, "it is the sheer number of evidential instances ...

that makes one ... induction better than another". Bacon, and more recently Mill, emphasized the importance of variational induction, which takes account of "the value of variety, as distinct from mere multiplicity, in experimental instances" (Cohen 1970b, 300). Cohen suggests that a theory of language learning based on variational induction may not share the defects of theories based on enumerative induction. He argues that his theory of variational induction invalidates the usual arguments against empiricism, and in favor of the innateness of syntactic universals.

Three questions face us in attempting to understand Cohen's paper: What is Cohen's theory of variational induction? How does he propose to apply it to the explanation of language learning? And in what way is variational induction supposed to remedy the defects of enumerative theories of language learning?

On the first question, the concept of a (finite) sequence of canonical tests is central to Cohen's theory. Given certain information about the field of research within which an hypothesis lies, it is possible to calculate the design of this sequence of tests. Each test in the sequence is more thorough than its predecessor. If an hypothesis passes every test in this (finite) sequence, it is said to enjoy full inductive support. A fully supported hypothesis may, of course, be false.

In order to calculate the sequence of canonical tests for a given hypothesis H, we need two pieces of information (Cohen

1970b, 301; 1970a, 51-60). The first is a specification of the vocabulary used to formulate hypotheses in the same field of research as H. This vocabulary will most likely include many predicates that do not appear in H itself, of course. Any hypothesis that can be formulated in this vocabulary is said to be <u>materially similar</u> to H.[17]

Second, we need a list of variables that have proven to be <u>inductively relevant</u> to hypotheses materially similar to H. This list of variables is taken to be disjoint (initially, at least) with the vocabulary used to define material similarity to H. A variable is inductively relevant to a set of hypotheses just in case manipulation of that variable has led to the refutation of at least one hypothesis in the set. A variable V has greater inductive relevance than variable W to a set S of hypotheses, just in case manipulation of V has refuted more hypotheses in S than has manipulation of W. Our (finite) list of variables is to be ranked in order of decreasing inductive relevance. The list begins with that variable having the greatest inductive relevance to H-like hypotheses, and ends with the variable which has the poorest track record of falsified hypotheses. Also, the values of each variable in the list are to be "observable circumstances" (Cohen 1970b, 301).

Given this list of inductively relevant variables, we are at last in a position to calculate the canonical test sequence for H. For the first test in the sequence, we manipulate only the first variable in our list, while holding all other variables constant. In the second test, we manipulate (independently, of course) the first two variables in our list, holding the others constant. And so on, until we reach the final test in the sequence, in which we simultaneously manipulate all the variables of known relevance, from the greatest unto the least.

Now if H happens to pass all these tests, it has full inductive support. But to see one of the alleged advantages of variational induction, let us consider a case in which an

(11) All hares are gray.

hypothesis gets refuted in the last test of its sequence. The hypothesis Cohen discusses is (11), and he assumes that "variation from arctic to temperate zones was considered the least relevant variable for hypotheses about fur-color" (Cohen 1970b, 302). Naturally, a white arctic hare turns up when we finally manipulate

(12) All temperate-zone hares are gray.

this variable, and we must reject (11). But variational induction tells us how to convert (11) into a slightly more complex hypothesis, (12), that has full inductive support, without any further observations. This modification in effect removes the arctic/temperate variable from the list of variables to be manipulated, thus shortening the sequence of canonical tests, and speeding the confirmation of a putatively correct hypothesis.

Such modifications need not, of course, involve the last variable on our list for a given hypothesis. We can deal in similar fashion with counterexamples that arise earlier in the test sequence. Such modifications will not result in hypotheses with full support, of course, but even here, variational induction facilitates the process of confirmation. This is because "the more we modify and complicate our hypothesis, in one of the ways described, the smaller is the number of evidential instances, or individual experiments, that we need to obtain full support for a modified version of the hypothesis" (Cohen 1970b, 302). One of the advantages of variational induction, then, is supposed to be that it leads more rapidly to the (full) confirmation of true hypotheses, because of the hypothesis-modification procedure just described.

Though Cohen does not claim any further advantages in the paper under discussion (Cohen 1970b), I gather that variational induction is also meant to promote quicker refutation of false hypotheses. This surely is part of the reason Cohen instructs us to begin a test sequence by manipulating the variable having the greatest inductive relevance. These, then, are the two ways in which variational induction is supposed to be more efficient than enumerative induction: it speeds the refutation of false hypotheses, and the confirmation of true ones.

We come now to our second question: how might variational induction be used to explain language learning? First, how would "an inductive language-learning device...acquire the concepts in terms of which it could construct its syntactic hypotheses" (Cohen 1970b, 304)? Among these concepts "necessary for generalizations about surface structure" are concepts like <ins>noun</ins>, <ins>verb</ins>, and <ins>sentence</ins>. All these concepts

> can be reduced to two primitive ones ("sentence" and "nominal") by assuming a mode of derivation for the others like that used in categorial grammars. Then the acquisition of these two primitive concepts has to be supposed to stem, as to "sentence," from experience of utterances formed by varying recombinations of a stock of recognizably recurring components, and, as to "nominal," from experience of the utterance of some such components in isolation from others (in the presence, no doubt, of what Quine called "conspicuously segregated objects"). (Cohen 1970b, 304)

This, then, is the process by which the child is supposed to acquire the vocabulary in terms of which he defines the class of grammatical hypotheses, an equivalence class with respect to the relation of material similarity. He uses this vocabulary initially to formulate "generalizations about surface structure", e.g. the hypothesis that "any string of the form noun-verb-noun is grammatical in English" (Cohen 1970b, 302).

But a variational language-learning device is not "confined to ... listing labeled bracketings of non-deviant surface structures" (Cohen 1970b, 303). As it begins to find counterexamples to its initial hypotheses, which <ins>are</ins> about surface structure, "a store of relevant variables can gradually be amassed: e.g., the singular-plural variable for nouns preceding verbs, the singular-plural variable for verbs, the transitive-intransitive variable for verbs, and so on" (Cohen 1970b, 302). The device

> (13) Any noun-verb-noun sequence in which both the verb and the preceding noun are singular, is grammatical in English.

(14) Any noun-verb-noun sequence in which both the verb and the preceding noun are plural, is grammatical in English.

mentions specific values of these variables in modifying its initial hypotheses. Its modified hypotheses might include (13) and (14), for instance. Then

> its nisus toward generalization will lead it to formulate [second-order, correlational] hypotheses about relationships between these variables which will subsume and explain, as it were, the more elementary hypotheses that have already been established: e.g., "The singular/plural variable for verbs varies directly with the singular/plural variable for preceding nouns," or "The insertability of a <u>by</u>... phrase after the verb varies inversely with the active/passive variable." (Cohen 1970b, 303)

Number agreement and passivization are two well-known transformational processes. I take it, from the passage just quoted, that Cohen regards transformations as second-order (correlational) hypotheses that express relationships between (and thus quantify over) surface-structure variables, such as number in verbs and nouns.

This, then, is how Cohen applies variational induction to the theory of language acquisition. We come now to our third question: how is the variational theory of language learning supposed to remedy the defects of theories based on enumerative induction?

To review our discussion of Harman (pp. 228-229), the principal difficulty with inductivist language-learning devices was their inability to generate hypotheses about deep structures and transformational processes. Now on Cohen's theory, a child's <u>initial</u> hypotheses are "generalizations about surface structure". But the benefits claimed for variational induction have to do with hypothesis testing and modification, not with the generation of initial hypotheses. And it is in fact the testing and modification process which leads to the development of transformational hypotheses, according to Cohen. Refutation of hypotheses leads us to recognize the inductive relevance of various syntactic variables, and thus makes it possible to formulate (transformational) generalizations about relationships among these variables.

Cohen is not the first empiricist to suggest that transformations are absent from a child's first grammatical hypotheses. Martin Braine (1963, 1965) thought that a child first learned to use sentences whose deep structures were identical with their surface structures. Later, they learned to use sentences transformationally derived from these elementary sentences. But as Bever, Fodor, and Weksel (1965a,b) pointed out in their critique of Braine's theory, there <u>are</u> no sentences whose surface and deep structures coincide; the derivation of <u>every</u> sentence involves transformational processes.

This observation in itself does not defeat Cohen's (or Braine's) theory. There is no reason why a child's <u>first</u> hypothesis must be correct; Cohen only needs to explain how the child comes to adopt the correct grammar as his <u>final</u> hypothesis. And of course Cohen does suggest a process by which transformation rules might enter the child's grammar.

Unfortunately, Cohen (like Braine before him) misconceives the nature of transformation rules. A grammatical transformation does not derive the surface structure of a sentence from another surface structure, not even from the surface structure of some other sentence. Rather, transformations operate on deep structures, and no deep structure is identical with any surface structure. To be sure, the present concept of a transformation rule appears to have developed out of work by Zellig Harris (1957), in which he spoke of "a formal [transformational] relation among sentences", i.e. surface structures (Harris 1957, 283; emphasis added). But since at least 1957, transformations have been conceived as operations on abstract deep structures, or underlying phrase markers (cf. Chomsky 1957, 44-48; Chomsky 1965, 128-129).

Now Cohen's theory tells us how children might learn Harris-style transformations, but that is not the accomplishment which a theory of language learning needs to explain. The switch to a variational theory of induction cannot help the empiricist to get at deep structures and transformational operations on them, for a very simple reason. In variational induction, we modify a refuted hypothesis by constraining an inductively relevant variable. Such variables take "observable circumstances" as their values. The child acquires them by "perception of a difference between his own utterance and some adult speaker's utterance" (Cohen 1970b, 303).

But neither deep structures nor transformations are perceptible. Consequently, if there is no reference to deep structure in the hypothesis to be modified, then the modified hypothesis will not contain any such reference either. The strength of variational induction supposedly lies in the modification process; but that process only adds references to new surface-structure variables. If Cohen's theory is to explain how a child learns a transformational grammar, then reference to deep structure must appear in the child's _first_ hypothesis, before the alleged advantages of variational induction can be realized.

Cohen's paper contains a remark that might be construed as the beginning of an answer to this criticism. He suggests that "the more abstract concepts of transformational grammar" may be innate, rather than acquired, and that these concepts are "capable of non-linguistic as well as of linguistic realization" (Cohen 1970b, 304). Now, he might argue, some of the concepts needed to describe deep structure, and to formulate transformations, are acquired: _NP_, _plural_, etc. The rest of the necessary concepts are innate, but not specifically linguistic. (This latter restriction expresses the empiricist ban on task-specific learning mechanisms.) With all the requisite concepts at hand, why can't the child's _first_ hypothesis be a conjecture about deep structure, and about the operations that map it into surface structure?

This proposal claims that the child learns all his specifically linguistic concepts from his experience with surface structure. Unfortunately, some deep-structure constituents (e.g., Imperative) never appear in surface structure, so the child could not learn the corresponding concepts from experience. But the proposal has a far more serious defect. Let us suppose that concepts like deletion, permutation, etc. are innate, and "capable of non-linguistic... realization". Still, all we have thus far is a collection of conceptual materials. How is the child supposed to distinguish between promising and unpromising ways of putting these materials together into hypotheses? Not, of course, by observation, since transformational processes are not observable. Moreover, the child's first hypothesis is bound to be refuted. When this happens, how does the child know what changes to make in the refuted hypothesis? Again, observation is not likely to help, for transformational grammars are hypotheses about unobservable operations on unobservable entities (deep structures).

The point is that substantive universals, linguistic or otherwise, do not tell us what form a first hypothesis (or a modified hypothesis) must take in order to be plausible. For this, we need formal and organizational universals (cf. Katz 1972, 30-31), plus an evaluation procedure that defines the relation more plausible than on the class of admissible hypotheses. Cohen has of course suggested a method of hypothesis modification, and he claims that that method is applicable to non-linguistic learning tasks. We have reason to doubt, however, that Cohen's modification procedure can make appropriate changes in hypotheses about unobservable entities and processes.

As for the heuristic task of formulating a plausible initial hypothesis, Cohen's modification procedure can be of no help in this process. It would be inimical to Cohen's empiricist principles to invoke an innately (rather than environmentally) determined heuristic process (cf. pp. 234-236 above), let alone a plausibility metric defined only on grammatical hypotheses. But since the heuristic device must generate an hypothesis about unobservables, it is unlikely that the operation of the child's heuristic depends on experience as heavily as empiricists have supposed. And Cohen's paper gives us no reason to suppose that any general, multipurpose heuristic can generate grammatical hypotheses of the appropriate sort. With or without variational induction, Cohen's theory of language acquisition is inadequate, and cannot be used to vindicate empiricism.

In summary, the rationalist account of language learning does not require the postulation of infinite storage capacity, as Morgenbesser argues. Contrary to Harman's contention, rationalism is empirically distinct from empiricism. And empiricism seems unable to account for language learning, even when enriched by Cohen's variational theory of induction.

Footnotes

*I wish to thank Sylvain Bromberger and Ned Block for their help in revising an earlier version of this paper.

1 The entity Morgenbesser has in mind is the nervous system. Morgenbesser goes on to suggest a way to justify making state S a knowledge state. That is, he puzzles over how Fodor might justify a claim that some part of the nervous system knows something. Cf. Morgenbesser 1969, 463-464.

2 The concept of tacit action raises many issues in the theory of action, e.g.: When one of my little men performs an action, is that action *ipso facto* an action (tacit or otherwise) of mine? I shall have nothing to say about such issues.

3 For discussion of a similar argument, cf. Chapter 2 of this thesis, pp. 116-123; also Smart 1963, 38, and Katz 1972, 29.

4 In an earlier draft, my discussion of this point was excessively complex. I wish to thank my wife for a radical simplification of this discussion.

5 Similar diagrams can be drawn for systems that perform tasks involving sequential, rather then simultaneous, occur of elementary operation. The added complexity would only complicate the discussion, however. Despite its simplicity, Figure 1 seems to me to do full justice to Nagel's point.

6 By picturing the nervous system as a sort of Hobbesian

leviathan, we have also made it necessary to specify the little man for whom a given operation is elementary. We can no longer simply say that an operation is elementary for the organism, or for its nervous system as a whole. Elementarity is thus relative to systems as well as to tasks. An operation that is elementary for a superintendent may set his subordinates to performing tasks of considerable complexity. It may thus be somewhat misleading to say (as Fodor does) that an elementary operation is without "theoretically relevant internal structure" (Fodor 1968, 629), because we may give the superintendent credit for an entire operation, including the labors of his underlings.

7 In "correlating" sentence (1) with sentence (2), the idea is, I take it, that we mentally transform sentence (1) into sentence (2). We can easily interpret (2), because all semantical constituents are explicitly represented in (2). We give sentence (1) the semantic interpretation we get for sentence (2), on the assumption that "correlation" preserves meaning. Katz's question is: how do we know which correlations (transformations) preserve meaning, and which do not?

8 Katz (1966, 260-261) is somewhat unfair, I think, to the empiricist's correlational theory about the interpretation of subjectless imperatives. He mistakenly derives from that theory the requirement that in order to interpret a particular subjectless imperative S, we must first have encountered the actual sentence S' with which S is correlated. (The sentence S' begins with a <u>you</u> subject, but thereafter it is identical, constituent for constituent, with S.)

9 Notice, by the way, that a child would have to know a principle of conversational implicature in order to draw the appropriate conclusion from this fact: the principle that people do not usually issue commands which obviously cannot be executed. The child reasons: "Surely, he must know we can't all hide behind the easy chair, so he can't be addressing all of us." The earlier argument against correlating (1) with (3) may require appeal to another such principle: a speaker does not usually address someone whom he knows to be absent. "He can't be talking to any John Jones, because no John Jones is present."

10 Cf. note 9 above. Also, to provide a manageably small set of hypotheses, it may be necessary to resort to an innate evaluation measure that is applicable only to the evaluation of grammatical hypotheses; cf. Chomsky 1965, 203, note 22. Also, in case other than that of the imperative rule, considerations of simplicity of the general (vs.grammar-specific) sort may not dictate the correct choice among hypotheses compatible with a given body of data. But in general, an incorrect choice based on simplicity considerations may be rectified by gathering further data.

11 I have presupposed what Fodor (1971) calls the "strong psychological reality" position, but a refutation of that position would clearly do nothing to enhance the tenability of empiricism. However we recover the deep structure of a sentence from its phonetic representation, the process is clearly not an observable one.

12 This argument is not intended to show that resourceful empiricism is false. It goes rather to the character of the bias a resourceful empiricist must postulate, if he wants a theory that is empirically adequate. This argument may, then, tend to show, not that resourceful empiricism is not true, but that it is not _empiricism_.

13 Even outside linguistics, confirmability apparently depends on more than entrenchment of predicates. Davidson (1966) has pointed out, for example, that the statement _All emeroses are gred_ is, despite its use of ill-entrenched predicates, a lawlike statement. (Note, by the way, that we can confim this statement by observing rubies before _t_, and emeralds after _t_, though such rubies are not, and such emeralds may not be, emerubies.) Also, an hypothesis containing ill-entrenched predicates (e.g., _All emerphires are grue, and all sappheralds are bleen_) may be logically equivalent to an hypothesis of similar complexity, in which no such predicates occur (_All emeralds are green, and all sapphires are blue_).

14 Earlier (pp. 223ff.) I pointed out that empiricists look to experience, rather then to any innate schematism, to perform the heuristic function of generating hypotheses, which the learner then tests against further experience. But Chomsky (1965, 203, note 22) has pointed out that an evaluation measure may play a key role in selecting hypotheses for testing. The bias postulated by the "resourceful empiricist" may offend genuine empiricists, then, by its innateness as well as by its task- and species-specificity.

15 My argument in this paragraph assumes that Harman's list of all grammars, learnable and unlearnable, must be of order type ω. For an argument that does not rely on this assumption, cf. pp. 240-241 below.

16 The set of all possible sentences is defined in universal phonetic theory; cf. Chomsky 1965, 31.

17 It is not clear to me just how one would go about deciding which predicates belong to this vocabulary and which do not. This problem appears to be a crucial one for Cohen's theory, since the constitution of this vocabulary affects the design of canonical test sequences and thus the assessment of inductive support for hypotheses. I shall have nothing to say about this problem, however.

References

Bever, T., J. Fodor, and M. Garrett (forthcoming) Language Structure and Verbal Behavior.

Bever, T., J. Fodor, and W. Weksel (1965a) "On the Acquisition of Syntax: A Critique of "Contextual Generalization"," Psychological Review 72, 467-482.

Bever, T., J. Fodor, and W. Weksel (1965b) "Is Linguistics Empirical?," Psychological Review 72, 493-500.

Bloomfield, L. (1916) "Subject and Predicate," Transactions of the American Philological Association 47, 13-22. Reprinted in Hockett (1970), 70-77.

Braine, M. (1963) "On Learning the Grammatical Order of Words," Psychological Review 70, 323-348.

Braine, M. (1965) "On the Basis of Phase Structure: A Reply to Bever, Fodor, and Weksel," Psychological Review 72, 483-492.

Chomsky, N. (1957) Syntactic Structures, Janua Linguarum No. 4, Mouton & Company, The Hague.

Chomsky, N. (1965) Aspects of the Theory of Syntax, MIT Press, Cambridge, Mass.

Chomsky, N. (1969a) "Linguistics and Philosophy," in Hook (1969).

Chomsky, N. (1969b) "Comments on Harman's Reply," in Hook (1969).

Chomsky, N. and Halle, M. (1968) The Sound Pattern of English, Harper and Row, New York.

Chomsky, N. and Miller, G. "Finitary Models of Language Users," in Luce, Bush, and Galanter (1963).

Cohen, L. (1970a) The Implications of Induction, Methuen & Company, London.

Cohen, L. (1970b) "Some Application of Inductive Logic to the Theory of Language," American Philosophical Quarterly 7, 299-310.

Davidson, D. (1966) "Emeroses by Other Names," Journal of Philosophy 63, 778-780.

Fodor, J. (1968) "The Appeal to Tacit Knowledge in Psychological Explanation," Journal of Philosophy 65, 627-640.

Fodor, J. (1971) "Current Approaches to Syntax Recognition," in Horton and Jenkins (1971).

Gleason, H. (1958) Introduction to Descriptive Linguistics, Holt, Rinehart, & Winston, New York.

Goodman, N. (1965) Fact, Fiction, and Forecast, Bobbs-Merrill, New York.

Graves, C., J. Katz, Y. Nishiyama, S. Soames, R. Stecker, and P. Tovey (in press) "Tacit Knowledge," Journal of Philosophy.

Harman, G. (1967) "Psychological Aspects of the Theory of Syntax," Journal of Philosophy 64, 75-87.

Harman, G. (1969) "Linguistic Competence and Empiricism," in Hook (1969).

Harris, Z. (1957) "Co-occurrence and Transformation in Linguistic Structure," Language 33, 283-340.

Harris, Z. (1963) Methods in Structural Linguistics, University of Chicago Press, Chicago.

Hockett, C. (1970) A Leonard Bloomfield Anthology, Indiana University Press, Bloomington.

Hook, S., ed. (1969) Language and Philosophy, New York University Press, New York.

Horton, D. and Jenkins, J., eds. (1971) The Perception of Language, E. Merrill, Columbus.

Hume, D. (1739) A Treatise of Human Nature, L. A. Selby-Bigge, ed., Oxford University Press, London, 1968.

Hume, D. (1748) An Enquiry Concerning Human Understanding, C. W. Hendel, ed., Bobbs-Merrill, New York, 1955.

Katz, J. (1966) The Philosophy of Language, Harper & Row, New York.

Katz, J. (1972) Semantic Theory, Harper & Row, New York.

Locke, J. (1690) An Essay Concerning Human Understanding, A. Fraser, ed., Dover, New York, 1959.

Longacre, R. (1964) Grammar Discovery Procedures, Janua Linguarum No. 33, Mouton & Company, The Hague.

Luce, R., R. Bush, and E. Galanter, eds. (1963) Handbook of Mathematical Psychology, v. 2, John Wiley & Sons, New York.

Morgenbesser, S. (1969) "Fodor on Ryle and Rules," Journal of Philosophy 66, 458-472.

Nagel, T. (1969a) "Linguistics and Epistemology," in Hook (1969).

Nagel, T. (1969b) "The Boundaries of Inner Space," Journal of Philosophy 66, 452-458.

Putnam, H. (1967) "The 'Innateness Hypothesis' and Explanatory Models in Linguistics," Synthèse 17, 12-22.

Skinner, B. (1957) Verbal Behavior, Appleton-Century-Crofts, New York.

Smart, J. (1963) Philosophy and Scientific Realism, Routledge and Kegan Paul, London.

Appendix

Gettier, Idiolects, and Linguistic Rationalism*

According to Graves, Katz, _et_ _al_., "establishing that speakers know some principles of linguistic theory does not show that they know that ... the principles in question are principles of every possible grammar" (Graves, Katz, _et_ _al_. in press). By "principles _of_ every possible grammar" I take it these authors mean "principles _about_ all possible grammars". That is, I take them to be denying that we have innate knowledge of any universal principles which quantify over possible grammars. I wish in this appendix to argue, on the contrary, that we do know many such principles. The argument for this thesis will bring to light a problem with the rationalist account of linguistic knowledge, a problem related to Gettier's refutation of the traditional analysis of knowledge.

It will be convenient to begin the argument by considering a passage from another recent publication by Katz:

> ... asking "What is analyticity in English?" or "What is synonymy in Chinese?" is no less foolish than asking "What is a toothache for Englishmen?" or "What is the relation of being sicker than among Chinese?" ...the concepts of analyticity, synonymy, and so on [are not to be] conceived of as relative notions, to be explicated with respect to the particular structure of one or another natural language... Rather, they are to be conceived of as absolute notions, as semantic properties and relations exhibited by words, phrases, clauses, and sentences in each and every natural language. (Katz 1972, 11-12)

There is a danger that this passage may be misinterpreted. Katz's point, I take it, is that one can define the relation _S is analytic in L_, or _S and T are synonymous in L_, for variable L as well as variable S and T. That it, we need not take _L_ in these schemata to refer to any _particular_ language; rather, it is to be construed as a variable ranging over all possible human languages. This point is quite correct. The theory of transformational grammar meets the challenge, laid down by Quine in "Two Dogmas of Empiricism", to "explain 'S is analytic for L' generally for variable 'L'" (Quine 1961, 34).

But Katz's remarks might be taken to imply something stronger than this. They might be read as claiming that synonymy, for instance, is strictly speaking a two-place relation, which takes

(1) _Oculist_ and _eye doctor_ are synonymous in English.

(2) _Oculist_ and _eye doctor_ are synonymous.

expressions as its arguments. On this view, sentence (1) is a needlessly verbose way of stating the fact expressed by (2). There is nothing incomplete or elliptical about (2), on this view; in particular, no reference to a language (or to a grammar) is necessary in order to render (2) fully explicit.

This view might seem to gain further support from the

(3) _Messer_ and _couteau_ are synonymous.

observation that (3) cannot be rendered "explicit" by adding a reference to any single language. _Messer_ and _couteau_ are not synonymous in German, because _couteau_ is not a German word. Nor are they synonymous in French, since _Messer_ is not a French term. They are, apparently, synonymous _simpliciter_.

I wish to argue that synonymy is not merely a two-place relation, that grammatical is not strictly speaking a one-place predicate, etc. Rather, I maintain that a sentence is grammatical with respect to a specifiable language or grammar, that expression pairs are synonymous relative to a pair of languages or grammars, etc. (Throughout this discussion, by the way, I shall identify and distinguish sentences not by their deep or surface structure, but only by their phonetic characteristics.) I thus maintain that sentences like (2) and (3) above are

(4) German Messer and French couteau are synonymous.

(5) The cat is on the mat is grammatical.

(6) The cat is on the mat is grammatical in English.

elliptical versions of (1) and (4) respectively. Similarly, I regard (6) as a more explicit version of (5). None of this seems incompatible with the intent of Katz's remarks, quoted above.

The need for reference to a language is fairly obvious when

(7) The cat is on the mat is grammatical in Spanish.

we are dealing with "one-place" predicates, such as grammatical. Thus The cat is on the mat is not grammatical tout court, irrespective of language; sentence (7), for example, is false. The point is somewhat more elusive, I think, when we consider relation terms, such as synonymous. To see that synonymy, like grammaticality, is relative to language, let us consider a sentence which has been said (I forget by whom) to be grammatical in both English and German. The sentence, in its English spelling, is _Euripides leaped_; a German would write _Euripides liebt_. Let us adopt a language-independent phonetic spelling: [yuwripidiyz liypt].

Now a question arises as to "the" meaning of this sentence: is it or is it not synonymous with the English sentence, _Euripides loves_? Clearly, there is no answer to this question; or if you like, there are two answers to it:

(8) German [yuwripidiyz liypt] is synonymous with (English) _Euripides loves_.

(9) English [yuwripidiyz liypt] is not synonymous with (English) _Euripides loves_.

The example is not altogether satisfying, since _love_ and _lieben_ are not object-deleting verbs. Nonetheless, the general point is clear. The theory of universal grammar specifies a set of possible sentences (Chomsky 1965, 31), phonetically represented. A given possible sentence will generally be grammatical in some possible languages (_i.e._, it will be generated by some possible grammars), and ungrammatical in others. Not all the grammars that generate a given possible sentence will necessarily assign it the same structural descriptions or semantic imterpretations. The grammatical and semantic properties of a sentence are not, then, properties it has in and of itself; a sentence has these properties by virtue of its relationship to this or that grammar or language.

This same point may be brought out in a somewhat different way by considering the situation of a child who is learning English. The child, by virtue of his knowledge of universal grammar, considers only those grammatical hypotheses which are grammars of possible human languages. His experience thus far has led him to adopt, as his tentative grammar of English, a certain possible grammar which we may call G.

Now Katz (1966, 251-261) has argued that the meaning of a sentence often depends on unobservable grammatical features of that sentence. One must have recourse to the appropriate grammar in order to be able to "fill in" the unobservable grammatical features of a sentence, on the basis of its observable features.

(10) You must not put pins in your mouth.

(11) You need not put pins in your mouth.

(12) You may not put pins in your mouth.

Thus the child's grammar G assigns to sentence (10) certain unobservable grammatical features, and eventually, a semantic interpretation based in part on those features. In particular, G marks (10) as synonymous with (11), and as heteronymous with (12).

As adult speakers of English, we know that the child's grammar G has led him to misinterpret sentence (10). Evidently G is not the appropriate grammar for the child to use in interpreting the utterances of his elders. The child will eventually discover that G is not the grammar underlying the speech of his community, no doubt, and he will then abandon G in favor of a better approximation, say G'. He may learn, for instance, that some sentence S which is generated by his grammar G is not grammatical in English. There is a (possible) language in which _S is_ grammatical, of course; that is the language L, generated by the rejected grammar G. But the child is not interested in grammaticality (or synonymy) in just any possible language. He wants to be able to recognize grammaticality and synonymy, _inter alia_, in _English_.

In claiming that sentences have unobservable grammatical features with semantic relevance, then, Katz means to imply that some features of the structure of sentences can be determined only with reference to a particular (possible) grammar or language. Apparently any given possible sentence may be generated by a variety of possible grammars. Consequently, when someone describes a sentence as grammatical, or a pair of sentences as synonymous, it is often in order to ask, "Grammatical (synonymous) with respect to which language (or grammar)?"

In what follows, then, I shall assume that the native speaker's knowledge of "the grammatical properties of individual expressions of [his] language" (Graves, Katz, _et al_. in press)

(13) _The cat is on the mat_ is grammatical in English.

consists in his knowing that such propositions as (13) are true. Graves, Katz, _et al_. (in press) seem prepared to go along with this. Now according to the rationalist account, a native speaker comes to know propositions like (13) by deducing them from his general knowledge of the grammar of his language. The speaker's knowledge of principles of particular grammar is postulated precisely because this postulation contributes to an explanation of the speaker's knowledge of such propositions as (13).

Evidently one's knowledge of particular grammar can be formulated in some proposition or other. Graves, Katz, _et al_. explain that their argument for knowledge of particular grammar implicitly assumes the following principle : "a proposition is tacitly known if it is ... one of the propositions from which the person has tacitly deduced a (in this case explicitly) known proposition" (Graves, Katz, _et al_. in press). Their appeal to this principle clearly implies that we have _propositional_ knowledge of particular grammar. I wish now to inquire what sort of proposition might express our knowledge of English grammar.

The answer, I think, is adumbrated in a footnote to the paper by Graves, Katz, _et al_. Those writers indicate that "x is grammatical in L =df, an optimal grammar of L generates x" (Graves, Katz, _et al_. in press). This suggests that our

(14) Grammar G_e is the (optimal) grammar of English.

knowledge of particular grammar may be formulated in some such proposition as (14). Our knowledge of universal grammar gives us "a method for determining what each [possible grammar] implies with respect to each [possible] sentence" (Chomsky 1965, 30). In particular, we are innately endowed with the ability to determine whether a given grammar generates a specified possible sentence. We are thus able to discover the truth of (15). But

(15) Grammar G_e generates The cat is on the mat.

(16) The optimal grammar of English generates The cat is on the mat.

from (14) and (15), we can deduce (16), and according to the definition quoted above, (16) means just the same thing as (13). We were concerned to explain the speaker's knowledge of (13). We have done so by postulating knowledge of (14), which (I maintain) expresses the speaker's knowledge of "the particular grammatical rules of [his] language" (Graves, Katz, et al. in press). Also, we had to assume an innate ability to discern the consequences of a grammatical hypothesis.

One might ask why the grammar G_e itself cannot be taken to formulate what an English speaker knows. The answer is that the grammar G_e does not have the status of a proposition; a grammar, in and of itself, is neither true nor false. The grammar G_e is rather more like a definite description, which is true of English, and false of every other possible language. This is why I maintain that our knowledge of English grammar is expressed in a proposition that mentions G_e (namely (14)), rather than by G_e itself.

It is not altogether satisfactory to view a grammar as a definite description of a possible language, for on that view, we ought to get a true statement if we filled the blank in the following schema with the grammar of English:

(17) __________ is the English language.

Filling this blank with the grammar of English would not give us a truth, however. Unless a grammar is a gigantic noun phrase, the result will not even be syntactically well-formed, and surely it would be odd to regard a grammar as an NP. Perhaps a grammar should be viewed instead as an open sentence having roughly the following force:

(18) In the ____________ language, sounds and meanings are paired according to rules R_1, R_2, ..., R_n.

Like a definite description, such an open sentences would be true of one (possible) language, and false of all others. With its blank filled in, (18) would have much the same import as (14); until its blank is filled in, (18) does not express a proposition that one could know, since (18) is merely an open sentence, neither true nor false in itself, but only true of this language, and false of that.

Apparently, then, it is not the grammar of English that expresses our knowledge of English, but rather some statement such as (14), which mentions the grammar of English (or else a filled-in version of (18), which would contain the grammar of English, rather than a term referring to it). Thus in acquiring our knowledge of English, the proposition we deduce from universal grammar plus primary linguistic data must be something like (14). We are now in a position to ask what sort of propositions must be supposed to constitute the theory of universal grammar, in order for (14) to be deducible from that theory, in conjunction with primary linguistic data. (Graves, Katz, et al. in press). I wish to argue that unless certain principles of universal

> (19) Every (psychologically) possible human language has a grammar with property P.

grammar have roughly the force of schema (19) (which appears in the body of Chapter 2, p.99, as (13)), these principles will not be strong enough to play their role in the deduction of propositions like (14). If this is correct, then we must admit that some of our innate linguistic knowledge consists in knowing that certain properties are properties of every possible grammar. If we refuse to admit this, then the explanatory power disappears from the postulation of innate linguistic knowledge. That explanatory power depends, after all, on the deducibility of statements like (14) from primary linguistic data plus universal grammatical principles.

On the rationalist theory of language acquisition, "a child who is capable of language learning must have [among other things] some initial delimitation of a class of possible hypotheses about language structure" (Chomsky 1965, 30). Otherwise put, he must have "a definition of "generative grammar"", or a "specification of the class G_1, G_2, ... of possible generative grammars" (Chomsky 1965, 31). This characterization of a possible human grammar (and indirectly of a possible human language) is available to a child by virtue of his knowledge of linguistic universals.

Katz distinguishes formal, substantive, and organizational universals.

> Formal universals constrain the form of the rules in a grammar; substantive universals provide a theoretical vocabulary from which the constructs used to formulate the rules of particular grammars are drawn; organizational universals ... specify the interrelations among the rules and among the systems of rules within a grammar. (Katz 1972, 30-31)

How might these linguistic universals be expressed? I suggest that in general their form is that of (19). More specifically, formal universals might be expressed by such

(20) Every possible grammar contains transformation rules as well as phrase-structure rules.

(21) The terms in any rule in any possible grammar are a subset of the following vocabulary: _S_, _NP_, _VP_, ..., [_continuant_], ..., (_Human_), ...

(22) Every possible grammar contains a system of phonological rules that operate on the output of the transformation rules in the syntactic component.

statements as (20); substantive and organizational universals are exemplified in sentences (21) and (22) respectively.

The theory of universal grammar contains, in addition to principles having the form of (19) (linguistic universals), an evaluation procedure or plausibility metric. "With this metric the child can rank the set of initial hypotheses prior to any linguistic experience, and thus can obtain a best hypothesis or set of best hypotheses given no evidence about the language community he is in" (Katz 1966, 277). The linguistic universals define a set of possible grammars, and the evaluation procedure (partially) orders this set of grammars with respect to _a priori_ plausibility. This is the contribution of universal grammar to the process of language acquisition, the deduction of a particular

(23) In order of plausibility, the possible grammars are: G_1, G_2, ..., G_e, ...

grammar. Universal grammar can thus be seen as contributing to the deduction of a particular grammar some such premise as (23).

At this point, primary linguistic data enters the deduction. The data which a child has encountered up to a given moment will be logically compatible with some of the grammars in the list of (23), and logically inconsistent with other grammars in the list.

(24) _The cat is on the mat_ is grammatical in English.

(25) _The cats is on the mats_ is not grammatical in English.

These data will consist of such propositions as (6) (repeated here for convenience as (24)) and (25). Grammar G_1 is logically inconsistent with (24)[1] if it fails to generate _The cat is on the mat_; it is inconsistent with (25) if it generates _The cats is on the mats_. In either case, G_1 can be eliminated from the list of hypotheses. Thus "the role of experience is primarily to ... eliminate false hypotheses about the rules of a language" (Katz 1966, 278n.).[2] The child's grammatical hypothesis at any given moment is "the maximally [plausible] hypothesis all of whose predictions are compatible with the facts ... available to him ... _up to that time_" (Katz 1966, 278; emphasis Katz's).

In short, the primary linguistic data will allow the child to deduce such results as (26) and (27). Eventually, the child's

(26) G_1 is not compatible with the data.

(27) G_2 is not compatible with the data.

(28) G_e is the most plausible grammar compatible with the data.

data refutes every grammatical hypothesis that precedes G_e in the plausibility ranking of (23). At that point, the child reaches conclusion (28) via a deduction from his innate linguistic knowledge plus his primary linguistic data. This justifies the child in believing, and acting on, proposition (29)

(29) G_e is the (optimal) grammar of English.

(which appeared earlier as (14)).

Having sketched the overall structure of the deduction of (28), I wish now to return to premise (23), which summarizes the innate knowledge that enters into the language learning process. If (28) is to follow from (23), in conjunction with refutation reports such as (26) and (27), then (23) must entail three things. First, (23) must entail that the list of grammars it contains is arranged in order of decreasing plausibility; second, that the list contains no psychologically impossible grammars; and third, that the list contains all the psychologically possible grammars.

If the third requirement is not met, then there is some possible grammar, say G_i, which does not appear in the list.

A given body of data may fail to refute G_i, while refuting every listed grammar that precedes G_j, one of the grammars on the list. If we knew that G_j was more plausible than G_i, we could conclude that G_j was the most plausible unrefuted grammar.[3] But since G_i does not appear in the list, it might be (for all we know) that G_i is more plausible than G_j, and thus that G_j is <u>not</u> the most plausible unrefuted possible grammar.

How can we guarantee that the list in premise (23) includes all the possible grammars? So far as I can see, the only way

(30) All possible grammars have properties P_1, P_2, ..., P_n.

(31) All grammars that have properties P_1, P_2, ..., P_n are psychologically possible grammars.

to guarantee this, in a theory that already contains a principle like (30) (the conjunction of all the linguistic universals), is to include in our theory some such principle as (31). But (31) quantifies over possible grammars. So, for that matter, do the language universals; (30) says on the face of it that certain properties are properties of <u>all</u> possible grammars.

Of course, Graves, Katz, <u>et al</u>. do not deny that there are <u>properties</u> which are innately known to belong to every possible grammar. They rather deny the existence of any <u>principles</u> that are innately known to occur in every possible grammar. Their example is the A-over-A principle, a "condition for extracting phrases in certain contexts" (Graves, Katz, <u>et al</u>. in press). But if the A-over-A principle is really a principle we know

(32) In any possible grammar, a transformation that applies to constituents of category A may never be applied to an A that is embedded in some other A.

innately, then it must be stated as a constraint on transformations in all possible grammars, as in (32). Innate knowledge of the A-over-A principle, then, would consist in knowing that a certain principle governs all possible grammars.[4]

I take Graves, Katz, *et freres* to be maintaining that the principles of universal grammar, which we know innately, do not quantify over possible grammars. The foregoing discussion has, I think, sufficiently refuted that view. But in exploring the nature of the deduction of a particular grammar from universal grammar plus linguistic data, we have, I think, stumbled onto a problem for the rationalist explanation of our knowledge about individual sentences. The problem is related to Gettier's counterexamples to the traditional analysis of Knowledge.

(33) *The cat is on the mat* is grammatical in English.

(34) G_e is the grammar of English.

English speakers know such propositions as (33). They come to know these propositions, according to the rationalist, by inferring them from (34), another proposition they believe for good reasons. Now it might be argued that this explanation of our knowledge of (33) is at best incomplete, in the light of Gettier's counterexample. That counterexample showed that inference from justified premises does not always produce

(35) The man who will get the job has ten coins in his pocket.

(36) Jones will get the job, and Jones has ten coins in his pocket.

knowledge, even when the conclusion happens to be true. Smith was justified in believing (36), but when he reasoned from (36) to (35), the result was not knowledge, despite the truth of (35) (Gettier 1963, 122). To explain someone's <u>knowledge</u> of (33), then, the rationalist must show that the inference from (34) to (33) meets some additional condition, not satisfied by the inference from (36) to (35), which guarantees that the resultant belief in (33) will count as knowledge.

There does seem to be one relevant difference between these two inferences. The premise in Gettier's example, proposition (36), is false, and hence it is not something Smith <u>knows</u>. Premise (34), however, is true, and presumably the native speaker

(37) If A knows *p*, and A infers *q* from *p*, then A knows *q*.

knows (34). The rationalist can argue, then, that Smith's inference does not satisfy the antecedent of principle (37), but that the native speaker's inference does satisfy the antecedent of (37).

Unfortunately, (37) is false. We can easily modify Gettier's

(38) The president of the company says that Jones will get the job, and Jones had ten coins in his pocket ten minutes ago.

example to refute (37). Smith's mistaken belief in (36) was based on something he knew, namely (38). Suppose that instead of inferring (35) from (36), Smith had inferred (35) directly from (38). In this case, Smith would have inferred (35) from something he knew, namely (38), but Smith would still not know (35). This refutes (37).

The rationalist still has resources, however. The native speaker's inference from (34) to (33) is deductively valid, unlike the inference from (38) to (35). This means that the rationalist

> (39) If A knows p, and A validly deduces q from p, then A knows q.

can appeal to principle (39) in support of this explanation of the native speaker's knowledge of (33). Smith's inference does not satisfy the antecedent of (39), but the native speaker's inference does. That is why the native speaker's inference, unlike Smith's, results in knowledge. Principle (39) does not seem vulnerable to any Gettier-style counterexamples, such as the one that refuted (37).

Here the rationalist's problem appears to end, but this appearance is deceiving. We have been assuming that the native speaker knows premise (34), but this assumption cannot be justified by appeal to principle (39), even if we grant innate knowledge of universal grammar. For premise (34) does not follow

> (40) G_e is the most plausible grammar compatible with the available data.

deductively from the conjunction of primary linguistic data with the theory of universal grammar. What does follow deductively is (40) (previously designated as (28)). But (40) says only that the grammar G_e is the most likely candidate relative to the particular body of data which the language learner happens to have encountered. The possibility remains that this body of data is in some way deficient, and that G_e is not the grammar of English after all. If this is the case, then (34) is false, and the native speaker's inference from (34) to (33) does not satisfy the antecedent of (39). If (34) is false, then the native speaker's inference is precisely like Smith's inference from (36) to (35), and the rationalist is faced with Gettier's original counterexample.

This is no mere logical possibility. Sylvain Bromberger has suggested to me that perhaps no one acquires precisely the grammar of standard English. The phenomenon of idiolect variation is well attested, and it may be that every speaker of English has an idiolect that differs in some respect from standard English. If G_m is the grammar of my idiolect, and my idiolect differs slightly from standard English, then (41) is, strictly speaking,

(41) G_m is the grammar of English.

false. But (41) is the premise from which I deduce my judgments about the grammatical properties of English sentences.

One might ask what sense there can be in talking about standard English, if no one speaks it. In answer to this objection, we may adopt something like (42) as a convention

(42) If sentence S is grammatical in a (substantial) majority of English idiolects, then S is grammatical in standard English.

defining grammaticality in standard English, and similar conventions for synonymy and other properties and relations among sentences. There is no logical impossibility in both (43) and

(43) Every sentence in E_s is grammatical in a majority of idiolects $E_1, E_2, \ldots, E_n$.

(44) Each of the idiolects $E_1, E_2, \ldots, E_n$ lacks at least one sentence of E_s.

(44) being satisfied simultaneously. For instance, let n = 3, and let a, b, c, and d be the sentences of E_s (standard English), where $a \neq j$, $b \neq h$, $c \neq g$, and $d \neq f$. Then the family of idiolects

(45)

$$E_s = \{a,b,c,d\}$$

$$E_1 = \{a,b,c,f\} \qquad E_3 = \{a,h,c,d\}$$

$$E_2 = \{a,b,g,d\} \qquad E_4 = \{j,b,c,d\}$$

in (45) satisfies both (43) and (44). Thus it is possible for standard English (in the sense of (42)) to exist, even if no one happens to speak it. And it seems compatible with all we know about idiolect variation to suppose that this sort of situation really does obtain.

It is an empirical question whether idiolect variation is as extensive as (44) suggests. But however that question is eventually answered, it seems clear that many of us, though native speakers of English, speak idiolects that differ in minor ways from standard English. (Not all of us can be "right", since we differ over the grammaticality of certain constructions.) We know propositions like (33), and we deduce them from whatever grammar we have acquired (cf. note 1). But since our grammars are not altogether correct, why should such deductions result in knowledge? The rationalist account seems to provide no explanation of our knowledge about English sentences.

One might attempt to circumvent this difficulty be defining some sense of approximate truth in which grammars of our idiolects are "almost" true. Smith's premise (36) must not be anywhere near the truth, under this definition of approximate truth. A tempting suggestion is to say that a grammar of English (say G) is approximately correct if the class of sentences concerning which G yields true consequences is larger than the class of sentences about which G yields false consequences. This suggestion seems doomed to fail, however, because both sets are likely to be infinite.

The grammar of an idiolect may differ from the standard grammar in a variety of ways. It may have a deviant rule or lexical entry, or it may have a particular rule in the right form, but incorrectly ordered with respect to other rules. Since the grammar treats many sentences correctly, evidently this rule or lexical entry applies correctly (or does not apply at all) to these sentences. But since the length of a sentence is in general grammatically irrelevant, the length of a given sentence S is not likely to have any bearing on whether the grammar G treats S correctly or incorrectly. There is in general no upper bound on the length of sentences in which a lexical item may appear, nor on the length of sentences to which a given rule or set of rules may apply. Thus both the set of sentences which G handles properly, and the set of sentences it does not handle properly, are likely to be infinite sets.

So far from providing a ready-made defense of the rationalist account of linguistic knowledge, Gettier's work poses a serious problem for that account. Perhaps the only way to solve this problem would be to provide a general analysis of knowledge that escapes Gettier's refutation.

Footnotes

* The idea for this paper grew out of a conversation I had with Sylvain Bromberger. I also wish to thank him and Ned Block for helping me to revise an earlier version of this paper.

1 Strictly speaking, of course, G_1 is only an open sentence, and hence it cannot literally be incompatible with (24). In saying that a grammar is inconsistent with a datum about English, we mean to say that that datum refutes the claim that that grammar is the optimal grammar of English.

2 Katz implies that the child's linguistic data does not affect the operation of the evaluation procedure. This is not an essential feature of all rationalist hypotheses about language learning, however. For example, a given datum might (on another rationalist hypothesis) bring about the rejection of G_1, even though the datum is not logically incompatible with G_1. That datum might bring about a revision in the plausibility ranking of possible grammars, so that G_1 trades places with G_i, a grammar that was highly implausible up until the time when the new datum became available.

Any hypothesis in which the evaluation procedure applies only to the ranking of linguistic hypotheses is a rationalist hypothesis. There might even be rationalist hypotheses in which the relative plausibility of two grammatical hypotheses cannot be assessed at all, save on the basis of some linguistic data, though Katz of course presents quite a different sort of hypothesis.

3 This assumes that G_i is the _only_ unlisted possible grammar.

4 In an earlier passage, the authors mentioned two other examples of "principles of particular grammars which are also principles of linguistic theory". These are "the base rule for the recursive expansion of NP to contain sentence-structures" and "the transformation rule that forms a relative clause" (Graves, Katz, _et al_. in press). But if every possible grammar does contain a base rule that generates complex NP's, and a relative-clause transformation rule, the details of these rules may vary from one language to another. If such variation does occur, then the theory of universal grammar obviously cannot contain "the" rules which occur in each particular grammar. Universal grammar can at best _describe_ the rules, in broad outline.

But even if all possible grammars contained precisely the same relative-clause transformation, the theory of universal grammar still could not _contain_ that rule. Universal grammar is not a grammar of a possible human language, not a theory about the structure of sentences in such a language. Rather, universal grammar is a theory about grammars of possible human languages. A statement of the theory of universal grammar might thus _mention_ the (putatively universal) relative-clause rule, or describe it, but a statement of linguistic theory would not _use_ or _contain_ that rule. The relative-clause rule itself has nothing to say about the structure of grammars, but deals only with the structure of _sentences_. To shed light on the structure of possible grammars, we must say something _about_ that rule, namely, that it occurs in every possible grammar.

References

Chomsky, N. (1965) _Aspects of the Theory of Syntax_, MIT Press, Cambridge, Mass.

Gettier, E. (1963) "Is Justified True Belief Knowledge?," _Analysis_ 23, 121-123.

Graves, C., J. Katz, Y. Nishiyama, S. Soames, R. Stecker, and P. Tovey (in press) "Tacit Knowledge," _Journal of Philosophy_.

Katz, J. (1966) _The Philosophy of Language_, Harper and Row, New York.

Katz, J. (1972) _Semantic Theory_, Harper and Row, New York.

Quine, W. (1961) "Two Dogmas of Empiricism," in _From a Logical Point of View_, Harper and Row, New York.

Biographical Note

In June, 1968, Mr. Mellema received the A.B. degree in Philisophy from Calvin College. He spent the 1966-67 academic year in Oaxaca, Mexico, working as a language survey technician under the auspices of the Wycliffe Bible Translators. Following his graduation from college, Mr. Mellema spent a year participating in Wycliffe language surveys throughout Canada, and in central Mexico. He has taught a course in English for Spanish-speaking people as part of the Grand Rapids (Michigan) Adult Basic Education program, and assisted in teaching introductory transformational grammar at the 1970 Summer Institute of Linguistics at the University of North Dakota.

Chapter 1 of this thesis, "A Brief Against Case Grammar," is to appear in _Foundations of Language_.